Journal of Molecular Structure (Theochem), 255 (1992) 45–91
Elsevier Science Publishers B.V., Amsterdam

Efficient new methods for the determination of integrated atomic properties via atom specific electron density functions based on subsets of selected localized molecular orbitals and the reduction of the space of the primitives[α,β]

Rainer Glaser[1] and Benjamin L. Harris

Department of Chemistry, University of Missouri-Columbia, Columbia, MO 65211 (USA)

(Received 16 May 1991)

Abstract

Atomic properties are central to discussions of bonding and reactivity, and topological electron density analysis provides a powerful framework for their rigorous determination. Whereas the topological method can be applied routinely to comparatively small molecules, its application to large molecules is somewhat impeded by the rather considerable amounts of computer time required. Here, we describe two new methods for the determination of integrated atomic properties that greatly reduce the integration times while maintaining the generality, the rigor, and the accuracy of the topological partitioning method. The principle of the methods consists in reducing the total molecular electron density function to atom-specific electron density functions that accurately describe the electron density distribution in the basin of the atom whose properties are being determined. It is shown that this task can be accomplished either by a reduction of the space of the primitives in which the wave function is expanded or via electron density functions defined by atom-specific subsets of selected localized molecular orbitals. With the current integration algorithm the latter method is more efficient. The two methods can be combined. The theoretical principles and the computational implementation of the methods are discussed. Their performances have been tested for a series of polyines $C_{2n}H_2$ and polynitriles $C_nN_nH_{n+2}$ ($n=1$–5 and 10) and the topological characteristics and the integrated properties are found to be in excellent agreement with results obtained by the conventional technique. Importantly, it is shown that with the subset selection, integration time requirements approach an upper limit as the size of the molecule increases. The methods perform equally well for all kinds of basis sets, allowing for the analysis of large molecules described by spliced basis sets. The rigorous electron density analysis of very large molecular systems becomes feasible.

[α]Part of the projected Ph.D. dissertation of Benjamin L. Harris.

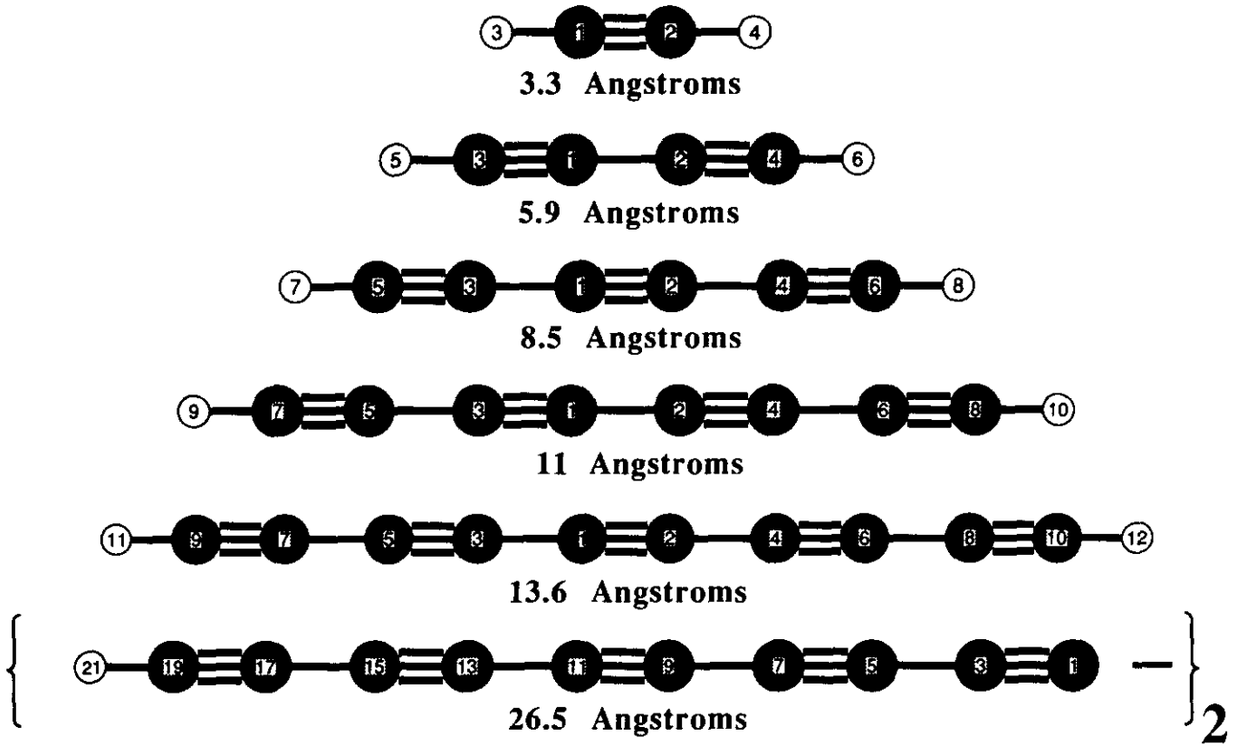
[β]Presented in part at the 201st ACS National Meeting, Atlanta, Georgia, April 1991, and at The St. Louis Regional Gathering on Computer-Aided Molecular Design and Computational Chemistry, University of Missouri-St. Louis, May 1991.

[1]Author to whom correspondence should be addressed.

INTRODUCTION

Atomic properties are central to discussions of bonding and of intra- and intermolecular interactions, to explain the reactivity of molecules, and to discuss reaction mechanisms. For a recent discussion of various population analysis methods see ref. 1. Basis set partitioning and electron density integration techniques have emerged as the two fundamentally different approaches to deriving atomic properties. Among the population analyses based on basis set partitioning are the Mulliken population analysis [2], the SEN method [3], and the natural population method [4], to name a few. Density integration techniques differ fundamentally from the basis set partitioning methods, in that they consider only the (experimentally observable) electron density. Problems arising in the basis set partitioning techniques related to the assignment of density to the basis function centers do not occur. The problem of partitioning the electron density between the nuclei remains, and different methods have been proposed to resolve it [5–7]. For planar molecules integrated populations derived from electron density projection functions approximate the Bader populations in an efficient way. Bader's theory of atoms in molecules defines the most rigorous partitioning scheme based exclusively on the topology of the electron density distribution [8,9].

The topological method has two major advantages compared to all other methods. First, it is the only method in which the partitioning of the molecular system into atomic regions is done in a rigorous way that is based on the axioms of quantum mechanics [8]. All other methods involve further assumptions. Secondly, the topological method allows for the determination of a variety of atomic properties. With the atomic region in the molecule defined, the charge, for example, can be derived by numerical integration of the electron density within that space. Other interesting properties that can be derived include the atomic moments, the atomic kinetic energy and some thirty other properties. The one limiting disadvantage of the topological method relates to its extreme requirement of computer time. The time required for the determination of the integrated atomic properties of a molecules greatly exceeds the time required for the determination of the wave function. The determination of these integrated atomic properties essentially involves three steps [10]. We thank Professor Bader for the programs EXTREME and PROAIM. EXTREME and PROAIM were ported to the Silicon Graphics Personal Iris by R. Glaser. In the first step, the critical points—the extrema—of the electron density distribution are located and their characteristic values—the so-called topological properties including ρ value, gradient, curvatures, and similar properties — are determined. This step is fast, even for very large molecular systems. For a report on the determination of molecular graphs and topological properties of rather large systems, see, for example, ref. 11. Next, the partitioning surfaces are determined by tracing the paths of steepest descent in the electron density distri-



Length of repeating unit about 2.6 Angstroms
Number of electrons NE = 12n+2

Fig. 1. Molecular models of the optimized structures of the polyines H-$(CC)_n$-H (n=1-5 and 10) examined.

bution. The surface definitions start at the bond critical points and in a variety of directions perpendicular to the bond paths. With the partitioning surfaces determined, the integrated atomic properties can then be determined by numerical integration within the atomic regions, the basins. This integration is the most computer time intensive step. Even for a molecule of moderate size, the time required for the determination of the integrated atomic properties of each atom is of the same order as the time required for the geometry optimization of the molecule. It is this feature that limits the application of this powerful method to comparatively small molecules, and essentially prohibits the electron density analysis of large systems. In ab initio quantum chemistry, even a molecule with as few as ten non-hydrogen atoms has to be considered as "large" and it becomes thus quite obvious that the time limitation is indeed a serious disadvantage, especially with a view toward applications in QSARs.

In this article we present two new methods that allow for an efficient determination of integrated atomic properties. With these methods, the integration times required for the determination of integrated atomic properties can be drastically reduced, while maintaining the generality, the rigor, and the accuracy of the topological method. The principles of these methods consist in reducing the total molecular electron density function to atom specific functions that accurately describe the electron density distribution in the basin of the

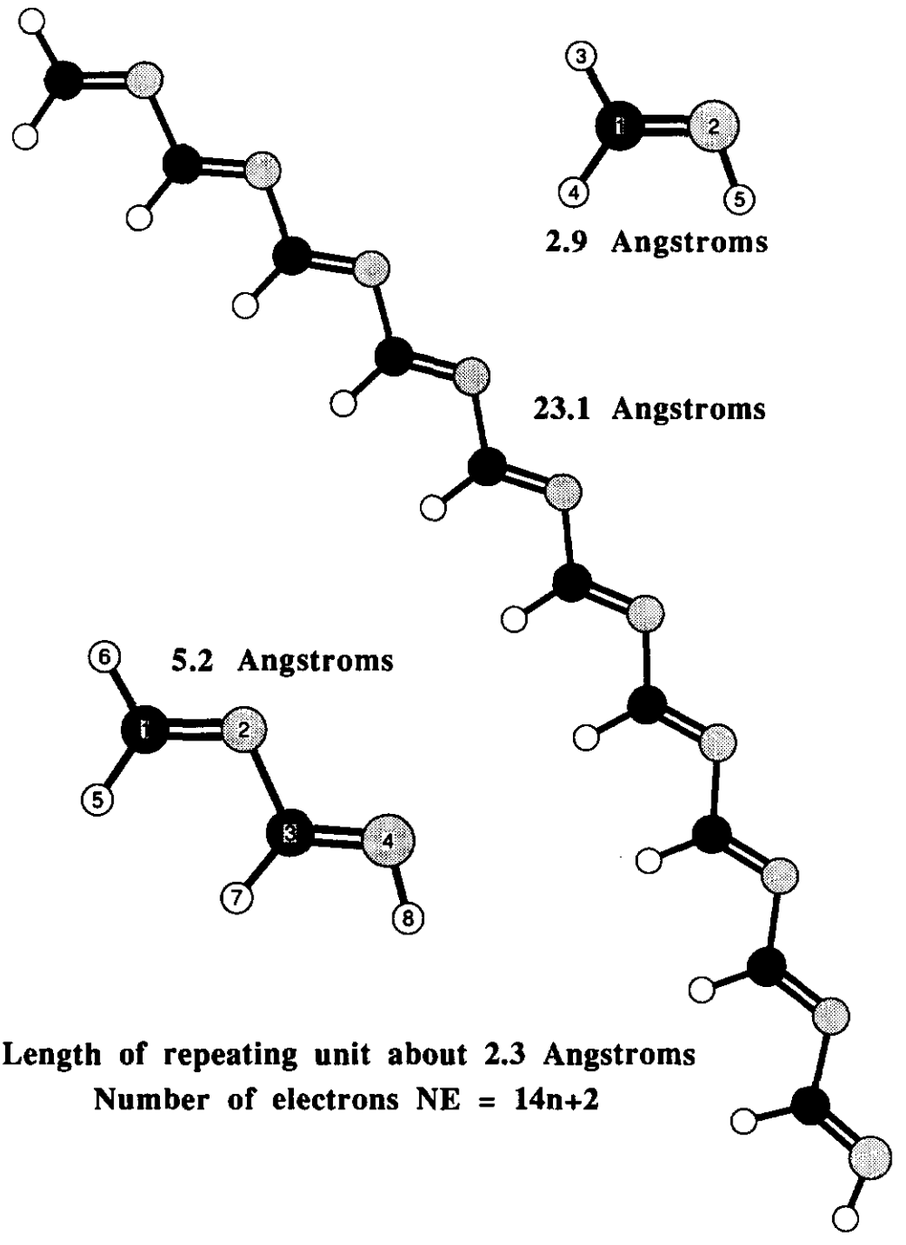


Fig. 2 (opposite and above). The optimized structures of the planar all-E polynitriles H-(HCN)$_n$-H (n=1-5, and 10).

atom whose properties are being determined. The theoretical principles and the computational implementation are discussed. With these methods, not only will the analysis of the electronic structures, the atomic properties and the reactivities of large molecules be greatly facilitated but, moreover, it will also be significant in examining and judging important concepts in molecular modeling. Quantitative structure–activity relationships of natural products, biologically and/or pharmaceutically important molecules, and the properties of polymers are usually approached with molecular mechanics or with semi-empirical procedures, but rarely at the ab initio level. It is hoped that efficient density integration techniques will make it possible to derive rigorous atomic properties of large molecules at the ab initio level and, in so doing, will make

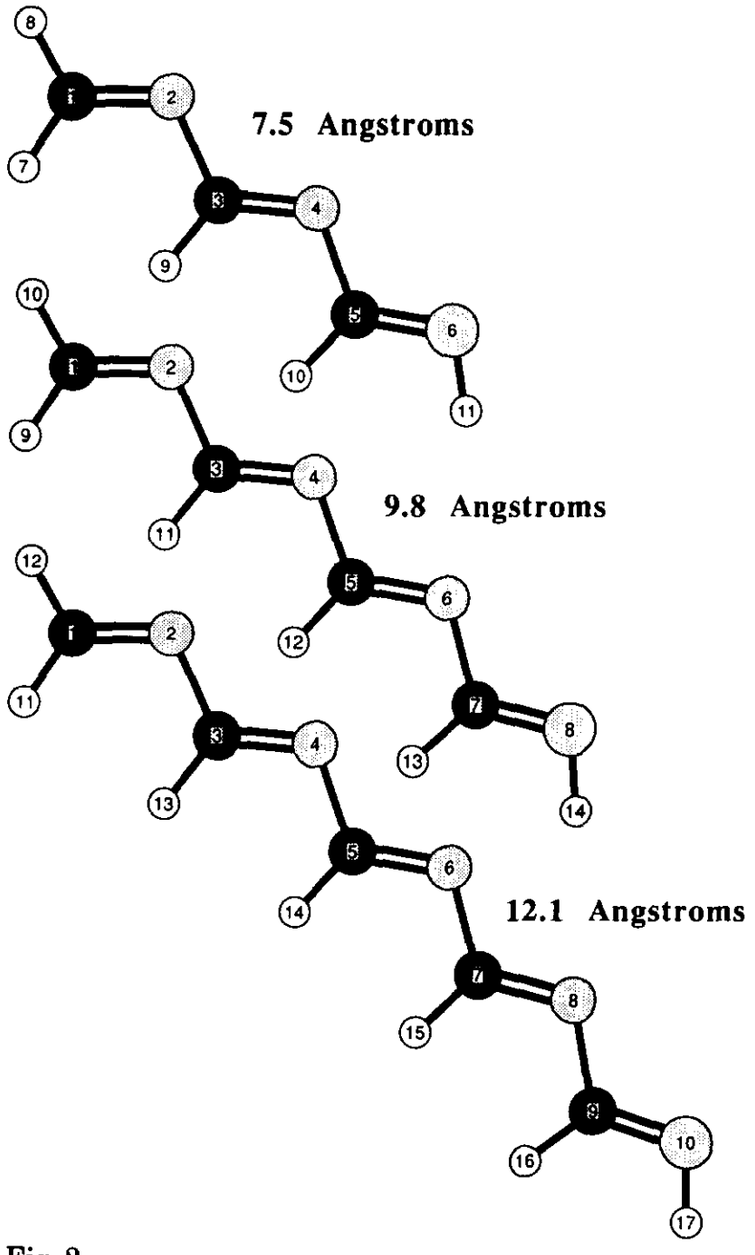


Fig. 2.

it possible to evaluate and refine the capabilities of modern molecular modeling techniques.

A series of polyines H–$(CC)_n$–H and a series of N-substituted polyacetylenes, the polynitriles H–$(HCN)_n$–H, were examined to test the new methods. In this article we restrict the discussion of these systems exclusively to the methodological aspects. The largest molecules in these series have in excess of 120 electrons and are longer than 23 Å. All of these polymers were optimized at the restricted Hartree–Fock (RHF) level with the minimal basis set STO-3G [12]. The resulting structures are shown in Figs. 1 and 2 and structural parameters are documented in Tables 1 and 2. This theoretical level was chosen with a view toward studies of rather large systems and, moreover, this level is perfectly adequate for systematically testing the procedure, with the least investment of computer time. A few systems were also examined with larger,

TABLE 1

Bond lengths (in Å) of polyines[a]

Parameter	C_2H_2	C_4H_2	C_6H_2	C_8H_2	$C_{10}H_2$	$C_{20}H_2$
C1-C2	1.1684	1.4083	1.1816	1.3979	1.1839	1.3956
C1-C3		1.1746	1.4034	1.1827	1.3968	1.1846
C3-C5			1.7156	1.4027	1.1829	1.3956
C5-C7				1.1758	1.4028	1.1845
C7-C9					1.1759	1.3959
C9-C11						1.1843
C11-C13						1.3968
C13-C15						1.1831
C15-C17						1.4025
C17-C19						1.1760
C-H	1.0654	1.0660	1.0665	1.0667	1.0666	1.0669

[a]All structures optimized with $D_{\infty h}$ symmetry with the minimal basis set.

split-valence and polarized basis sets [13], including the basis sets [14–16] 3-21G [14] and 3-21G*, 6-31G [15] and 6-31G*, and 6-311G [16] and 6-311G*. Sets of six Cartesian single d functions with exponent 0.8 were added to all first row atoms in the non-standard 3-21G* basis set. The success of the methods greatly depends on the accuracy of the partitioning surfaces in the bonding regions. The selected polymers present excellent test cases in this regard because they include single, double, and triple bonds in addition to polar and unpolar bonds. The polyines are of particular value because of very subtle features in the bonding regions of the triple bonds. It is known that the electron density distribution in triple bonds are rather flat [17,18] and, in some cases, such bonds exhibit small maxima in the electron density in the bonding regions. Such non-nuclear attractors have also been found in other systems [19]. We report the topological properties of the critical points and a small selection of integrated atomic properties, to demonstrate the quality of the results obtained with the new methods. The topological parameters selected include the distances of the bond critical points from the adjacent atoms, and the values of the electron density and of the principal curvatures of the electron density at the location of the bond critical point. The three integrated properties discussed throughout are the atom populations, the atomic dipole moment, and the atomic kinetic energy. Because of their definitions, these properties exhibit different dependencies on the accuracy of the partitioning method and they thus provide excellent parameters for the present task. Moreover, these quantities were selected because we find them the most useful integrated properties in our studies [20–23] of deamination reactions and their role in the alkylation chemistry of biopolymers.

TABLE 2

Structures of N-substituted polyacetylenes: the polynitriles H-$(HCN)_n$-H[a]

Parameter	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=20$[b]
Bond length						
C1-N2	1.2727	1.2765	1.2777	1.2781	1.2782	1.2784
N2-C3		1.4644	1.4598	1.4588	1.4585	1.4582
C3-N4		1.2752	1.2798	1.2813	1.2817	1.2821
N4-C5			1.4603	1.4549	1.4537	1.4530
C5-N6			1.2762	1.2806	1.2822	1.2830
N6-C7				1.4601	1.4545	1.4525
C7-N8				1.2760	1.2807	1.2830
N8-C9					1.4601	1.4526
C9-N10					1.2760	1.2830
H[c]-C1	1.0888	1.0917	1.0883	1.0883	1.0883	1.2830
H[d]-C1	1.0907	1.0881	1.0919	1.0919	1.0920	1.0884
H-C3		1.0966	1.0976	1.0977	1.0977	1.0920
H-C5			1.0964	1.0973	1.0974	1.0979
H-C7				1.0963	1.0972	1.0976
H-C9					1.0963	1.0974
H[e]-N	1.0483	1.0475	1.0478	1.0479	1.0479	1.0479
Bond angles						
C1-N2-C3	114.299	114.300	114.274	114.264	114.257	
N2-C3-N4	118.770	119.041	118.955	118.945	118.911	
C3-N4-C5		114.208	114.170	114.156	114.147	
N4-C5-N6		118.648	118.944	118.892	118.858	
C5-N6-C7				114.179	114.157	114.128
N6-C7-N8				118.654	118.910	118.796
C7-N8-C9					114.199	114.138
N8-C9-N10					118.629	118.811
H[c]-C1-N2	119.118	119.217	119.083	119.070	119.059	119.058
H[d]-C1-N2	125.396	124.725	124.690	124.694	124.694	124.706
H-C3-N2		116.243	116.900	116.967	117.003	117.042
H-C5-N4			116.444	117.038	117.141	117.223
H-C7-N6				116.490	117.057	117.218
H-C9-N8					116.493	117.241
H[e]-N-C	109.102	108.536	108.507	108.524	108.540	108.549

[a]All structures optimized in C_s symmetry. Bond lengths in Å and angles in Deg.
[b]See Appendix for a complete set of structural parameters.
[c]H trans with N3.
[d]H cis with N3.
[e]The terminal imine hydrogen atom.

BASIC PRINCIPLES

The molecular electron density function in real Cartesian space, $\rho(x,y,z)$, is the product of the wave function and its complex conjugate. If the wave func-

tion of the system is given by the self-consistent set of real molecular orbitals Ψ_i, the electron density distribution is defined as

$$\rho_{\text{mol}}(x,y,z) = 2\sum_i \Psi_i \Psi_i$$

The factor of 2 accounts for double occupancy and the summation is over the occupied orbitals only. Here we are discussing RHF wave function only, but the methods are also all applicable to cases with orbital occupancies other than 2. The molecular orbitals (MOs) Ψ_i are expressed as linear combinations of Slater-type, atom-centered atomic orbitals. Each atomic orbital is usually described by several basis functions, with the appropriate second quantum number and with the appropriate exponent, and the atomic orbitals themselves are (fixed) contractions of primitive functions ϕ_ν. The MOs can thus be expressed as either a linear combination of the basis functions ϕ'_ν with coefficients $c'_{\nu i}$ or as a

$$\Psi_i = \sum_\nu c'_{\nu i} \phi'_\nu \qquad \text{(basis function expansion)}$$

$$\Psi_i = \sum_\nu c_{\nu i} \phi_\nu \qquad \text{(primitive function expansion)}$$

linear combination of the primitive functions where the $c_{\nu i}$ result from the combination of the coefficients $c'_{\mu i}$ and the contraction coefficients. This yields for the molecular electron density function as expressed in the primitives

$$\rho_{\text{mol}}(x,y,z) = \sum_\mu \sum_\nu P_{\mu\nu} \phi_\mu \phi_\nu$$

where $P_{\mu\nu}$ is the one-electron density matrix

$$P_{\mu\nu} = 2\sum_i c_{\mu i} c_{\nu i}$$

The determination of the atomic properties by the density integration technique requires this function to be evaluated for small volume elements within the atomic basins. The time required for the evaluation of $\rho_{\text{mol}}(x,y,z)$ will increase with the size of the molecule for three reasons: (a) The number of MOs increases, (b) the number of basis functions increases, and (c) the number of primitives increases. For the determination of the wave function, the number of MOs cannot be reduced, of course, and the number of basis functions and the associated number of primitives should definitely not be reduced to guarantee the determination of properties of excellent quality. However, after the wave function is determined, the integration problem can be reduced in two ways. This equation for $\rho_{\text{mol}}(x,y,z)$ suggests two methods for a more efficient

numerical integration within the basin of a given atom. For the determination of the properties of a specific atom, all that is required is an electron density function that is the same as ρ_{mol} within the basin of that atom, but this function might differ outside the basin without consequences for the accuracy of the computed atomic properties. These two methods consist in the reduction of the numbers of primitives considered (ν reduction) and in the reduction of the number of MOs considered (i reduction).

SELECTION AND REDUCTION OF PRIMITIVE FUNCTIONS

Theoretical principles

In the expression for the molecular electron density function $\rho_{\mathrm{mol}}(x,y,z)$ the index ν denotes one primitive function among all of the primitive functions, without regard to the location at

$$\rho_{\mathrm{mol}}(x,y,z) = 2\sum_i [\sum_\nu c_{\nu i}\phi_\nu]^2$$

which the function is centered. Instead of indexing all the primitive functions in this way, we may collect the primitives that are centered at a given nucleus N into N sets and use k to index the primitives within each of these sets to define the functions Ψ_{iN}. The summation of all Ψ_{iN} yields of course the molecular orbital Ψ_i. The index k depends on the atom type and the basis set selected for atom N.

$$\Psi_i = \sum_\nu c_{\nu i}\phi_\nu$$

$$\Psi_{iN} = \sum_k c_{Nki}\phi_{Nk} \qquad (k=f(N))$$

$$\Psi_i = \sum_N \Psi_{iN}$$

We thus obtain for the molecular electron density function

$$\rho_{\mathrm{mol}}(x,y,z) = 2\sum_i [\sum_N \{\sum_k c_{Nki}\phi_{Nk}\}]^2$$

The primitive functions decrease with the distance from the atom at which they are centered. If the atom M is far away from atom K, then the electron density of atom K will not affect the electron density function at the location of atom M. For the correct description of the electron density distribution in the basin of atom M, the contributions made by the primitives of atom K are negligible. Thus, the determination of the atomic properties of an atom M may be carried out with an atom-specific electron density function $\rho'_{\mathrm{M}}(x,y,z)$ defined by

$$\rho'_M(x,y,z) = 2\sum_i [\sum_N \{\sum_k c_{Nki}\phi_{Nk}\}]^2 \qquad \text{where } N \in \{\text{K close to M}\}$$

that contains only the contributions to the electron density distribution of the atoms in the proximity of atom M. Electron density distributions of atoms in the proximity of atom M can be described by the set of all MOs containing only primitive coefficients for atoms in the proximity of atom M. The atoms that need to be considered for the evaluation of this function depend on the exponents of the primitive functions and on the distances between these atoms and atom M. For the types of extended systems considered here, the selection of these atoms can be done simply based on connectivity. More generally, this selection is accomplished by a criterion based on internuclear distances. Based on connectivity, we can easily determine the next neighbors of atom M and the second-next neighbors of M (the neighbors of the next neighbors) and so forth. We refer to the subset of atoms {K close to M} that contains all of the neighbors including the nth-next neighbors as $NNN = n$. The choice of NNN will of course affect the atom specific electron density function ρ'_M and we thus write

$$\rho'_M(x,y,z,NNN) = 2\sum_i [\sum_N \{\sum_k c_{Nki}\phi_{Nk}\}]^2 \qquad \text{where N} \in \{\text{K close to M}\}$$

Performance of the method

The use of this equation for the determination of integrated atomic properties has been examined for the polyines and the results obtained for $C_{10}H_2$ are discussed as an illustrative example. The atom-specific electron density functions $\rho'_M(x,y,z,0)$ have been determined and the functions $\rho'_M(0,0,z,0)$ are depicted graphically in Fig. 3. The molecule is aligned with the z axis and the inversion center is located at the origin. A logarithmic scale is used to emphasize the tails of the functions ρ'_M. In the specific case considered here, it is found that the functions ρ'_M become less than 10^{-4} e a.u.$^{-3}$ as the distance from the nucleus M becomes greater than 4 Å. In general, the fall-off of these functions will depend on the basis set and on the molecular system. Molecules that require diffuse functions for their proper description will lead to such functions ρ'_M that extend further away from the nuclei M.

If all of the primitives of the atoms K that are within 4 Å of atom M are considered in the computation of the atom specific electron density function $\rho'_M(x,y,z,NNN)$, then the determination of the integrated atomic properties of atom M should be rather exact. The functions $\rho'_{C9}(0,0,z,NNN)$ shown in Fig. 4 were determined with NNN values of 1 (dotted), 2 (short dashes), and 3 (long dashes). The molecular electron density function is superimposed as a solid line. The largest deviations between ρ_{mol} and ρ'_M are expected to occur in

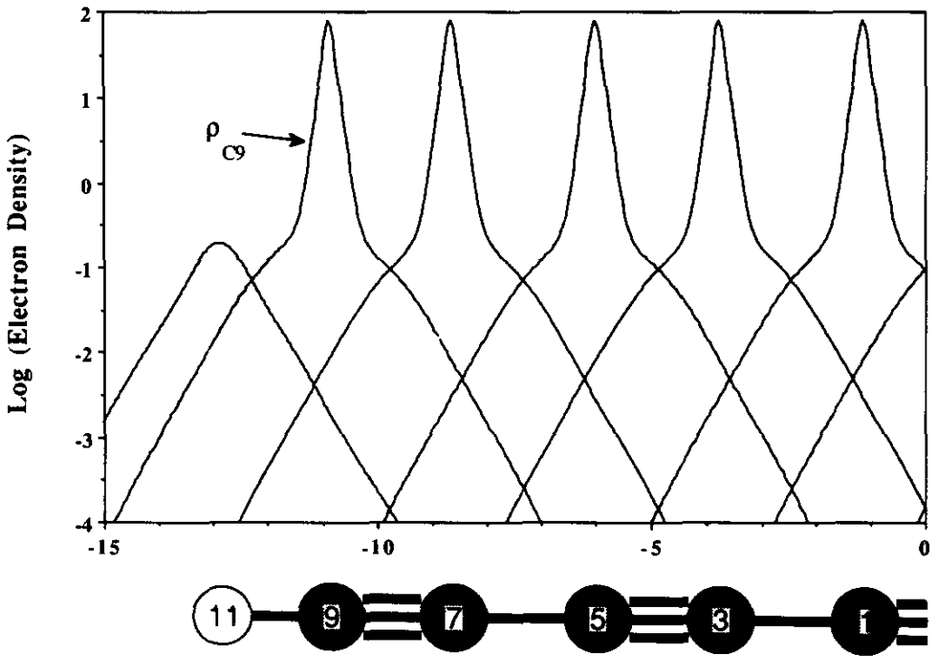


Fig. 3. The electron density functions $\rho_M(0,0,z,0)$ associated with the primitive functions of each atom in $C_{10}H_2$ plotted logarithmically versus the distances from the atom at which they are centered.

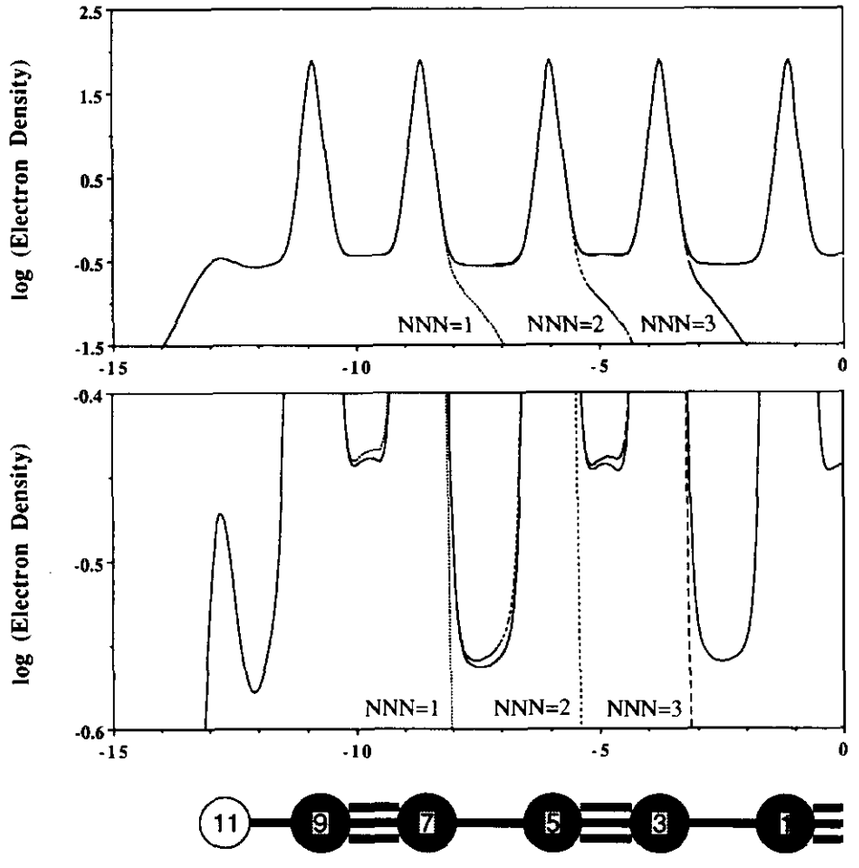


Fig. 4. Top, the molecular electron density function ρ_{mol} of $C_{10}H_2$ shown together with the corresponding electron density functions $\rho_{C9}(NNN)$ that result from the expansion of the wave function in the primitives of atoms H-C9-C7 ($NNN=1$, dotted), H-C9-C7-C5 ($NNN=2$, short dashes), and H-C9-C7-C5 ($NNN=3$, long dashes) only. Bottom, the region $-0.6<\log\rho<-0.4$ shown in an expanded fashion. Small deviations between the electron density functions expanded in all primitives or subsets of primitives become apparent in the bonding regions. The occurrence of non-nuclear attractors in the triple bonds is also clearly shown.

TABLE 3

Elimination of primitive functions: topological properties of C9 in $C_{10}H_2$ polyine

Entry No.[a]	A–B Bond	r_A[b]	r_B[b]	ρ[c]	$\lambda_1=\lambda_2$[d]	λ_3[d]
1[e]	C9–C7	0.4468	0.7291	0.3602	−0.2976	0.4901
		0.6054	0.5705	0.3641	−0.2896	−0.1381
		0.7253	0.4506	0.3621	−0.2834	0.3863
	H11–C9	0.4210	0.6456	0.2640	−0.5906	0.4636
9[f]	C9–C7	0.4468	0.7291	0.3602	−0.2976	0.4899
		0.6053	0.5706	0.3640	−0.2896	−0.1382
		0.7253	0.4505	0.3621	−0.2833	0.3870
	H11–C9	0.4210	0.6456	0.2640	−0.5906	0.4636
10[g]	C9–C7	0.4426	0.7334	0.3623	−0.3056	0.5629
		0.6390	0.5369	0.3680	−0.2979	−0.0925
		0.7002	0.4757	0.3677	−0.2970	0.1570
	H11–C9	0.4210	0.6456	0.2640	−0.5906	0.4635

[a]Entry numbers correspond to those in Table 4 (for an explanation of entry Nos. see text). For the C9–C7 bond, the three critical points characterized are the bond critical point near C9, the non-nuclear attractor in the C9–C7 bonding region, and the bond critical point near C7, in this order.
[b]Distances (in Å) of the critical points from atoms A and B.
[b]Electron density at the critical point (in e a.u.$^{-3}$).
[d]Curvatures in the density perpendicular (λ_1 and λ_2) and parallel (λ_3) to the molecular axis (in a.u.).
[e]Total electron density.
[f]Considering the primitive functions of fragment H–C9–C7–C5 only.
[g]Considering the primitive functions of fragment H–C9–C7 only.

the bonding regions and the plot shown at the bottom of Fig. 4 emphasizes the dependence of ρ'_M on *NNN* in those regions. Note that the non-nuclear attractors in the triple bond regions are clearly manifested. Considering only the next neighbors for the calculation of ρ'_{C9} reproduces the total electron density function in the C–H bonding region excellently, but small deviations do occur in the C9–C7 bonding region. With the selection of $NNN=2$, a function ρ'_{C9} is obtained that perfectly matches the total electron density function of the molecule; further increase of *NNN* has no effect.

The characteristic values of the critical points in the C9–H and C9–C7 bonding regions allow for a more quantitative comparison of the functions ρ_{mol} and ρ'_{C9}. In Table 3, the topological properties for these bonding regions are presented. Each critical point is characterized by its distance from the adjacent attractors A (r_A) and B (r_B), by the value of the electron density at that point, and by the principal curvatures of the electron density perpendicular ($\lambda_1=\lambda_2$) and parallel (λ_3) to the bond path. The C9–C7 bonding region contains two bond critical points and one non-nuclear attractor and the characteristic values of all these critical points are given in Table 3. The integrated atomic pop-

TABLE 4

Elimination of primitive functions: integration results for C9 in $C_{10}H_2$ polyine

Entry No.	Eliminated primitives[a]	IP[b]	μ_z[c]	T[d]	t[e]
1	None	5.375	0.745	36.984703	3.50
1	None[f]	5.375	0.745	36.984702	3.49
2	H12	5.375	0.745	36.984701	3.49
3	C10-12	5.375	0.745	36.984702	3.47
4	C8-12	5.375	0.745	36.984702	3.43
5	C6-12	5.375	0.745	36.984702	3.43
6	C4-12	5.375	0.745	36.984701	3.44
7	C2-12	5.375	0.745	36.984702	3.43
8	C1-12	5.375	0.745	36.984702	3.37
9	C3-12	5.376	0.745	36.984754	3.23
10	C5-12	5.352	0.764	36.968049	2.87

[a]Integrations are for atom C9, using all of the MOs after elimination of all of the primitives of the atoms shown in column 2 (see Fig. 1 for atom numbering).
[b]Integrated populations (in e.).
[c]Integrated dipole moment (in e a.u.$^{-5}$).
[d]Integrated kinetic energies in (a.u.), $C_{10}H_2$, E(RHF/STO-3G) = −374.819358 a.u. with $-V/T = 2.00664975$. The kinetic energies are corrected for the virial defect of the wave function.
[e]Time taken for integration (in h.).
[f]Values determined using all of the localized orbitals.

ulations, dipole moments, and kinetic energies of C9, derived with these functions ρ'_{C9}, are given in Table 4, together with the definition of the entry number. The results given for entry 1 are those determined with the molecular electron density function, and as the Entry No. increases, the more sets of primitives are eliminated, beginning with the elimination of the primitive functions centered at H12. Entry No. 10 thus refers to ρ'_{C9} with $NNN=1$ and Entry No. 9 to ρ'_{C9} with $NNN=2$ and so on. Table 4 and Fig. 5 show that the integrated properties are exactly reproduced for all entries 1–8. The same is true for the topological properties of these entries. For Entry 9 ($NNN=2$) the population and the dipole moment are reproduced virtually exact, whereas the kinetic energy is slightly (0.033 kcal mol^{-1}) too high. More significant differences occur as expected, if only the next neighbors are considered. The bond critical point in the proximity of C9 occurs closer to C9 and the populations and kinetic energy of C9 are consequently underestimated. Note that $\log[\rho'_{C9}(NNN)]$ is larger than $\log[\rho_{mol}]$ in the bonding regions because increasing NNN will add "tails" of electron density, for which $\log[\rho] < 0$.

These results clearly demonstrate that numerically exact integrated atomic properties can be derived from the atom specific electron density functions $\rho'_{C9}(NNN)$ so long as NNN is sufficiently high (here $NNN > 2$). Even with

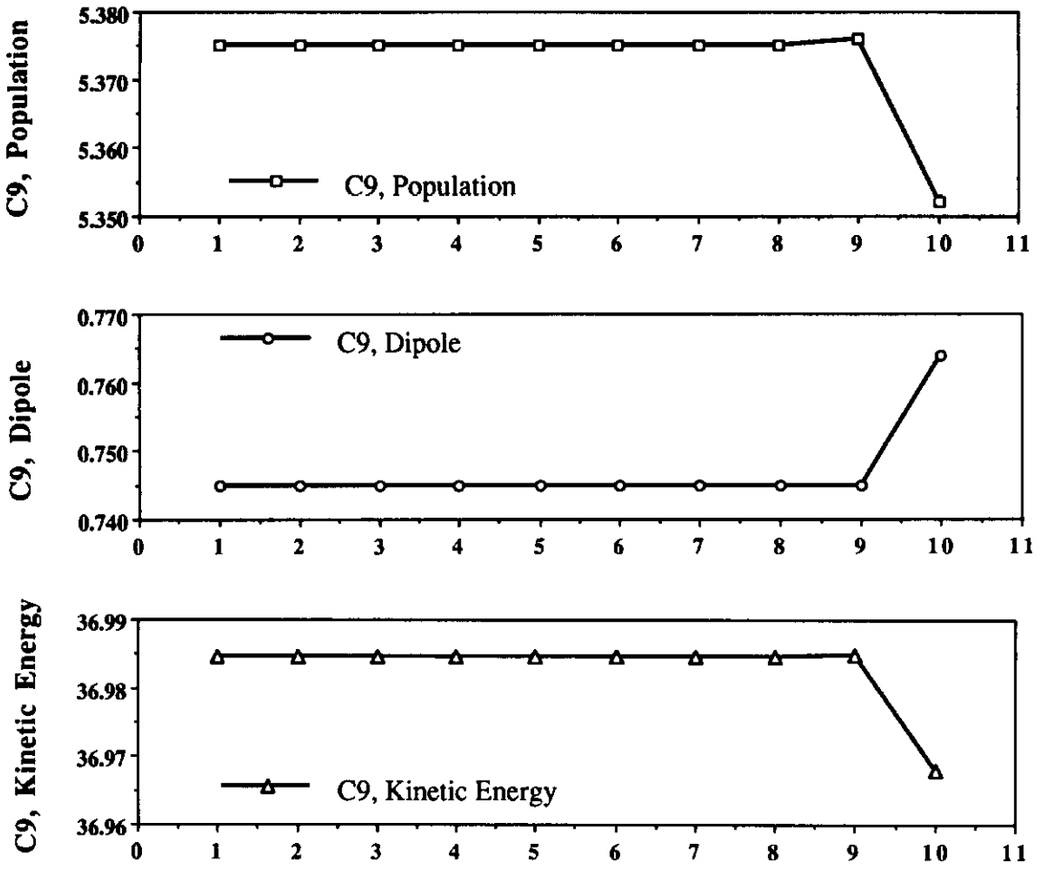


Fig. 5. Plots of the integrated populations, dipole moments, and kinetic energies of C9 in $C_{10}H_2$ versus the extent of elimination of atom-specific sets of primitive functions (see Entry Nos. in Table 4).

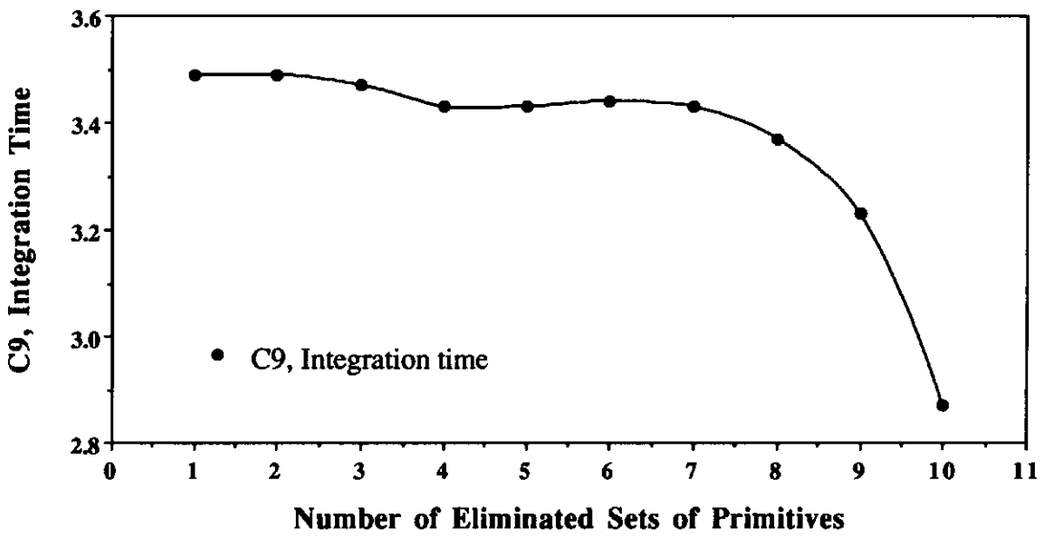


Fig. 6. Integration times for C9 in $C_{10}H_2$ plotted versus the extent of elimination of atom-specific sets of primitive functions (see Entry Nos. in Table 4).

NNN=2, populations and dipole moments can be reproduced in an excellent fashion. The time requirements for the integrations of the C9 basins with the functions ρ'_{C9} (*NNN*) are given in Table 4 and they are shown graphically in Fig. 6. It is found that the primitive elimination method leads to only relatively modest savings of computation time with the current integration program. We

point out that this method of selecting and reducing the number of primitives considered in the integration has the potential of greatly speeding up the determination of atomic properties if the appropriate changes are made to the integration program. Preliminary results support this conclusion [24].

SELECTION OF SUBSETS OF LOCALIZED MOLECULAR ORBITALS

Theoretical principles

The self-consistent molecular orbitals are the result of linear combination of atomic orbitals (LCAO) and this LCAO procedure yields a set of delocalized orthogonal MOs. It is important to note that these MOs are delocalized, that is they involve contributions from many or all atoms of the molecule. In Fig. 7, a contour plot of a delocalized σ-type orbital of butadiine is shown. It is obvious that this orbital contributes electron density to each atom; therefore this orbital needs to be considered for the determination of the atomic properties of each atom in the molecule. The same situation occurs with all other delocalized MOs.

With a suitable unitary transformation of the delocalized MOs it is possible to obtain localized MOs (LMOs) without any overall change of the electron density distribution of the molecule. In Fig. 8, a contour plot is shown of one

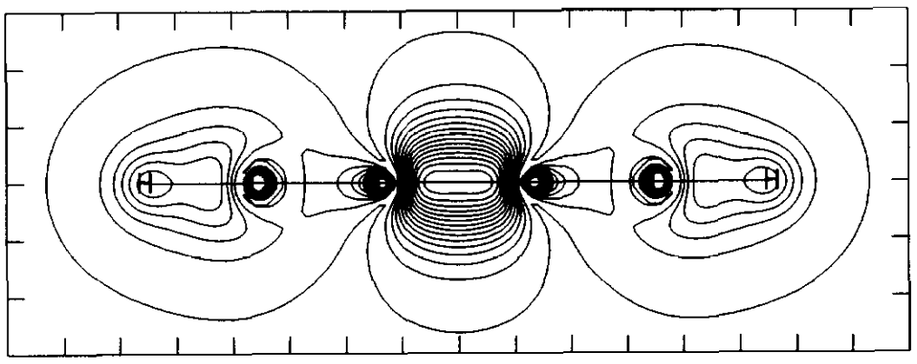


Fig. 7. Contour plot of a delocalized molecular orbital of butadiine.

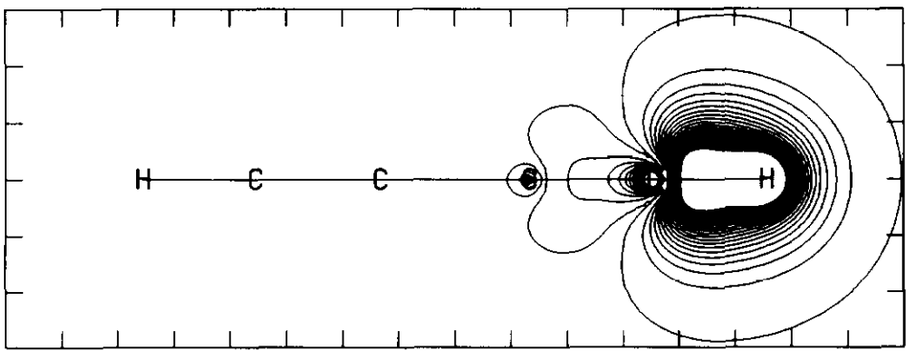


Fig. 8. Contour plot of a localized molecular orbital of butadiine.

of the resulting LMOs. It is clear simply by inspection that this LMO contributes electron density to only a few atoms in the molecule.

Thus: With an algorithm for the selection of the subset of LMOs that suffices to fully describe the electron density distribution in the proximity of a given atom, the determination of atomic properties could be made highly efficient.

Suppose the localization was performed. The molecular electron density is given by

$$\rho_{\mathrm{mol}}(x,y,z) = 2\sum_i [\sum_\nu c_{\nu i}\phi_\nu]^2$$

where the coefficients $c_{\nu i}$ are now those of the LMOs. We are interested in finding an electron density function ρ''_{M} that describes the electron density distribution in the basin of atom M in a simpler fashion than the function ρ_{mol} does. The molecular orbitals are now localized, so it is clear that there will be many LMOs that contribute essentially nothing to the molecular electron density function ρ_{mol} in the basin of atom M. If these LMOs could be identified, they could be omitted in the determination of the integrated properties of atom M. To accomplish this task, we first determine the atom-specific LMOs. The set of atom-specific LMOs of atom M contains all of the LMOs which do have significant contributions from the basis functions centered at atom M. These sets have been determined with a selection algorithm that examines the coefficients of the LMOs as described below. The determination of the sets of atom-specific LMOs is made for all of the atoms in the molecule and it results in N lists of LMO identifiers. To perform the integration within the basin of atom M we need to use an electron density function that accurately describes the electron density in that basin, that is an electron density function expressed in terms of all of those LMOs that contribute to ρ_{mol} in the basin of M. We denote these LMOs required for the property evaluation of atom M as the subset of LMOs of atom M, SSLMO(M). The smallest conceivable subset SSLMO(M) must contain all of the LMOs that occur in either of the lists of atom-specific LMOs of atom M and its next neighbors. Increased accuracy can be achieved if the SSLMO(M) also contains LMOs that have been identified as important for higher-order neighbors. With the definition of *NNN* given above, we thus define an atom-specific electron density function

$$\rho''_{\mathrm{M}}(x,y,z,NNN) = 2\sum_i \Psi_i^{\mathrm{loc}}\Psi_i^{\mathrm{loc}} \qquad (\text{where } i \in \mathrm{SSLMO(M},NNN)$$

that may be used to determine the integrated atomic properties of atom M. Note that the subset SSLMO(M,0) contains only the atom-specific LMOs of atom M, and $\rho''_{\mathrm{M}}(x,y,z,0)$ is the associated electron density function.

The evaluation of the atomic properties of M via the function $\rho''_{\mathrm{M}}(x,y,z,NNN)$ appears rather attractive because the number of LMOs that need to be consid-

ered is not only reduced but, very importantly, there will be an upper limit to the number of LMOs in the subsets SSLMO(M), no matter what the size of the molecular system.

Graphical illustration

The principle of this method can be conveniently shown graphically. We begin with an inspection of the atom-specific LMOs. In Fig. 9, the resulting electron density functions $\rho''_M(0,0,z,0)$ are shown for $C_{10}H_2$. The function $\rho''_{C5}(0,0,z,0)$ is emphasized in Fig. 9 and its shape is fairly typical. This function has a large maximum at the position of the C5 nucleus ($>70\ e$ a.u.$^{-3}$), it extends over the entire bonding region to the next neighbors, it exhibits significant maxima at the positions of the nuclei of the next two neighbors, and decreases rapidly in the regions that are further than two bonds away from C5. Note that $\rho''_{C5}(0,0,z,0)$ and $\rho''_{C7}(0,0,z,0)$, for example, essentially coincide in the C5–C7 bonding region. This is of course a consequence of the procedure because the important LMOs for this bond will be contained in the SSLMO lists of both of these atoms. In Fig. 10, the total electron density of $C_{10}H_2$ is shown graphically, together with several electron density functions

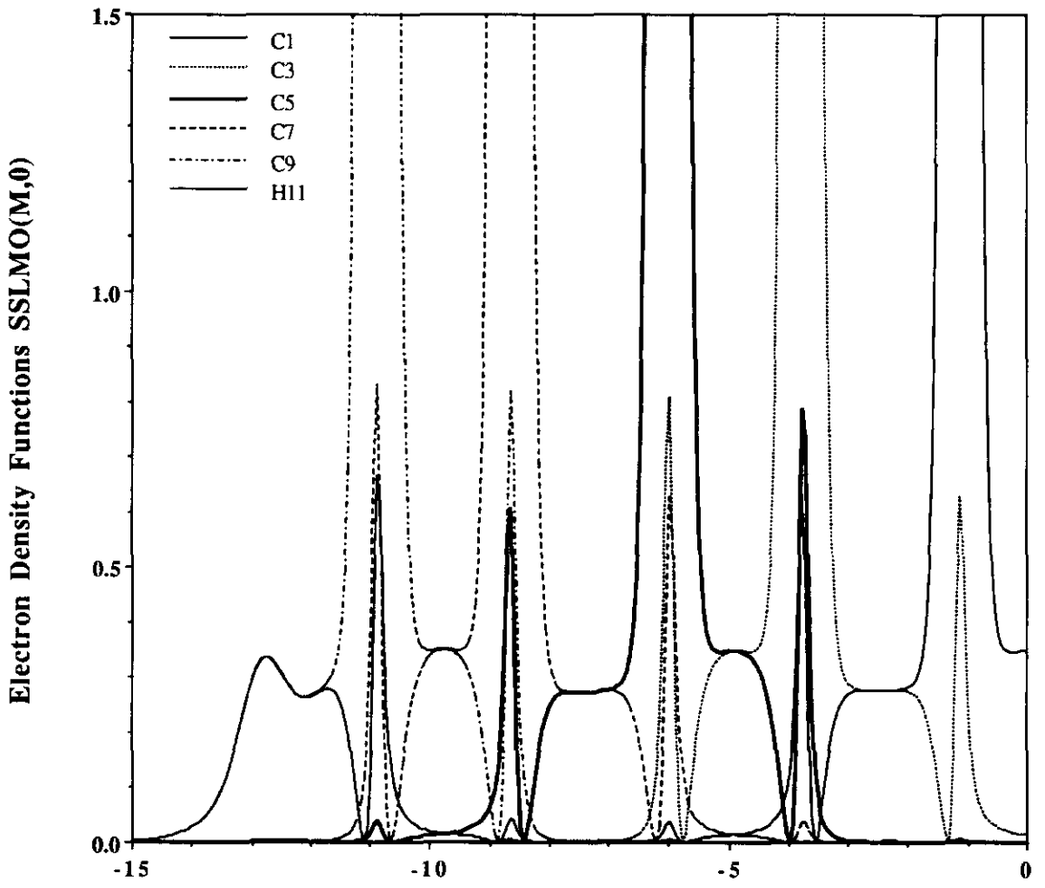


Fig. 9. Electron density functions $\rho''_M(0,0,z,NNN=0)$ associated with the atom-specific localized molecular orbitals of the atoms in $C_{10}H_2$. The inversion center of $C_{10}H_2$ is at the origin, and the molecule lies on the z axis (in Å). Values of ρ are in e a.u.$^{-3}$ (see Fig. 1 for atom numbering).

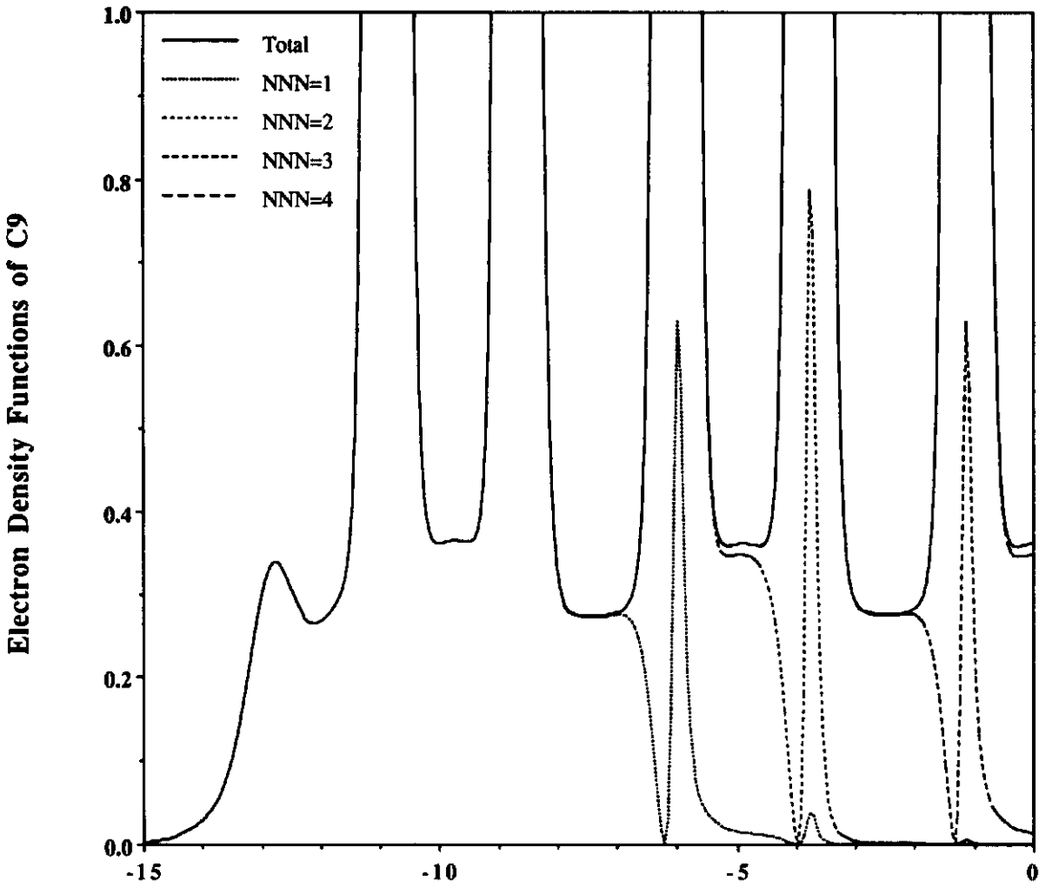


Fig. 10. The electron density functions $\rho''_{C9}(0,0,z,NNN)$ of $C_{10}H_2$ shown as a function of *NNN*. The molecular electron density function $\rho_{mol}(0,0,z)$ is superimposed as a solid line (see Fig. 9 for units and molecular orientation).

$\rho''_{C9}(0,0,z,NNN)$ associated with the subsets of selected localized molecular orbitals of atom C9, are illustrated as a function of *NNN*. Figure 10 shows in a compelling fashion that the electron density distribution in the basin of the C9 atom can indeed be excellently described by electron density functions of the type ρ''_M. Whereas there are no significant deviations apparent in the C9 basin in this Fig. 10, there might still be differences in other regions, and this possibility will be examined quantitatively below. Probably the best single graphical representation of the essentials of the method is illustrated in Fig. 11. On top in Fig. 11, a 3-dimensional contour plot is shown of the molecular electron density function $\rho_{mol}(x,y,z)$ of C_4H_2. This plot was generated with the programs PSICON ad PSI2/88 [25], which were easily adapted to contour electron density functions described by a 3-dimensional matrix of $\rho_{mol}(x,y,z)$ values computed with the program NETZ3D [26]. After localization and determination of the subsets of selected localized MOs, we can determine the atom-specific electron density functions $\rho''_M(x,y,z,NNN)$. The resulting electron density function for one of the hydrogens, $\rho''_H(x,y,z,1)$, is shown on the bottom of Fig. 11. Note that the electron density distribution in the basin of the hydrogen and in the CH bonding region in particular shows no obvious deviation from $\rho_{mol}(x,y,z)$. Quantitative comparisons between $\rho''_M(x,y,z,NNN)$ and $\rho_{mol}(x,y,z,)$ will be made in the next section, for the polyines and the polynitriles.

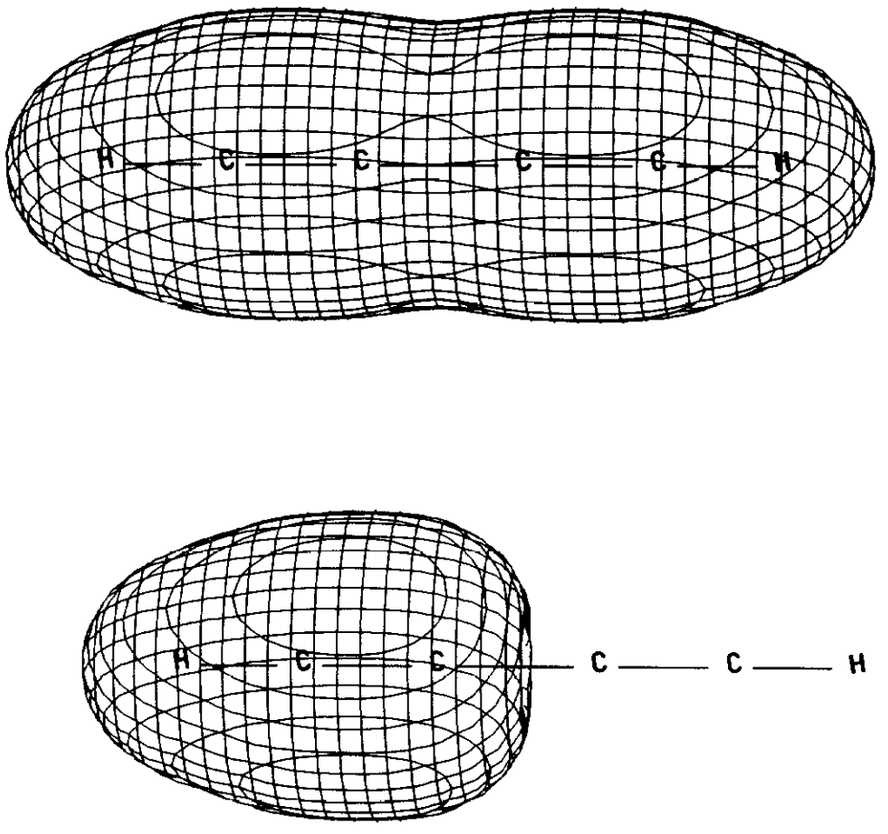
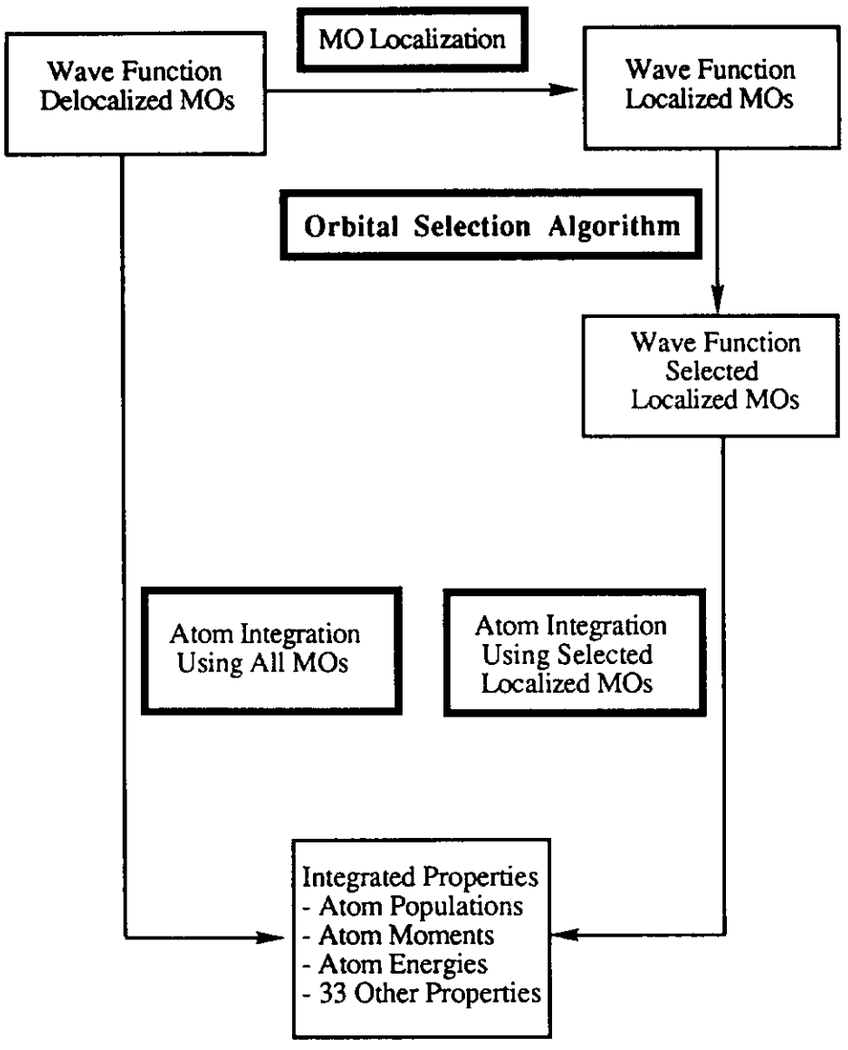


Fig. 11. Top, a three-dimensional contour plot of the electron density function of butadiine. Bottom, the electron density function of one of the hydrogen atoms that results after selection of the localized molecular orbitals with $NNN=1$ is contoured. In both cases, contours are drawn for $\rho=0.01\ e$ a.u.$^{-3}$.

Computational implementation

In Scheme 1, the procedure that has been implemented to determine the subsets SSLMO(M) and to determine the integrated atomic properties of atom M is schematically outlined. In the first step, the molecular orbitals are localized. We have employed the Boys localization [27]. The source codes of the programs BOYLOC and PRPLIB of the quantum-chemistry program GAMESS general atomic and molecular electronic structure system [28] were modified so that both the MO coefficients for the basis functions and for the primitives of the LMOs were written out to the GAMESS DAT file. The program EXTRACTLOCWFN was written to collect pertinent information regarding basis set and molecular geometry and to write it, together with the wave function composed of the LMOs and expanded in the primitives into a so-called WFN file, in a format suitable for the electron density analysis programs. Our program LOCALMOID was then used to inspect the LMOs and to produce lists of the atom-



Scheme 1. Program organization and data flow for wave function analysis.

specific LMOs for each atom. The program LOCALMOID will be described in detail elsewhere [29]. In short, this orbital selection algorithm operates on the wave function expanded in the basis functions. First, for each of the occupied LMOs the largest coefficient is searched for. Secondly, all basis function coefficients of the same magnitude are searched for and this LMO is assigned to the lists of all of those atoms that are associated with basis function coefficients of that magnitude. After all LMOs have been examined in this way and all of the lists of atom-specific LMOs are determined, the program LOCALMOID inquires into the definition of the neighborhood of M via the parameter *NNN* (or a distance criterion). The *N* atom-specific subsets of LMOs SSLMO(M) are then generated as lists of all of the LMOs that occur in either the list of the atom-specific LMOs of the atom M or any of its neighbors, with the neighborhood defined by *NNN*. These lists are written out to a so-called MID file. The program SELECT-MO-WFN subsequently uses the subset selection information contained in the MID file, together with the molecular wave function of the

WFN file of the LMOs, to produce one WFN file for each atom that defines the electron density functions $\rho''_M(x,y,z,NNN)$. With the atom-specific electron density functions $\rho''_M(x,y,z,NNN)$ determined in this way, the topological analyses and the numerical integrations are performed with EXTREME and PROAIM.

Performance of the method

Topological properties

In Table 5, the topological properties for all of the polyines, as determined with the entire wave function and with the subsets of selected LMOs using a minimal neighborhood, are given. Except for acetylene, three sets of topological properties are given for each bond. The first entry gives the values determined with ρ_{mol} and the other entries specify bond properties that were determined with the ρ''_M functions of the two bonded atoms, respectively, as indicated in column 2. The symmetry of the polyines intrinsically requires that the curvatures λ_1 and λ_2 determined with the total electron density function are identical. The subset selection may result in ρ''_M functions that are no longer perfectly symmetric in this regard. For all of those cases where λ_1 and λ_2 differ, both values are given in Table 5 and it is found that such deviations are essentially negligible. Non-nuclear attractors occur in all triple bonds at this level and the topological parameters show the electron density to be essentially flat in the central third of each triple bond. Despite these very small gradients of the electron density in the bonding regions, the agreement between the topological properties determined with the entire wave functions and those determined with the subsets of LMOs is excellent.

Integrated properties

Results obtained both with and without the new method are shown in Table 6. In Fig. 12, the integrated atomic dipoles determined both with the total electron density functions and with the electron density functions resulting after subset selection with $NNN=1$, are plotted versus each other. The agreement is excellent. Similar agreement was also found for the integrated populations and for the kinetic energies of the atoms. Linear regression resulted in intercepts of zero (± 0.004) and slopes of unity (± 0.01). The average error for the integrated population is thus less than 1% and that for the integrated energies is less than 0.1% of an atomic unit (0.6275 kcal mol^{-1}). The absolute errors are found to vary to some extent, as illustrated in Fig. 13. In Fig. 13 the absolute errors of the atom populations and the kinetic energies are shown. As can be seen, the integrated properties of the hydrogen are reproduced in an excellent fashion, even if only next neighbors are considered for the ρ''_H function. For the heavy atoms, the populations may deviate by up to 0.06 kcal mol^{-1}

TABLE 5

Topological properties for polyines[a]

A-B Bond	Atom[b]	r_A[b]	r_B	ρ	λ_1	λ_2	λ_3
C_2H_2	$H\text{-}C_1{\equiv}C_2\text{-}H$						
C-C	All	0.4486	0.7198	0.3651	−0.2853		0.4359
BCP	C	0.4486	0.7198	0.3651	−0.2853		0.4359
C-C	All	0.5842	0.5842	0.3678	−0.2826		−0.1415
NNA	C	0.5842	0.5842	0.3678	−0.2826		−0.1415
C-H	All	0.4279	0.6374	0.2643	−0.5806		0.4549
C_4H_2	$H\text{-}C_3{\equiv}C_1\text{-}C_2{\equiv}C_4\text{-}H$						
C1-C2	All	0.7041	0.7041	0.2714	−0.4088		0.0531
	C1	0.7019	0.7064	0.2713	−0.4083		0.5284
C3-C1	All	0.4462	0.7285	0.3607	−0.2921		0.4963
BCP at C3	C1	0.4462	0.7285	0.3607	−0.2921		0.4964
	C3	0.4461	0.7285	0.3606	−0.2986	−0.2942	0.4975
C3-C1	All	0.6049	0.5698	0.3646	−0.2874		−0.1391
NNA	C1	0.6049	0.5698	0.3646	−0.2873		−0.1391
	C3	0.6050	0.5697	0.3646	−0.2879		−0.1393
C3-C1	All	0.7239	0.4508	0.3626	−0.2836		0.3847
BCP at C1	C1	0.7239	0.4507	0.3626	−0.2835		0.3850
	C3	0.7240	0.4507	0.3626	−0.2833		0.3852
H5-C3	All	0.4236	0.6424	0.2641	−0.5861		0.4600
	C3	0.4236	0.6424	0.2641	−0.5889		0.4600
	H	0.3938	0.6722	0.0567	−0.0746		0.0106
C_6H_2	$H\text{-}C_5{\equiv}C_3\text{-}C_1{\equiv}C_2\text{-}C_4{\equiv}C_6\text{-}H$						
C1-C2	All	0.4482	0.7334	0.3582	−0.2913		0.4454
BCP at C1	C1	0.4480	0.7336	0.3581	−0.2967		0.4475
C1-C2	All	0.5908	0.5908	0.3611	−0.2946		−0.1368
NNA	C	0.5911	0.5905	0.3611	−0.2950		−0.1372
C3-C1	All	0.6730	0.7304	0.2731	−0.4081		0.0510
	C1	0.6751	0.7283	0.2730	−0.4096		0.0506
	C3	0.6709	0.7325	0.2729	−0.4095		0.0511
C5-C3	All	0.4465	0.7291	0.3603	−0.2955		0.4936
BCP at C5	C3	0.4465	0.7291	0.3603	−0.2962		0.4937
	C5	0.4464	0.7292	0.3602	−0.3035		0.4947
C5-C3	All	0.6056	0.5700	0.3642	−0.2890		−0.1383
NNA	C3	0.6056	0.5699	0.3642	−0.2891		−0.1383
	C5	0.6058	0.5699	0.3642	−0.2899		−0.1385
C5-C3	All	0.7248	0.4508	0.3622	−0.2839		0.3838
BCP at C3	C3	0.7248	0.4508	0.3622	−0.2838		0.3840
	C5	0.7249	0.4508	0.3622	−0.2836		0.3842
H7-C5	All	0.4221	0.6444	0.2640	−0.5885		0.4621
	C5	0.4221	0.6444	0.2640	−0.5917		0.4621
	H7	0.4219	0.6446	0.2639	−0.5916		0.4618

TABLE 5 (continued)

A-B Bond	Atom[b]	r_A[b]	r_B	ρ	λ_1	λ_2	λ_3
C_8H_2	H-C_7≡C_5-C_3≡C_1-						
C1-C2	All	0.6989	0.6989	0.2751	−0.4089		0.0465
	C1	0.6968	0.7011	0.2750	−0.4131		0.0463
C3-C1	All	0.4486	0.7341	0.3577	−0.2947		0.4411
BCP at C3	C1	0.4486	0.7341	0.3577	−0.2954		0.4413
	C3	0.4485	0.7342	0.3577	−0.3015	−0.2973	0.4424
C3-C1	All	0.5913	0.5915	0.3607	−0.2966		−0.1357
NNA	C1	0.5909	0.5918	0.3606	−0.2969		−0.1360
	C3	0.5914	0.5913	0.3606	−0.2973		−0.1360
C3-C1	All	0.7345	0.4482	0.3577	−0.2916		0.4498
BCP at C1	C1	0.7347	0.4481	0.3576	−0.2967		0.4471
	C3	0.7346	0.4482	0.3577	−0.2915	−0.2914	0.4455
C5-C3	All	0.6631	0.7396	0.2733	−0.4076		0.0514
	C3	0.6652	0.7376	0.2732	−0.4094		0.0508
	C5	0.6611	0.7417	0.2731	−0.4092		0.0515
C7-C5	All	0.4467	0.7291	0.3602	−0.2969		0.4914
BCP at C7	C5	0.4467	0.7292	0.3602	−0.2978	−0.2973	0.4915
	C7	0.4466	0.7292	0.3601	−0.3053	−0.3002	0.4925
C7-C5	All	0.6055	0.5703	0.3641	−0.2895		−0.1381
NNA	C5	0.6055	0.5703	0.3641	−0.2896		−0.1382
	C7	0.6056	0.5702	0.3641	−0.2905		−0.1384
C7-C5	All	0.7251	0.4507	0.3621	−0.2836		0.3851
BCP at C5	C5	0.7251	0.4507	0.3621	−0.2836		0.3854
	C7	0.7252	0.4506	0.3621	−0.2833		0.3856
H9-C7	All	0.4214	0.6453	0.2640	−0.5898		0.4631
	C7	0.4213	0.6453	0.2640	−0.5931		0.4631
	H9	0.4211	0.6455	0.2639	−0.5930		0.4628
$C_{10}H_2$	H-C_9≡C_7-C_5≡C_3-C_1≡						
C1-C2	All	0.4487	0.7352	0.3573	−0.2951		0.4404
BCP	C1	0.4486	0.7353	0.3572	−0.3016	−0.2975	0.4418
C1-C2	All	0.5919	0.5919	0.3602	−0.2987		−0.1344
NNA	C	0.5921	0.5918	0.3602	−0.2993	−0.2989	−0.1347
C3-C1	All	0.6880	0.7088	0.2755	−0.4091		0.0460
	C1	0.6902	0.7066	0.2754	−0.4139	−0.4122	0.0457
	C3	0.6858	0.7110	0.2754	−0.4138	−0.4122	0.0459
C5-C3	All	0.4489	0.7341	0.3577	−0.2961		0.4385
BCP at C5	C3	0.4488	0.7341	0.3577	−0.2971	−0.2966	0.4388
C5-C3	All	0.5910	0.5919	0.3606	−0.2971		−0.1355
NNA	C3	0.5907	0.5923	0.3606	−0.2975		−0.1358
C5-C3	All	0.7348	0.4481	0.3576	−0.2914		0.4466
BCP at C3	C3	0.4480	0.7349	0.3575	−0.2964		0.4487
H11-C9	All	0.4210	0.6456	0.2640	−0.5906		0.4636
	H11	0.4208	0.6458	0.2639	−0.5938	−0.5918	0.4632

TABLE 5 (continued)

A-B Bond	Atom[b]	r_A[b]	r_B	ρ	λ_1	λ_2	λ_3
$C_{20}H_2$	$H\text{-}C_{19}\equiv C_{17}\text{-}C_{15}\equiv C_{13}\text{-}C_{11}\equiv C_9\text{-}C_7\equiv C_5\text{-}C_3\equiv C_1\text{-}$						
C1-C2	All	0.6978	0.6978	0.2760	−0.4098		0.0451
	C1	0.6978	0.6978	0.2760	−0.4102		0.0451
C1-C3	All	0.7358	0.4488	0.3569	−0.2960		0.4397
BCP at C1	C1	0.7358	0.4488	0.3569	−0.2968	−0.2963	0.4397
C1-C3	All	0.5922	0.5925	0.3599	−0.2998		−0.1337
NNA	C3	0.5922	0.5925	0.3599	−0.3000		−0.1337
	C1	0.5922	0.5925	0.3599	−0.3000		−0.1337
C1-C3	All	0.4490	0.7357	0.3570	−0.2965		0.4381
BCP at C3	C3	0.4490	0.7357	0.3570	−0.2974	−0.2969	0.4381
C5-C3	All	0.6930	0.7026	0.2760	−0.4097		0.0452
	C3	0.6930	0.7026	0.2760	−0.4103	−0.4099	0.0452
H-C19	All	0.4205	0.6464	0.2639	−0.5913		0.4643
	H	0.4205	0.6464	0.2639	−0.5917	−0.5915	0.4643

[a]For an explanation of headings see footnotes to Table 3.
[b]For each bond A–B, the characteristics of bond critical points and the non-nuclear attractor are given, as determined with the entire wave function (all) and with the subset of the MOs determined for the specified atom.

and atom kinetic energies by up to 45 kcal mol^{-1}. Such deviations in the populations may be acceptable for many types of qualitative applications. Note that these deviations are an order of magnitude smaller than the typical differences between integrated populations and basis set partitioning populations. We show below that a much better and virtually exact agreement can be obtained if the parameter *NNN* is increased.

$$IP_{sel} = 0.991 IP_{all} + 0.0017 \qquad (R^2 = 1.000)$$

$$\mu_{sel} = 1.002 \mu_{all} + 0.0013 \qquad (R^2 = 1.000)$$

$$T_{sel} = 0.999 T_{all} + 0.0036 \qquad (R^2 = 1.000)$$

Time requirements

In Fig. 14, the time requirements are plotted as a function of the number of triple bonds (size parameter), both without (solid) and with (unfilled marks) the localized orbital selection algorithm, for the C1, C3, and H atoms of the polyines. Consider, for example, the time requirements for the C1 atoms. Whereas there is little benefit in using a subset of selected LMOs for the small butadiine molecule, the advantage of the new technique becomes apparent as n increases. In particular, it is instantaneously obvious that the time requirement does not increase linearly with the size of the system. It is this feature that makes this method particularly powerful for the rigorous electron density analysis of rather large molecules. The advantage of the subset selection method

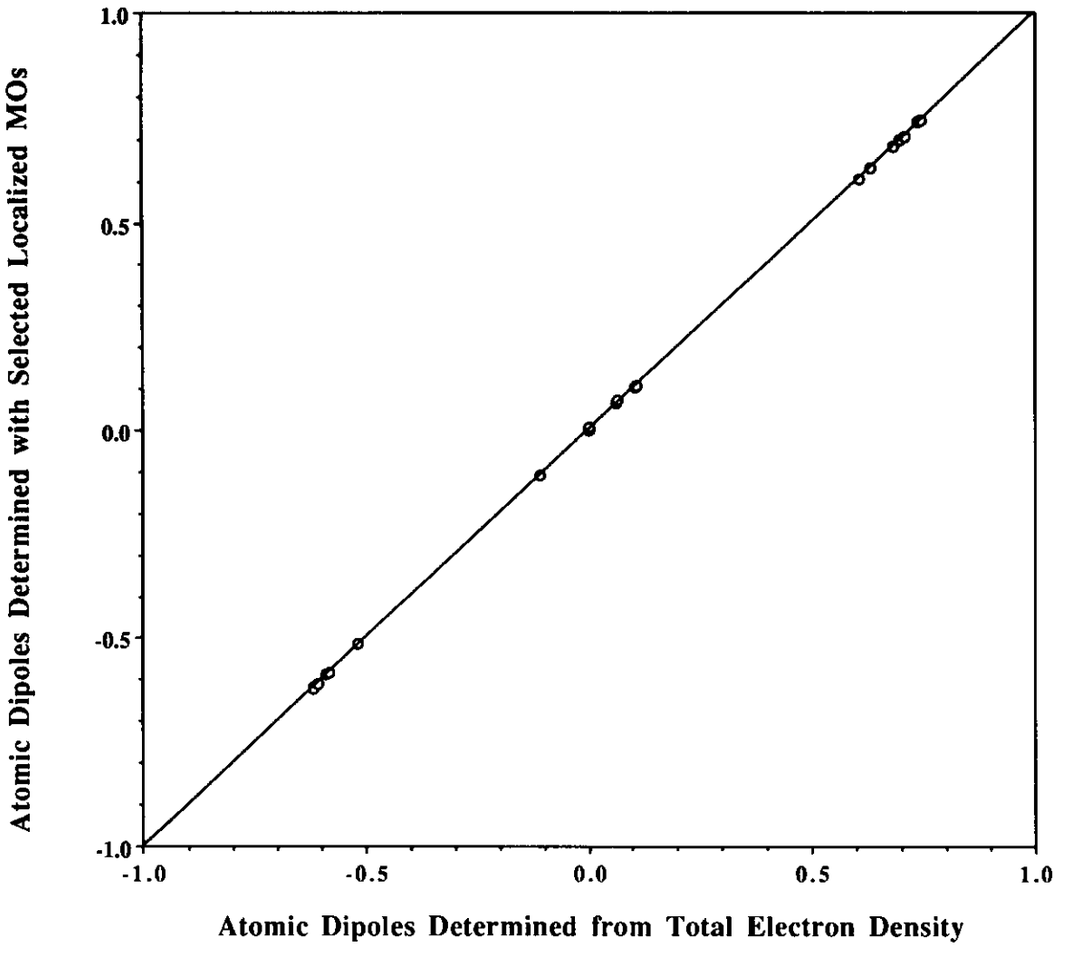


Fig. 12. Atomic dipole moment determined by integration of the electron density function resulting both from all and from subsets of localized molecular orbitals ($NNN=1$), correlated in a linear fashion with a slope of unity and an intercept of zero.

increases as the size of the system increases. Time savings are shown as a percentage in Fig. 15. Time savings amount to up to 80 percent for the largest system in this series.

Polynitriles

The analysis of the polynitriles gave results that lead to the same conclusions as the data discussed for the polyines. As an example, we present the results of the topological analysis of $C_3N_3H_5$ in Table 7, and the corresponding integrated properties and timing information are given in Table 8.

Neighborhood selection dependence

The deviations in the topological and integrated properties found for the polyines are related to the characteristics of the LMOs. LMOs may contain "tails", that is small local maxima may occur outside the region in which the MO is concentrated. For example, the LMO shown in Fig. 8 largely makes use of primitives centered on the terminal C3 carbon and H atoms, but it also

TABLE 6

Integrated properties of polyines[a-d]

Atom	IP[a]		μ_z[b]		T[c]		t[d]	
	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs
C_2H_2, E(RHF/STO-3G) = −75.856248 a.u. with −V/T = 2.003519923								
C1	5.415	5.414	0.696	0.697	36.922220	36.922085	0.82	0.82
H3	0.939	0.938	0.104	0.104	0.578044	0.576929	0.80	0.72
NNA 1-2	1.291	1.291	0.000	0.000	0.859158	0.859158	0.87	0.86
Σ	13.999	13.995			75.859684	75.857185	4.11	3.94
C_4H_2, E(RHF/STO-3G) = −150.595770 a.u. with −V/T = 2.00537723								
C1	5.413	5.410	−0.590	−0.588	36.927181	36.924579	1.62	1.46
C3	5.386	5.351	0.734	0.741	36.953679	36.909417	1.35	0.99
H5	0.923	0.923	0.107	0.106	0.568433	0.567712	1.38	0.92
NNA 1-3	1.274	1.291	0.062	0.067	0.850324	0.856935	1.35	1.00
Σ	12.995	12.975			149.462367	150.517287	11.40	8.74
C_6H_2, E(RHF/STO-3G) = −225.336858 a.u. with −V/T = 2.00606962								
C1	5.430	5.374	0.682	0.683	36.945853	36.879868	2.37	1.83
C3	5.366	5.363	−0.517	−0.514	36.921214	36.917475	2.22	1.71
C5	5.380	5.337	0.741	0.747	36.970037	36.916435	1.91	1.06
H7	0.918	0.918	0.108	0.107	0.565296	0.565296	1.77	1.10
NNA 1-2	1.259	1.285	0.000	0.006	0.839989	0.850118	2.13	1.44
NNA 5-3	1.271	1.290	0.065	0.072	0.848371	0.855819	1.94	1.08
Σ	37.989	37.849			225.341533	225.119906	22.55	15.00

C_8H_2, E(RHF/STO-3G)= −300.078097 a.u. with −V/T=2.00643103								
C1	5.384	5.329	−0.606	−0.610	36.928967	36.862231	3.28	1.75
C3	5.439	5.396	0.706	0.706	36.960321	36.931893	3.14	1.72
H9	0.915	0.915	0.108	0.108	0.563885	0.563154	2.30	1.22
NNA 1-3	1.254	1.274	0.002	0.007	0.837184	0.844959	2.94	1.68
$C_{10}H_2$, E(RHF/STO-3G)= −374.819358 a.u. with −V/T=2.00664975								
C1	5.396	5.349	0.632	0.634	36.939920	36.883866	4.43	1.98
C3	5.368	5.313	−0.581	−0.584	36.926998	36.860728	4.41	1.98
H11	0.914	0.914	0.108	0.108	0.563287	0.562570	2.77	1.37
NNA 1-2	1.249	1.272	0.000	0.004	0.834010	0.842764	3.78	1.79
$C_{20}H_2$, E(RHF/STO-3G)= −748.525436 a.u. with −V/T=2.00710254								
C1	5.376	5.327	0.606	0.608	36.940896	36.882364	10.49	2.26
C3	5.382	5.334	−0.618	−0.620	36.944164	36.885494	10.61	2.27
H	0.912	0.912	−0.109	−0.108	0.562288	0.561488	7.81	3.16
C1 (3)	5.376	5.371	0.606	0.607	36.940896	36.935152	10.49	3.77
C3 (3)	5.382	5.378	−0.618	−0.618	36.944164	36.938372	10.61	3.80
H (3)	0.912	0.912	−0.109	−0.109	0.562288	0.562307	7.81	1.38

[a]For an explanation of column headings see footnotes to Table 4.

[b]Integrated populations determined with all delocalized MOs or with selected (sel.) localized orbitals (in *e*).

[c]The integrated kinetic energies *T* are corrected for the virial defects.

[d]Carbon integrations on an SGI 4D/25 and H integrations on a VaxStation3100.

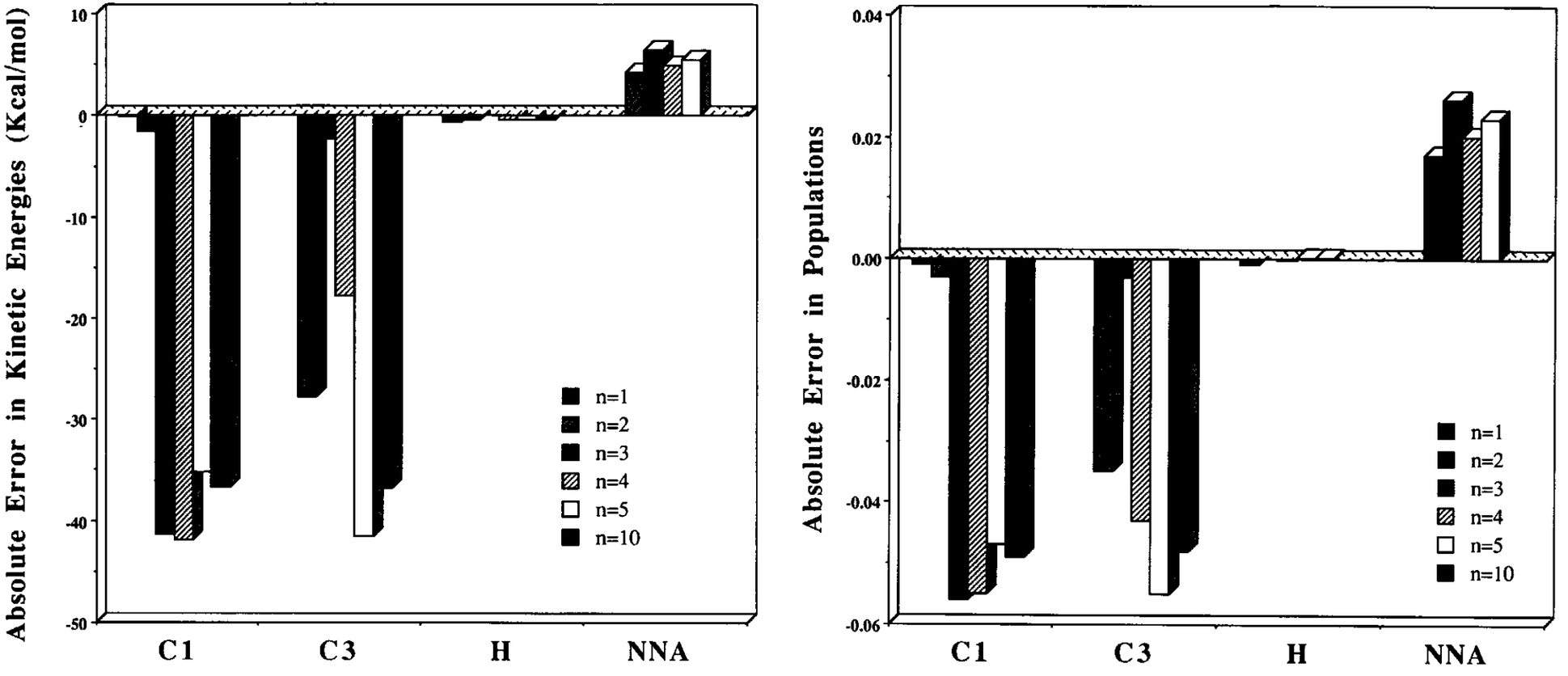


Fig. 13. Absolute deviations of the kinetic energy (kcal mol^{-1}) and of the population (right) of the atoms C1, C3, H and of the non-nuclear attractor (in the C1-C2 or C1-C3 bond) for the polyines $C_{2n}H_2$, as determined with the ρ_{mol} and the ρ''_M electron density functions.

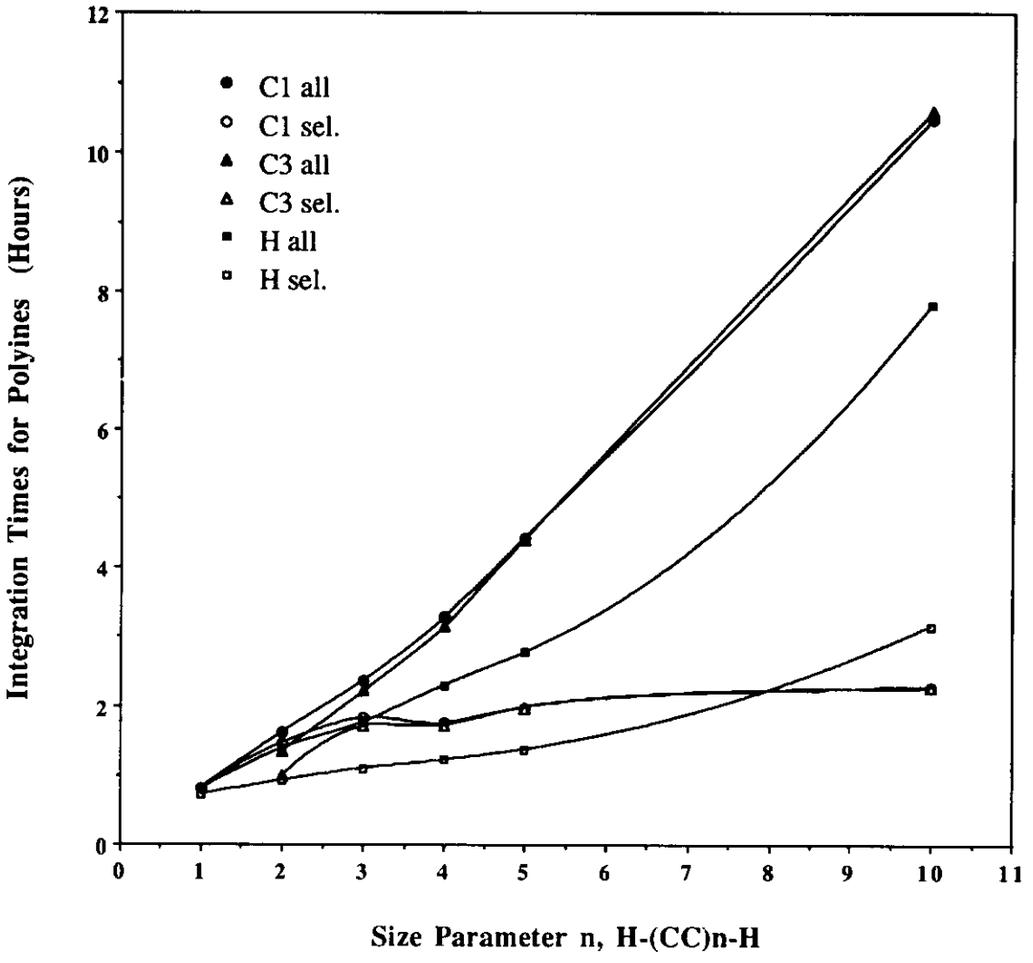


Fig. 14. Integration time plotted vs. the size parameter n for the polyines H-$(CC)_n$-H. Integration time is given in h required on the SGI 4D/25S for the heavy atoms; on the VaxStation3100 for hydrogen atoms. Solid marks refer to integrations of the entire wave functions and unfilled marks refer to integrations of electron density functions derived from subsets of selected localized molecular orbitals with $NNN=1$.

makes use of primitives centered at the C1 atom. The orbital selection algorithm will assign this LMO to the set of C3 but not to the set of C1. If *NNN* is chosen as 1, then this MO will not be considered in the integration of C2, although the small "tail" of this MO in the vicinity of C1 might affect the C1–C2 bonding region. These "tails" are, of course, also the reason for the small local maxima of the function shown in Fig. 9. Such tails in LMOs occur frequently and their locations cannot be predicted. However, graphical inspection of a large number of LMOs suggests that such "tails" extend no further than into the region of the second-order neighbors. It should thus be possible to eliminate this error by inclusion of the atom-specific subsets of LMOs of higher-order neighbors in the computation of the reduced electron density function of the atom that is being integrated.

The effects of the neighborhood selection on the accuracy of the integration results has been studied for atoms C3 and H of C_8H_2 and for atoms C3 and N10 of $C_5N_5H_7$. For each of these atoms, the subsets of LMOs used for their integration were determined by considering just next neighbors ($NNN=1$), all

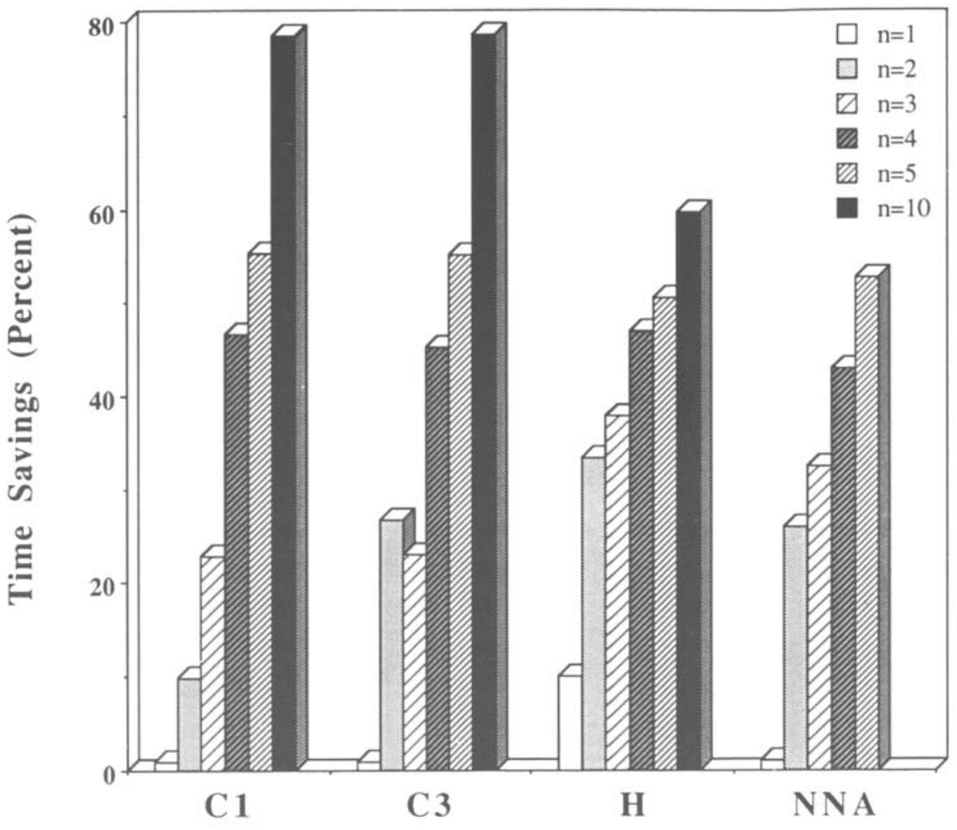
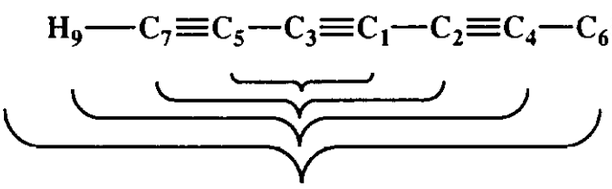


Fig. 15. Time saving for the determination of the integrated atomic properties of atoms C1, C3, H and the non-nuclear attractor for the polyines $C_{2n}H_2$ ($n=1$-5 and 10), based on ρ''_M electron density functions as compared to the total electron density.

next neighbors and their next neighbors ($NNN=2$), and so on, up to $NNN=4$. For C3 of C_8H_2, the four sets of selected localized molecular orbitals used for the integration of C3 contain the atom-specific LMOs of the atoms shown.



Inspection of the integrated properties listed in Table 9 shows that consideration of the second-next neighbors results in the most significant improvement in the integrated charges and dipole moments. Further increase of *NNN* has only an insignificant effect on these two properties. The dependence of the atomic kinetic energy on the neighborhood selection is illustrated in Fig. 16. In Fig. 16, the deviation of the integrated kinetic energy determined with subsets of selected LMOs and the corresponding numerically exact value is plotted versus the number of next neighbors considered. As with the charge and the dipole moment, the atomic kinetic energy is dramatically improved by consideration of the second-next neighbors, although the error may still be as large as 3–5 kcal mol^{-1}. This error is essentially eliminated entirely if *NNN* is increased to 4. Whereas charge and dipole moment can be determined to high accuracy by using $NNN=2$ (or even $NNN=1$), fragment energy analysis ap-

TABLE 7

Topological properties of $C_3N_3H_5$[a] and $C_5N_5H_7$[b c]

A-B Bond	Atom	r_A	r_B	ρ	λ_1	λ_2	λ_3
$C_3N_3H_5$							
C1-N2	All	0.4216	0.8574	0.3273	−0.6424	−0.4056	1.3347
	C1	0.4214	0.8576	0.3267	−0.6414	−0.4154	1.3430
	N2	0.4216	0.8574	0.3272	−0.6423	−0.4064	1.3350
N2-C3	All	0.9164	0.5442	0.2572	−0.4474	−0.3923	0.3143
	N2	0.9165	0.5441	0.2572	−0.4478	−0.3970	0.3141
	C3	0.9165	0.5442	0.2572	−0.4478	−0.3972	0.3142
C3-N4	All	0.4234	0.8573	0.3286	−0.6864	−0.4017	1.2811
	C3	0.4232	0.8575	0.3281	−0.6857	−0.4088	1.2892
	N4	0.4233	0.8573	0.3285	−0.6912	−0.4053	1.2812
N4-C5	All	0.9253	0.5360	0.2542	−0.4356	−0.3789	0.3268
	N4	0.9257	0.5355	0.2541	−0.4353	−0.3793	0.3272
	C5	0.9286	0.5327	0.2532	−0.4317	−0.3752	0.3301
C5-N6	All	0.4232	0.8537	0.3337	−0.7147	−0.4057	1.2699
	C5	0.4231	0.8537	0.3337	−0.7157	−0.4055	1.2699
	N6	0.4231	0.8537	0.3336	−0.7195	−0.4076	1.2699
H8-C1	All	0.4423	0.6460	0.2631	−0.5930	−0.5726	0.4857
	H8	0.4422	0.6461	0.2631	−0.5932	−0.5756	0.4854
	C1	0.4422	0.6461	0.2631	−0.5932	−0.5757	0.4855
H7-C1	All	0.4496	0.6423	0.2605	−0.5711	−0.5562	0.4819
	H7	0.4502	0.6417	0.2603	−0.5697	−0.5580	0.4795
	C1	0.4497	0.6423	0.2604	−0.5711	−0.5592	0.4817
H9-C3	All	0.4492	0.6485	0.2585	−0.5721	−0.5467	0.4935
	H9	0.4503	0.6472	0.2581	−0.5697	−0.5489	0.4884
	C3	0.4492	0.6484	0.2585	−0.5720	−0.5490	0.4933
H10-C5	All	0.4503	0.6462	0.2586	−0.5712	−0.5447	0.4871
	H10	0.4509	0.6455	0.2584	−0.5701	−0.5453	0.4843
	C5	0.4504	0.6461	0.2586	−0.5712	−0.5446	0.4867
N6-H11	All	0.7076	0.3402	0.3082	−0.9508	−0.9212	0.7951
	N6	0.7078	0.3401	0.3082	−0.9534	−0.9215	0.7940
	H11	0.7090	0.3387	0.3078	−0.9544	−0.9229	0.7871
$C_5N_5H_7$							
C1-N2	Al	0.4218	0.8577	0.3270	−0.6408	−0.4083	1.3301
	C1	0.4218	0.8577	0.3270	−0.6408	−0.4095	1.3304
	N2	0.4218	0.8577	0.3270	−0.6408	−0.4084	1.3302
N2-C3	All	0.9144	0.5449	0.2583	−0.4506	−0.3969	0.3128
	N2	0.9144	0.5449	0.2583	−0.4506	−0.3975	0.3128
	C3	0.9144	0.5449	0.2583	−0.4506	−0.3975	0.3128
C3-N4	All	0.4241	0.8586	0.3276	−0.6818	−0.4054	1.2678
	C3	0.4241	0.8586	0.3276	−0.6818	−0.4065	1.2680
	N4	0.4241	0.8586	0.3276	−0.6818	−0.4055	1.2678
N4-C5	All	0.9214	0.5331	0.2578	−0.4458	−0.3846	0.3260
	N4	0.9214	0.5331	0.2578	−0.4459	−0.3851	0.3260
	C5	0.9214	0.5331	0.2578	−0.4459	−0.3851	0.3261
C5-N6	All	0.4243	0.8588	0.3275	−0.6836	−0.3995	1.2596
	C5	0.4243	0.8588	0.3275	−0.6836	−0.4003	1.2598
	N6	0.4243	0.8588	0.3275	−0.6838	−0.3997	1.2596

TABLE 7 (continued)

A-B Bond	Atom	r_A	r_B	ρ	λ_1	λ_2	λ_3
N6-C7	All	0.9239	0.5316	0.2568	−0.4426	−0.3794	0.3295
	N6	0.9239	0.5316	0.2567	−0.4426	−0.3794	0.3295
	C7	0.9239	0.5315	0.2567	−0.4425	−0.3794	0.3296
C7-N8	All	0.4238	0.8577	0.3284	−0.6875	−0.3962	1.2696
	C7	0.4238	0.8577	0.3284	−0.6875	−0.3962	1.2696
	N8	0.4238	0.8577	0.3284	−0.6877	−0.3964	1.2696
N8-C9	All	0.9270	0.5342	0.2538	−0.4340	−0.3758	0.3297
	N8	0.9270	0.5341	0.2538	−0.4340	−0.3759	0.3297
	C9	0.9271	0.5340	0.2538	−0.4339	−0.3758	0.3298
C9-N10	All	0.4231	0.8536	0.3337	−0.7156	−0.4042	1.2727
	C9	0.4230	0.8536	0.3337	−0.7156	−0.4042	1.2727
	N10	0.4230	0.8536	0.3337	−0.7158	−0.4043	1.2727
H11-C1	All	0.4492	0.6428	0.2605	−0.5716	−0.5575	0.4827
	H11	0.4492	0.6428	0.2605	−0.5716	−0.5579	0.4827
	C1	0.4492	0.6428	0.2605	−0.5716	−0.5579	0.4827
H12-C1	All	0.4413	0.6470	0.2632	−0.5944	−0.5748	0.4872
	H12	0.4413	0.6471	0.2632	−0.5944	−0.5751	0.4872
	C1	0.4413	0.6471	0.2632	−0.5944	−0.5751	0.4872
H13-C3	All	0.4485	0.6493	0.2585	−0.5727	−0.5487	0.4947
	H13	0.4485	0.6493	0.2585	−0.5727	−0.5490	0.4947
	C3	0.4485	0.6493	0.2585	−0.5727	−0.5490	0.4947
H14-C5	All	0.4487	0.6487	0.2587	−0.5735	−0.5473	0.4941
	H14	0.4487	0.6487	0.2587	−0.5736	−0.5476	0.4941
	C5	0.4487	0.6487	0.2587	−0.5735	−0.5475	0.4941
H15-C7	All	0.4491	0.6481	0.2587	−0.5733	−0.5458	0.4933
	H15	0.4491	0.6481	0.2587	−0.5734	−0.5459	0.4933
	C7	0.4491	0.6481	0.2587	−0.5733	−0.5458	0.4933
H16-C9	All	0.4502	0.6461	0.2587	−0.5716	−0.5444	0.4870
	H16	0.4502	0.6461	0.2587	−0.5716	−0.5445	0.4870
	C9	0.4502	0.6461	0.2587	−0.5716	−0.5444	0.4870
N10-H17	All	0.7084	0.3396	0.3081	−0.9523	−0.9224	0.7942
	N10	0.7084	0.3396	0.3081	−0.9526	−0.9224	0.7941
	H17	0.7084	0.3395	0.3081	−0.9526	−0.9225	0.7940

[a] $NNN=1$ for $C_3N_3H_5$.
[b] $NNN=3$ for $C_5N_5H_7$.
[c] For an explanation of column headings see footnotes to Table 3.

pears considerably more sensitive to the selection of the neighborhood and the choice of $NNN=4$ appears mandatory.

Similar results are obtained for the polynitriles, and the molecules $C_3N_3H_5$ and $C_5N_5H_7$ may serve as an example. In Table 8, the integrated atomic properties that were determined with the electron density functions ρ''_M ($NNN=1$) for $C_3N_3H_5$ and ρ''_M ($NNN=3$) for $C_5N_5H_7$, and also with ρ_{mol} in each case, are shown. Absolute errors in populations and kinetic energies are compared graphically in Fig. 17. Whereas there is considerable deviation when the min-

TABLE 8

Integrated properties of *N*-substituted polyacetylenes: polynitriles[a]

Atom	IP[b]		μ_z		T[c]		t[d]	
	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs
$C_3N_3H_5$, E(RHF/STO-3G)= −276.212573 a.u. with −V/T=2.00959478; NNN=1								
C1	5.159	5.131	0.7610	0.7617	36.452260	36.413355	2.75	1.55
N2	8.166	8.167	0.3287	0.3278	53.874839	53.874097	2.81	2.02
C3	4.823	4.800	0.6783	0.6804	36.234604	36.201772	3.33	2.53
N4	8.170	8.161	0.3215	0.3376	53.878349	53.837887	2.96	2.29
C5	4.806	4.796	0.6797	0.6813	36.215551	36.201322	3.46	2.23
N6	8.003	7.988	0.4774	0.4925	53.769448	53.718654	2.88	1.43
H7	1.036	1.046	0.0978	0.1096	0.658552	0.667192	2.55	1.11
H8	1.010	1.011	0.0989	0.0983	0.636098	0.635839	2.20	0.95
H9	1.036	1.058	0.1010	0.1128	0.664721	0.685464	3.04	1.38
H10	1.038	1.048	0.0994	0.1093	0.664447	0.675895	2.86	1.30
H11	0.757	0.751	0.1640	0.1632	0.537657	0.531462	2.14	0.92
Σ	44.004	43.957			273.586526	273.442939	30.95	17.71
$C_5N_5H_7$, E(RHF/STO-3G)= −459.601827 a.u. with −V/T=2.01000174; NNN=3								
C1	5.159	5.156	0.763	0.763	36.450795	36.447092	4.99	2.35
N2	8.165	8.165	0.330	0.330	53.871187	53.871187	5.06	2.92
C3	4.823	4.820	0.678	0.678	36.231696	36.228133	6.11	3.85
N4	8.170	8.170	0.318	0.318	53.867777	53.867764	5.68	4.34
C5	4.811	4.809	0.680	0.680	36.220408	36.217663	6.41	5.18
N6	8.162	8.161	0.311	0.312	53.859890	53.856524	5.75	4.57
C7	4.811	4.811	0.678	0.678	36.221402	36.220986	6.31	4.69
N8	8.160	8.158	0.313	0.314	53.864441	53.860130	5.50	3.48
C9	4.805	4.805	0.682	0.682	36.214067	36.213659	5.61	3.18
N10	8.000	7.998	0.476	0.477	53.764191	53.759559	4.65	2.21
H11	1.034	1.034	0.098	0.098	0.664014	0.664026	4.12	1.63
H12	1.007	1.007	0.099	0.099	0.640052	0.640053	3.67	1.48
H13	1.034	1.034	0.101	0.101	0.669629	0.669640	4.83	2.74
H14	1.035	1.035	0.101	0.101	0.670548	0.670561	5.15	3.06
H15	1.037	1.037	0.101	0.101	0.671583	0.671588	4.96	2.99
H16	1.038	1.038	0.099	0.099	0.670893	0.670896	4.42	2.04
H17	0.755	0.754	0.164	0.164	0.541363	0.541221	3.85	1.43
Σ	72.006	71.992			455.093936	455.070682	87.07	52.14

[a]For an explanation of column headings see footnotes to Table 4.

[b]Integrated populations determined with all delocalized MOs (IP) or with selected localized orbitals (IPSLO) are (in *e*).

[c]The integrated kinetic energies are not corrected for the virial defects.

[d]All integrations on a SGI 4D/25S.

TABLE 9

Neighborhood selection effects on C_8H_2 and H-$(HCN)_5$-H integrated properties[a]

Atom	*NNN*[b]	*IP*	μ_z	*T*[c]	*t*[d]
C_8H_2[e]					
C3	1	5.396	0.632	36.931893	1.72
	2	5.434	0.707	36.954243	2.45
	3	5.434	0.707	36.954313	2.67
	4	5.439	0.706	36.960321	3.11
	All	5.439	0.706	36.960321	3.14
H	1	0.915	0.108	0.563154	1.22[f]
	2	0.916	0.108	0.564092	0.94
	3	0.915	0.108	0.563905	1.17
	4	0.915	0.108	0.563908	1.25
	All	0.915	0.108	0.563885	2.30
$C_5N_5H_7$[g]					
C3	1	4.794	0.681	36.553362	2.89
	2	4.820	0.678	36.590448	3.65
	3	4.820	0.678	36.590477	3.85
	4	4.823	0.678	36.593768	4.80
	All	4.823	0.678	36.594076	6.11
N10	1	7.984	0.491	54.249227	1.68
	2	7.998	0.477	54.296635	2.01
	3	7.998	0.476	54.297248	2.21
	4	8.000	0.476	54.301285	2.73
	All	8.000	0.476	54.301926	4.65

[a]For an explanation of column headings see footnotes to Table 4.
[b]The number of next neighbors considered in the selection of the localized molecular orbitals.
[c]*T* values are corrected for the virial defect.
[d]Integration times refer to CPU time on an SGI Personal Iris 4D/25S.
[e]*E*(RHF/STO-3G) = −300.078097 a.u. with $-V/T = 2.00643103$.
[f]This integration time on a VaxStation3100.
[g]$(CN)_5H_7$, *E*(RHF/STO-3G) = −459.601827 a.u. with $-V/T = 2.01000174$.

imal neighborhood is considered, selection of the subset with $NNN = 3$ dramatically reduces this error to acceptable values in all cases (< 0.03 for population and < 2 kcal mol^{-1} for kinetic energy).

The increased accuracy gained with the increase in NNN leads of course to higher integration time requirements, as is shown for a few atoms in Fig. 18. Figure 14 shows that the integration time requirement per atom increases steeply at the beginning and then shows only a very slow increase as the size of the system increases. This is equivalent to saying that there is a maximum number of LMOs that need to be considered for a given atom. With increasing *NNN*, this maximum number of LMOs becomes larger, and the advantage of the subset selection method becomes apparent only for larger molecules. Fig-

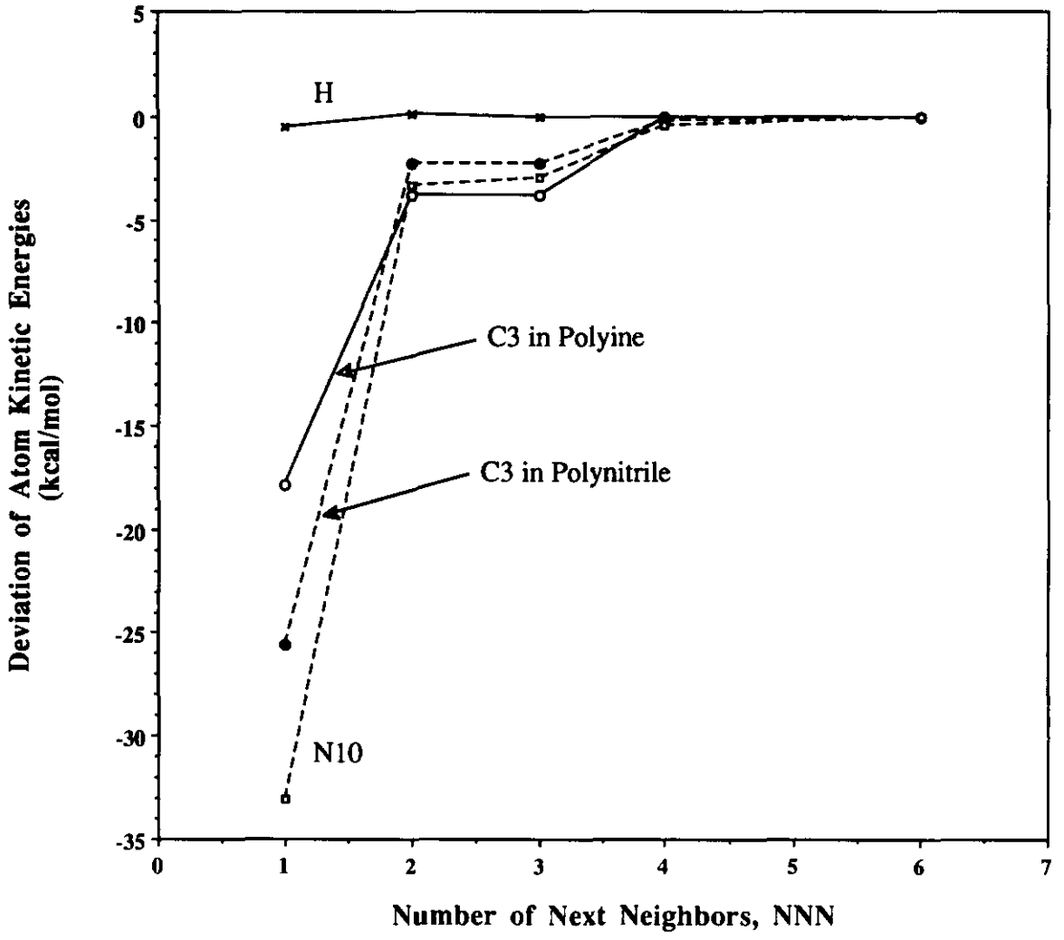


Fig. 16. Effect of the number of next neighbors on the integrated kinetic energy of atoms C3 and H in C_8H_2 and on atoms C3 and N10 in $C_5N_5H_7$. Energy is given relative to the value determined with the total electron density function (in kcal mol^{-1}).

ure 18 shows significant time savings still for $C_5N_5H_7$ but only modest ones for C_8H_2. The point is perfectly made by comparison of the integration times required for C1 in the polynitriles $C_nN_nH_{n+2}$ (n=1–5 and 10) with an *NNN* value of 4 (Fig. 19). This selection of *NNN* results in excellent agreement between values derived from ρ_{mol} and from ρ''_M (*NNN*=4). Topological properties of the C1 basins are documented in Table 10 and the integrated properties are listed in Table 11.

Applicability to general basis sets

Hexatriine was studied to test the performance of the orbital selection algorithm regarding its applicability to general basis sets. Minimal (STO-3G), split-valence (e.g. 3-21G, 6-31G), and triple-ζ type basis sets (e.g. 6-311G), in addition to several polarized basis sets (3-21G*, 6-31G*, 6-311G*) were considered. In each case, the electron density functions calculated for the optimized structures (Table 12) was analyzed. The topological properties are documented in Table 13; in Table 14 the integrated atomic properties of C1 and H are summarized. Integrated atomic properties were determined both with

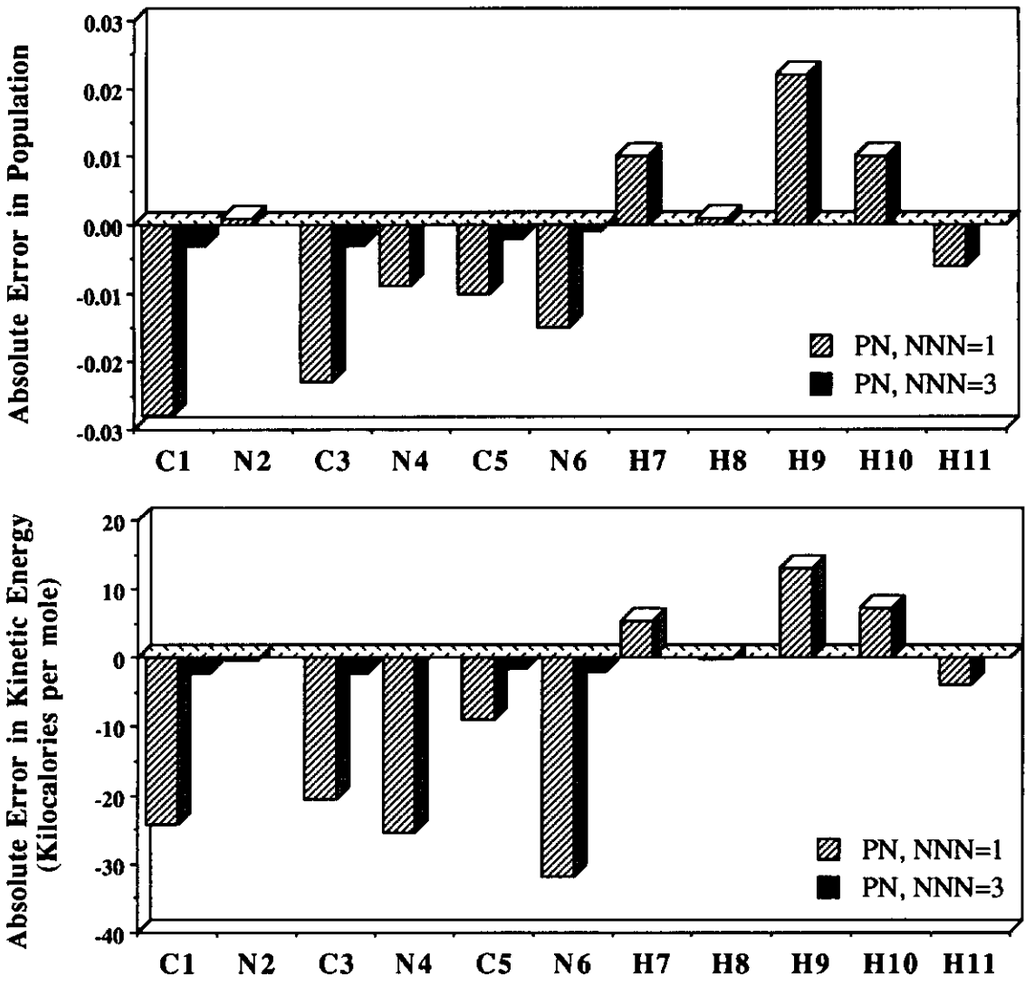


Fig. 17. Absolute error in the population and in the kinetic energy shown for $C_3N_3H_5$ and $C_5N_5H_7$, as determined with the electron density functions ρ''_M ($NNN=1$) and ρ''_M ($NNN=3$). Atom numbers are those for $C_3N_3H_5$ and the corresponding atoms in the larger molecule.

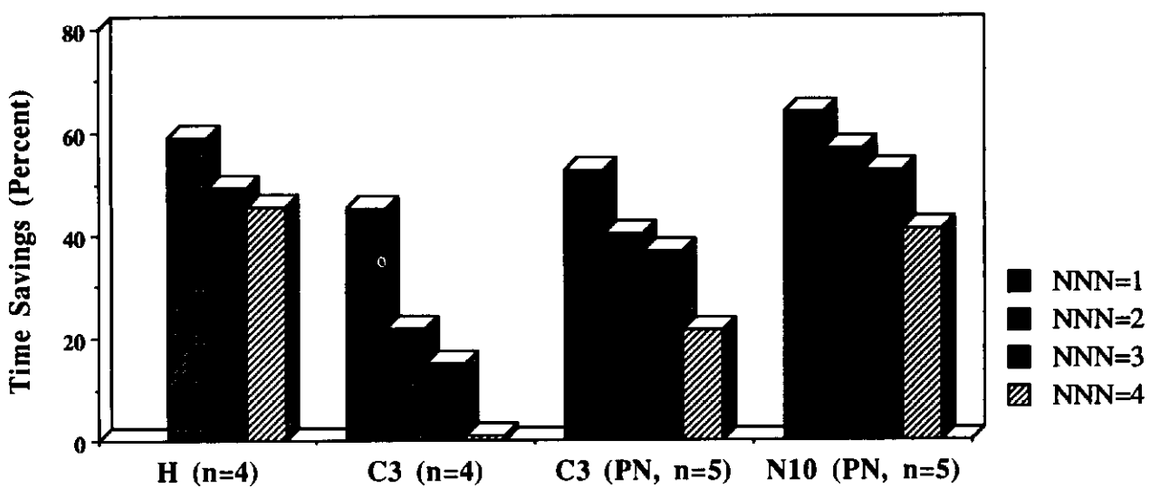


Fig. 18. Time saving for the determination of integrated atomic properties for selected atoms of C_8H_2 (H and C3) and $C_5N_5H_7$ (C3 and N10) as a function of the neighborhood selection.

the total electron density distribution r_{mol} and with the electron density functions ρ''_{C1} and ρ''_H, using a minimal neighborhood in each subset selection.

The characteristic values of the critical points are excellently reproduced

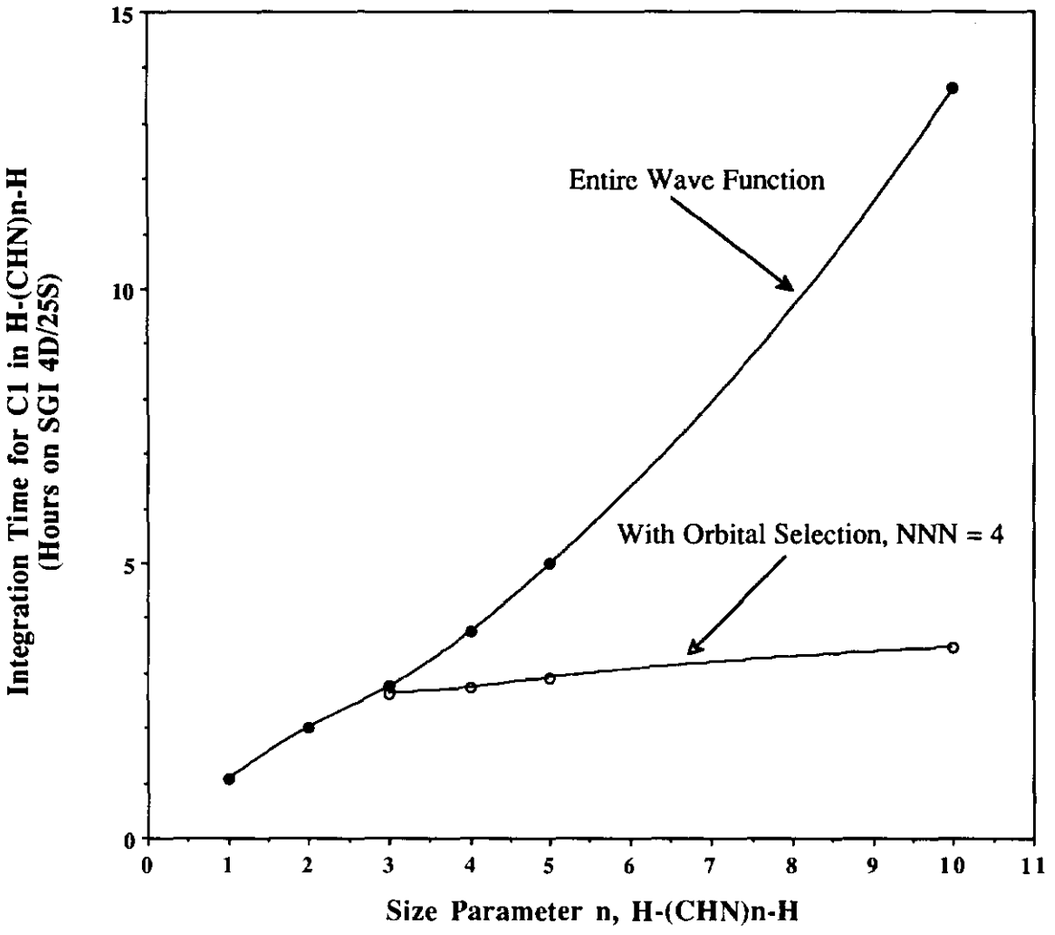


Fig. 19. Integration time requirements for the C1 atoms in the polynitriles $C_nN_nH_{n+2}$ ($n=1$-5 and 10), using the electron density functions ρ_{mol} and ρ''_{C1} ($NNN=4$).

(Table 13) at all theoretical levels. The topological parameters differ at most by a few thousand, a difference that is within the numerical accuracy of the methods by which the properties are computed. Absolute errors in atom population, atomic first moment, and atom kinetic energy are compared graphically in Fig. 20 for the different basis sets. The absolute deviations of the population and of the dipole moment are relatively small and of the same magnitude or less than the corresponding values discussed for the polyines with the minimal basis set (vide infra). Absolute deviation of the population is less than ± 0.05 and atomic dipole moments differ by less than $\pm 0.015\,e$ a.u. Differences in the kinetic energies may become as large as 20 kcal mol^{-1} with the split-valance basis sets, but the deviations generally are significantly smaller in the case of the split-valence basis sets, compared to the minimal basis set results. The accurate determination of the kinetic energy would require the inclusion of second-generation neighbors (vide supra). Absolute time requirements for the integrations do of course increase with the size of the basis sets (Table 14) but Fig. 21 shows that the relative time savings is comparatively independent of the specific basis set. The orbital selection algorithm performs with equal efficiency for all kinds of basis set.

This preliminary examination of basis set dependency suggests that the de-

TABLE 10

Topological properties for the C1 basins in H-$(HCN)_n$-H[a]

A-B bond	Atom	r_A	r_B	ρ	λ_1	λ_2	λ_3
n=*1*							
C-N1	All	0.4208	0.6148	0.3330	−0.6707	−0.4096	1.3391
H_t-C1	All	0.4471	0.6417	0.2630	−0.5857	−0.5649	0.4804
H_c-C1	All	0.4526	0.6382	0.2605	−0.5669	−0.5517	0.4730
n=*2*							
C-N1	All	0.4213	0.8565	0.3280	−0.6456	−0.4033	1.3404
H_t-C1	All	0.4503	0.6415	0.2605	−0.5704	−0.5546	0.4808
H_c-C1	All	0.4436	0.6445	0.2631	−0.5912	−0.5700	0.4837
n=*3*							
C1-N2	All	0.4216	0.8574	0.3273	−0.6424	−0.4056	1.3347
	C1	0.4216	0.8574	0.3273	−0.6424	−0.4056	1.3347
H_t-C1	All	0.4496	0.6423	0.2605	−0.5711	−0.5562	0.4819
	C1	0.4496	0.6423	0.2605	−0.5711	−0.5562	0.4819
H_c-C1	All	0.4423	0.6460	0.2631	−0.5930	−0.5726	0.4857
	C1	0.4423	0.6460	0.2631	−0.5930	−0.5726	0.4857
n=*4*							
C1-N2	All	0.4217	0.8576	0.3271	−0.6413	−0.4074	1.3315
	C1	0.4217	0.8576	0.3271	−0.6413	−0.4074	1.3315
H_t-C1	All	0.4493	0.6426	0.2605	−0.5715	−0.5571	0.4824
	C1	0.4493	0.6426	0.2605	−0.5715	−0.5571	0.4824
H_c-C1	All	0.4417	0.6467	0.2632	−0.5939	−0.5740	0.4866
	C1	0.4417	0.6467	0.2632	−0.5939	−0.5740	0.4866
n=*5*							
C1-N2	All	0.4218	0.8577	0.3270	−0.6408	−0.4083	1.3301
	C1	0.4218	0.8577	0.3270	−0.6408	−0.4084	1.3302
H_t-C1	All	0.4492	0.6428	0.2605	−0.5716	−0.5575	0.4827
	C1	0.4492	0.6428	0.2605	−0.5716	−0.5576	0.4827
H_c-C1	All	0.4413	0.6470	0.2632	−0.5944	−0.5748	0.4872
	C1	0.4413	0.6470	0.2632	−0.5944	−0.5748	0.4872
n=*10*							
C1-N2	All	0.4219	0.8579	0.3269	−0.6402	−0.4097	1.3284
	C1	0.8579	0.4219	0.3269	−0.6402	−0.4099	1.3284
H_t-C1	All	0.4490	0.6431	0.2605	−0.5718	−0.5581	0.4831
	C1	0.4490	0.6431	0.2605	−0.5718	−0.5582	0.4831
H_c-C1	All	0.4408	0.6477	0.2633	−0.5952	−0.5759	0.4881
	C1	0.4408	0.6477	0.2633	−0.5952	−0.5760	0.4881

[a]All orbital selections done with *NNN*=4. For an explanation of column headings see footnotes to Table 3.

TABLE 11

Integration time requirements for C1 in polynitriles H-$(HCN)_n$-H, both with and without orbital selection[a]

n	IP		μ		T[b]		t	
	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs
1	5.141		0.749		36.727055		1.03	
2	5.157		0.758		36.785493		2.02	
3	5.159	5.159	0.761	0.761	36.802011	36.802011	2.75	2.63
4	5.159	5.159	0.762	0.767	36.810299	36.809932	3.75	2.73
5	5.159	5.158	0.763	0.763	36.815366	36.814869	4.99	2.91
10	5.158	5.157	0.762	0.762	36.826510	36.825972	13.62	3.46

[a]All orbital selections done with $NNN=4$. For an explanation of column headings see footnotes to Table 4.
[b]Kinetic energies are corrected for the virial defect of the wave functions.

TABLE 12

Geometry and energy of hexatriine at various theoretical levels[a]

Basis set	H-C	C5-C3	C3-C1	C1-C2	Energy	$-V/T$
STO-3G	1.0665	1.1756	1.4034	1.1816	−225.336835	2.00606962
3-21G	1.0515	1.1908	1.3690	1.1923	−226.914828	2.00367761
3-21G*	1.0598	1.1877	1.3863	1.1909	−227.108734	2.02221608
6-31G	1.0535	1.1966	1.3766	1.1985	−225.336835	2.00606962
6-31G*	1.0573	1.1881	1.3851	1.1915	−228.178357	2.00119659
6-311G	1.0507	1.1898	1.3750	1.1921	−228.154910	1.99988710
6-311G*	1.0562	1.1848	1.3803	1.1872	−228.228444	2.00031764

[a]Distance in Å and energy in a.u.

termination of topological properties via the subsets of selected localized molecular orbitals generally performs better with the larger basis sets. The split-valence basis sets apparently result in "larger" LMOs (owing to the smaller exponents of the primitive functions) and, as a consequence, the lists of atom-specific LMOs include a larger number of MOs. More detailed studies of these features are in progress.

Interestingly, the topologically analyses of C_6H_2 show that the non-nuclear attractors found at the RHF/STO-3G level vanish when the split valence basis sets 3-21G, 6-31G, and 6-311G are employed. However, we find that these non-nuclear attractors reoccur when the split-valence basis sets are supplemented by polarization functions. Non-nuclear attractors with essentially the same

TABLE 13

Basis set effects on topological properties of C_6H_2[a]

A–B Bond	MOs	r_A	r_B	ρ	λ_1	λ_3[b]
STO-3G						
C1–C2	All	0.4482	0.7334	0.3582	−0.2913	0.4454
BCP at C1	C1	0.4480	0.7336	0.3581	−0.2967	0.4475
C1–C2	All	0.5908	0.5908	0.3611	−0.2946	−0.1368
NNA	C	0.5911	0.5905	0.3611	−0.2950	−0.1372
C3–C1	All	0.6730	0.7304	0.2731	−0.4081	0.0510
	C1	0.6751	0.7283	0.2730	−0.4096	0.0506
	C3	0.6709	0.7325	0.2729	−0.4095	0.0511
H7–C5	All	0.4221	0.6444	0.2640	−0.5885	0.4621
	C5	0.4221	0.6444	0.2640	−0.5917	0.4621
	H7	0.4219	0.6446	0.2639	−0.5916	0.4618
3-21G						
C1–C2	All	0.5941	0.5962	0.4036	−0.6598	0.1084
BCP at C1	C1	0.5962	0.5961	0.4036	−0.6597	0.1083
C1–C2[c]	All					
NNA[c]	C1					
C3–C1	All	0.6745	0.6945	0.2964	−0.5532	0.1639
	C1	0.6745	0.6945	0.2964	−0.5531	0.1639
H7–C5	All	0.3584	0.6931	0.2782	−0.7424	0.6043
	H7	0.3584	0.6931	0.2781	−0.7446	0.6043
*3-21G**						
C1–C2	All	0.4541	0.7368	0.4135	−0.6065	0.3472
BCP at C1	C1	0.4540	0.7369	0.4135	−0.6592	0.3476
C1–C2	All	0.5954	0.5954	0.4154	−0.5560	−0.0787
NNA	C1	0.5954	0.5954	0.4154	−0.6072	−0.7879
C3–C1	All	0.5697	0.8166	0.3201	−0.6392	0.7157
	C1	0.5694	0.8167	0.3200	−0.6352	0.7019
H7–C5	All	0.3447	0.7152	−0.2925	−0.8314	0.5598
	H7	0.3445	0.7153	0.2924	−0.8268	0.5601
6-31G						
C1–C2	All	0.5992	0.5992	0.3938	−0.6227	0.0341
BCP at C1	C1	0.5995	0.5990	0.3938	−0.6226	0.0340
C1–C2[c]	All					
NNA[c]	C1					
C3–C1	All	0.6841	0.6925	0.2912	−0.5339	0.3160
	C1	0.6841	0.6925	0.2912	−0.5339	0.3159
H7–C5	All	0.3647	0.6888	0.2782	−0.7498	0.5015
	H7	0.3647	0.6888	0.2781	−0.7515	0.5007

TABLE 13 (continued)

A–B Bond	MOs	r_A	r_B	ρ	λ_1	λ_3[b]
*6-31G**						
C1–C2	All	0.4403	0.7511	0.4119	−0.5969	0.4799
BCP at C1	C1	0.4403	0.7512	0.4119	−0.6467	0.4805
C1–C2	All	0.5958	0.5958	0.4147	−0.5964	−0.8459
NNA	C1	0.5957	0.5958	0.4147	−0.6445	−0.8465
C3–C1	All	0.6851	0.7001	0.3134	−0.6235	0.2221
	C1	0.6851	0.7001	0.3134	−0.6220	0.2221
H7–C5	All	0.3456	0.7117	0.2952	−0.8531	0.4273
	H7	0.3456	0.7117	0.29519	−0.8490	0.4273
6-311G						
C1–C2	All	0.5961	0.5960	0.3987	−0.6522	0.0818
BCP at C1	C1	0.5961	0.5960	0.3987	−0.6522	0.0818
C1–C2[b]	All					
	C1					
C3–C1	All	0.6813	0.6938	0.2945	−0.5464	0.2741
	C1	0.6813	0.6938	0.2945	−0.5464	0.2741
H7–C5	All	0.3556	0.6952	0.2819	−0.7765	0.5896
	H7	0.3556	0.6952	0.2819	−0.7786	0.5896
*6-311G**						
C1–C2	All	0.4976	0.6896	0.4187	−0.6449	0.0755
BCP at C1	C1	0.4976	0.6896	0.4187	−0.6921	0.0754
C1–C2	All	0.5936	0.5936	0.4189	−0.6425	−0.2174
NNA	C1	0.5934	0.5938	0.4189	−0.6893	−0.2177
C3–C1	All	0.6783	0.7020	0.3145	−0.6243	0.1917
	C1	0.6783	0.7020	0.3145	−0.6253	0.1916
H7–C5	All	0.3444	0.7118	0.2984	−0.8660	0.5323
	H7	0.3444	0.7118	0.2983	−0.8626	0.5322

[a]For an explanation of column headings see footnotes to Table 3.
[b]$\lambda_2 = \lambda_3$ in all cases.
[c]No non-nuclear attractor.

characteristics occur in all triple bonds at the RHF level with the 3-21G*, the 6-31G*, and the 6-311G* basis sets. This finding indicates that polarization functions play an important role in the correct description of triple bonds and second-order polarization functions, f-type functions, might indeed be crucial. Clearly, more detailed higher-level studies, with the best basis sets and including correlation corrections, are required to investigate this intriguing feature and its implications for the properties of polyines. With the methods described, we can now proceed to tackle this interesting problem with higher level electron density studies of the polyines in an efficient way, even for rather extended systems.

TABLE 14

Basis set dependency of integrated properties of C1 and H in C_6H_2[a b]

Basis set	No.[c]	*IP*		μ		T[d]		t[e]	
		All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs	All MOs	Sel. MOs
C1									
STO-3G	16	5.4296	5.3743	0.6816	0.6827	36.722959	36.657372	2.37	1.83
3-21G	16	6.0346	6.0341	−0.0012	−0.0006	37.509027	37.508323	2.40	2.08
3-21G*	15	5.5333	5.5340	1.0440	1.0320	36.516109	36.512872	3.00	2.66
6-31G	16	6.0219	6.0255	−0.0138	−0.0179	37.836266	37.837910	2.58	2.26
6-31G*	16	5.3129	5.3151	0.6408	0.6444	37.341850	37.338088	3.35	2.96
6-311G	16	6.0222	6.0234	−0.0043	−0.0054	37.847340	37.847985	2.77	2.43
6-311G*	16	5.6423	5.6179	0.3779	0.3899	37.608339	37.581112	3.54	3.11
H									
STO-3G	5	0.9175	0.9176	0.1080	0.1074	0.561875	0.561132	1.47	0.69
3-21G	7	0.7998	0.8007	0.1279	0.1271	0.522816	0.522651	1.57	0.77
3-21G*	5	0.7992	0.7635	0.1282	0.1387	0.522539	0.495759	1.56	0.82
6-31G	5	0.8279	0.8279	0.1178	0.1166	0.538410	0.537762	1.57	0.70
6-31G*	7	0.7985	0.7958	0.1286	0.1289	0.520349	0.519656	1.87	0.95
6-311G	6	0.8012	0.8019	0.1319	0.1310	0.526826	0.526628	1.85	0.89
6-311G*	6	0.7898	0.7875	0.1309	0.1310	0.517399	0.517098	2.01	0.98

[a]First generation neighbors considered only, $NNN=1$.
[b]For an explanation of column headings see footnotes to Table 4.
[c]Number of selected (sel.) localized molecular orbitals of hexatriine.
[d]Kinetic energies as integrated. For virial ratios, see Table 12.
[e]Integration times in Silicon Graphics Personal Iris 4D/25S.

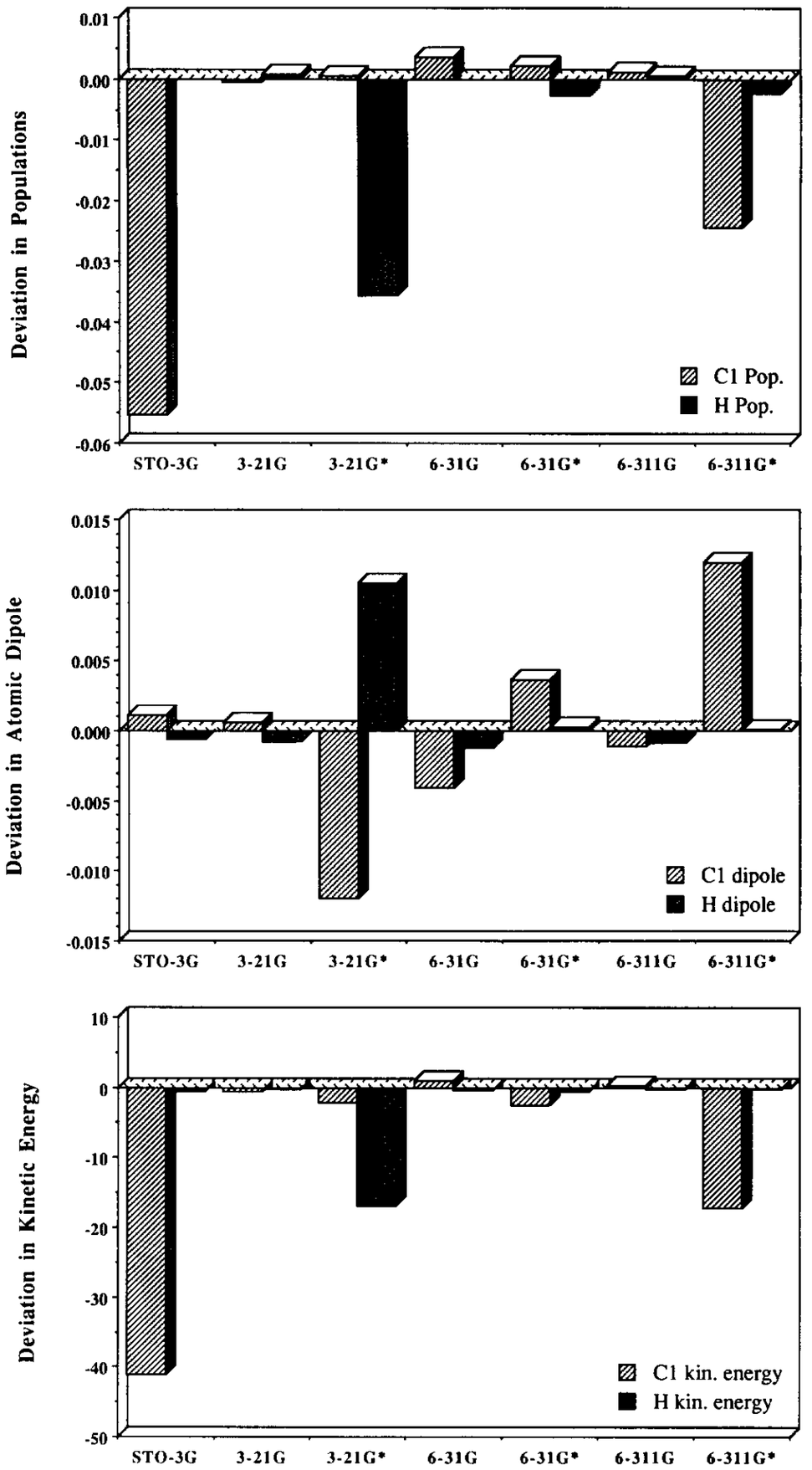


Fig. 20. Absolute deviation in the atom population, first atomic moment (in e a.u.), and atom kinetic energy (in kcal mol^{-1}), determined with the electron density functions ρ''_M ($NNN=1$) and ρ_{mol}, respectively, in dependence on the theoretical model.

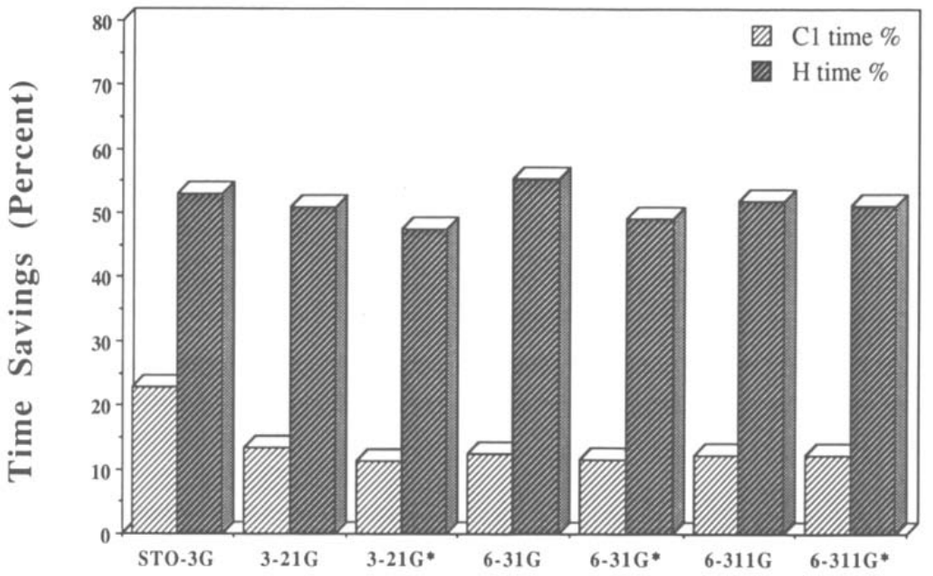


Fig. 21. Time savings (%) for determination of the integrated properties of C1 and H in C_6H_2, based on the electron density functions ρ''_M ($NNN=1$), in dependence on the basis set.

CONCLUSION

The theoretical principles of two methods that allow for a more efficient determination of integrated atomic properties have been described. The methods were computationally implemented and tested for a series of polyines $C_{2n}H_2$ and a series of polynitrils $C_nN_nH_{n+2}$ ($n=1$–5 and 10). Both of these methods are based on atom-specific electron density functions that accurately describe the electron density distribution in the basin of the atom whose properties are being determined.

The first of these methods aims at reducing the number of primitive functions considered in the integration and the atom specific electron density function $\rho'_M(x,y,z,NNN)$ has been introduced to achieve this aim. The second method is directed at reducing the number of MOs that need to be considered for the integration of a specific atom. The sequence involving MO localization, selection of atom-specific sets of LMOs, and subsequent determination of atom-specific subsets of LMOs has been developed. The electron density function $\rho''_M(x,y,z,NNN)$ determined by this sequence has been shown to produce atomic properties that are in excellent agreement with values derived from ρ_{mol} and that this method performs its task exceptionally well.

For both of these new methods the concept of the "neighborhood of an atom" is crucial for the determination of the electron density functions and the parameter *NNN* was introduced to define the neighborhood. Although *NNN* is easily defined via connectivity in the test cases discussed, this parameter is more generally defined by a distance criterion. For the first method, *NNN* follows from the exponents of the primitive functions and $NNN=2$ yields excellent results. For the second method, *NNN* needs to be determined in a heu-

ristic fashion. Consideration of all neighboring atoms, and of their neighbors too, yields rather satisfactory values of population and the atomic dipole moment. Kinetic energy values are more susceptible to the selection of *NNN*. An *NNN* value of 4 reproduces all integrated properties essentially exactly.

The method based on the selection of subsets of LMOs has been shown to significantly reduce the integration time requirements. In particular, we have emphasized that the time savings increase with the size of the system because there exists an upper limit to the number of LMOs that need to be considered for each atom. The primitive elimination method, in contrast, leads only to modest increases in efficiency with the current integration program. Nevertheless, this method has tremendous potential if the appropriate changes are made to the integration algorithm.

The preliminary study of basis set effects indicates that the subset selection technique performs equally well for all kinds of basis sets. This result is significant because it suggests that the method can be successfully applied to molecules that are described by "mixed" basis sets [30,31]. Especially for the study of very large molecules, such as pharmaceutically interesting drugs or for physiologically active biopolymers, for example, the "splicing" of basis sets might prove essential and recent systematic studies [32] are indeed encouraging.

Ultimately, it is intended to combine the methods of obtaining atom-specific electron density functions with the least number of localized molecular orbitals expanded in the least number of primitives. The results presented here demonstrate that the determination of integrated atomic properties could be carried out rather efficiently with such electron density functions and with an appropriate choice of the neighborhood. Greatly reduced integration times can be achieved while maintaining the generality, the rigor, and the accuracy of the topological method. It is emphasized that all of the manipulations that lead to the functions ρ'_M and ρ''_M are performed after the wave functions have been determined. The rigorous electron density analysis of very large molecules appears feasible.

ACKNOWLEDGMENTS

We thank the Campus Computing Center for a generous grant of computer time. Acknowledgment is made gratefully to the Donors of The Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. We thank Professor Bader for the programs EXTREME and PROAIM.

REFERENCES

1 R. Glaser, J. Comput. Chem., 10 (1989) 188-135.
2 R.S. Mulliken, J. Chem. Phys., 23 (1955) 1833, 1848, 2338, 2343.
3 (a) R.E. Davidson, J. Chem. Phys., 46 (1967) 3320.
(b) C. Ehrhardt and R. Ahlrichs, Theor. Chim. Acta, 68 (1985) 231.
4 A.E. Reed, R.B. Weinstock and F. Weinhold, J. Chem. Phys., 83 (1985) 735.
5 P. Politzer and P.H. Reggio, J. Am. Chem. Soc., 94 (1972) 8308, and references cited therein.
6 K.B. Wiberg, J. Am. Chem. Soc., 102 (1980) 1229.
7 (a) J.B. Collins, A. Streitwieser and J.M. McKelvey, J. Comput. Chem., 3 (1979) 165.
(b) R.S. McDowell, D.L. Grier and A. Streitwieser, J. Comput. Chem., 9 (1985) 165.
8 R.F.W. Bader, Atoms in Molecules—A Quantum Theory, Clarendon, Oxford, 1990.
9 R.F.W. Bader, Acc. Chem. Res., 18 (1985) 9.
10 F.W. Biegler-Koenig, R.F.W., Bader and T.H. Tang, J. Comput. Chem., 3 (1982) 317.
11 J. Cioslowski, S.T. Mixon and W.D. Edwards, J. Am. Chem. Soc., 113 (1991) 1083.
12 (a) W.J. Hehre, R.F. Stewart and J.A. Pople, J. Chem. Phys., 51 (1969) 2657.
(b) J.B. Collins, P.v.R. Schleyer, J.S. Binkley and J.A. Pople, J. Chem. Phys., 64 (1976) 5142.
13 W.J. Hehre, L. Radom, P.v.R. Schleyer and J.A. Pople, Ab Initio Molecular Orbital Theory, Wiley, New York, 1986.
14 (a) M.S. Gordon, J.S. Binkley, J.A. Pople, W.J. Pietro and W.J. Hehre, J. Am. Chem. Soc., 104 (1982) 2797.
(b) W.J. Pietro, M.M. Francl, W.J. Hehre, D.J. Defrees, J.A. Pople and J.S. Binkley, J. Am. Chem. Soc., 104 (1982) 5039.
15 (a) W.J. Hehre, R. Ditchfield and J.A. Pople, J. Phys. Chem., 56 (1972) 2257.
(b) P.C. Hariharan and J.A. Pople, Theor. Chim. Acta, 28 (1973) 213.
16 R. Krishnan, J.S. Binkley, R. Seeger and J.A. Pople, J. Phys. Chem., 72 (1980) 650.
17 R.F.W. Bader, T.S. Slee, D. Cremer and E. Kraka, J. Am. Chem. Soc., 105 (1983) 5061.
18 R. Glaser, J. Phys. Chem., 93 (1989) 7993.
19 (a) C. Gatti, P. Fantucci and G. Pacchioni, Theor. Chim. Acta, 72 (1987) 433.
(b) W.L. Cao, C. Gatti, P.J. MacDougall and R.F.W. Bader, Chem. Phys. Lett., 141 (1987) 380.
(c) R. Glaser, R.F. Waldron and K.B. Wiberg, J. Phys. Chem., 94 (1990) 7357.
20 R. Glaser, J. Phys. Chem., 93 (1989) 7993.
21 R. Glaser, J. Comput. Chem., 6 (1990) 663.
22 R. Glaser, G.S.-C. Choy and K. Hall, J. Am. Chem. Soc., 113 (1991) 1109.
23 R. Glaser, C. Horan, E. Nelson and M.K. Hall, J. Org. Chem., 57 (1992).
24 R. Glaser and B.L. Harris, in preparation.
25 D.L. Severance and W.L. Jorgensen, Department of Chemistry, Purdue University, West Lafayette, IN, 1988.
26 R. Glaser, University of Missouri-Columbia, 1990.
27 S.F. Boys, Rev. Mod. Phys., 32 (1960) 296.
28 M.W. Schmidt, K.K. Baldridge, J.A. Boatz, J.H. Jensen, S. Koseki, M.S. Gordon, K.A. Nguyen, T.L. Windus and S.T. Elbert, QCPE Bull. 10 (1990).
29 B.L. Harris and R. Glaser, in preparation.
30 W.W. Schmidt, R.O. Angus and R.P. Johnson, J. Am. Chem. Soc., 104 (1982) 6839.
31 L. Pardo, A.P. Mazurek and R. Osman, Int. J. Quantum Chem., 104 (1990) 701.
32 J.H. Jensen and M.S. Gordon, J. Comput. Chem., 12 (1991) 421.

APPENDIX: Z-MATRIX OF $C_{10}N_{10}H_{12}$

C\N, 1, CN1\C, 2, CN2, 1, A1\N, 3, CN3, 2, A2, 1, 180., 0\C, 4, CN4, 3, A3, 2, 180., 0\N, 5, CN5, 4, A4, 3, 180., 0\C, 6, CN6, 5, A5, 4, 180., 0\N, 7, CN7, 6, A6, 5, 180., 0\C, 8, CN8, 7, A7, 6, 180., 0\N, 9, CN9, 8, A8, 7, 180., 0\C, 10, CN10, 9, A9, 8, 180., 0\N, 11, CN11, 10, A10, 9, 180., 0\C, 12, CN12, 11, A11, 10, 180., 0\N, 13, CN13, 12, A12, 11, 180., 0\C, 14, CN14, 13, A13, 12, 180., 0\N, 15, CN15, 14, A14, 13, 180., 0\C, 16, CN16, 15, A15, 14, 180., 0\N, 17, CN17, 16, A16, 15, 180., 0\C, 18, CN18, 17, A17, 16, 180., 0\N, 19, CN19, 18, A18, 17, 180., 0\H, 1, H1, 2, AH1, 3, 0., 0\H, 1, H2, 2, AH2, 3, 180., 0\H, 3, H3, 2, AH3, 1, 0., 0\H, 5, H4, 4, AH4, 3, 0., 0\H, 7, H5, 6, AH5, 5, 0., 0\H, 9, H6, 8, AH6, 7, 0., 0\H, 11, H7, 10, AH7, 9, 0., 0\H, 13, H8, 12, AH8, 11, 0., 0\H, 15, H9, 14, AH9, 13, 0., 0\H, 17, H10, 16, AH10, 15, 0., 0\H, 19, H11, 18, AH11, 17, 0., 0\H, 20, H12, 19, AH12, 18, 180., 0\\AH1 = 124.706262\AH2 = 119.058347\AH3 = 117.041922\AH4 =117.222819\AH5 = 117.218103\AH6 = 117.241256\AH7 = 117.22364\AH8 = 117.196499\AH9 = 117.159884\AH10 =117.069291\AH11 = 116.459942\AH12 = 108.549184\A1 = 114.256748\A2 = 118.910588\A3 = 114.147033\A4 = 118.858381\A5 = 114.128474\A6 = 118.796024\A7 = 114.137551\A8 = 118.81138\A9 = 114.204265\A10 = 118.795881\A11 = 114.193688\A12 = 118.782361\A13 = 114.210306\A14 = 118.784794\A15 = 114.233486\A16 = 118.851862\A17 = 114.262205\A18 = 118.545731\H1 = 1.092026\H2=1.088401\H3 = 1.097887\H4 = 1.097596\H5 = 1.097519\H6 = 1.097393\H7 = 1.097308\H8 = 1.097255\H9 = 1.097146\H10 = 1.097026\H11 = 1.096217\H12 = 1.04789\CN1 = 1.278391\CN2 = 1.458152\CN3 = 1.282117\CN4 = 1.452962\CN5 = 1.283024\CN6 = 1.452547\CN7 = 1.282953\CN8 = 1.452609\CN9 = 1.28303\CN10 = 1.453041\CN11 = 1.282845\CN12 = 1.453102\CN13 = 1.28263\CN14 = 1.453527\CN15 = 1.282193\CN16 = 1.454818\CN17 = 1.28073\CN18 = 1.460393\CN19 = 1.275951