

# Calculation of Atomic Integration Data

FRIEDRICH BIEGLER-KÖNIG

*Fachbereich Mathematik und Technik, Fachhochschule Bielefeld, Am Stadtholz 24, 33609 Bielefeld, Germany*

*Received 7 February 2000; accepted 20 March 2000*

**ABSTRACT:** The calculation of average properties of atoms in molecules and interatomic surfaces is a difficult problem that requires the evaluation of two- and three-dimensional integrals over regions with nontrivial borders. A mathematical formalism is presented that maps these regions onto the whole of  $\mathbb{R}^2$  and/or  $\mathbb{R}^3$  and allows the construction of efficient and reliable numerical methods for the calculation of these integrals. These methods, which will be part of a forthcoming program package, are described and examples are given. © 2000 John Wiley & Sons, Inc. *J Comput Chem* 21: 1040–1048, 2000

**Keywords:** atoms in molecules; atomic basins; interatomic surfaces; average properties; integration over atomic basins

## Introduction

In this article we describe algorithms for the calculation of average properties of atoms in molecules. Atoms in molecules are defined by topological properties of the molecular electronic charge density,  $\rho(\mathbf{x})$ . The underlying theory<sup>1</sup> allows the identification of an atom with a well-defined region in three-dimensional space, the atomic basin. Atomic basins are separated by so-called zero-flux surfaces, which contain a  $(3, -1)$ -critical point, and satisfy the condition:

$$\nabla\rho(\mathbf{x})^T \cdot \mathbf{n}(\mathbf{x}) = 0 \quad \text{for all points } \mathbf{x} \text{ in the surface} \quad (1)$$

where  $\mathbf{n}(\mathbf{x})$  is the normal to the surface in  $\mathbf{x}$ .

A property,  $F$ , of an atom,  $A$ , with basin  $\Omega$  is defined as a three-dimensional integral over an in-

tegrable function  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ :

$$F(\Omega) = \int_{\Omega} f(\mathbf{x}) d\mathbf{x} \quad (2)$$

Similarly, we can write a property for the interatomic surface,  $S_{AB}$ , between the basins of atoms  $A$  and  $B$  as a two-dimensional surface integral:

$$F(S_{AB}) = \int_{S_{AB}} f(\mathbf{x}) d\mathbf{x} \quad (3)$$

The evaluations of  $F(\Omega)$  and  $F(S_{AB})$  are nontrivial, because the interatomic surfaces in general do not have a local definition or a simple geometrical structure. An interatomic surface is defined globally as the set of points that belongs to gradient paths of  $\rho$  going downward from a  $(3, -1)$ -critical point.

An obvious approach for calculating these properties is to approximate the interatomic surfaces. Volume integration is then usually done in a polar coordinate system centered on the nucleus. Concepts of this kind have been proposed in refs. 2–4.

Correspondence to: F. Biegler-König; e-mail: bieglerr@mathe.fh-bielefeld.de

Because modeling an interatomic surface is very complicated, especially in the case of ring or cage structures, programs based on such an approach are often inaccurate and unreliable.

The algorithms we present are based on an earlier idea<sup>5</sup> that uses a coordinate transformation to map an atom in a molecule onto the whole of  $\mathbb{R}^3$  or an interatomic surface onto the whole of  $\mathbb{R}^2$ . This approach is mathematically more elegant and more reliable in its computations, but has been abandoned due to numerical problems and high computational requirements. These problems have been overcome in a new implementation.

In what follows we revisit the approach of ref. 5 for volume integration and supplement it with some new mathematical ideas. We also present some examples for the calculation of properties of atoms in molecules. Then we generalize some of the findings and obtain an algorithm for surface integration. This algorithm is illustrated in the fifth section.

## Integration Over Atomic Basins

Let  $\mathbf{X}_A \in \mathbb{R}^3$  be the position of the charge density maximum of atom  $A$  in a molecule. We use an atomic coordinate system centered in  $\mathbf{X}_A$  (all units are atomic units).

### GRADIENT PATHS AND ATOMIC BASINS

The gradient path,  $g(\mathbf{x})$ , through a noncritical point,  $\mathbf{x} \in \mathbb{R}^3$  ( $\nabla \varrho(\mathbf{x}(s)) \neq 0$ ), is defined by an initial value problem:

$$g(\mathbf{x}) = \left\{ \mathbf{x}(s) \in \mathbb{R}^3 \mid \frac{d\mathbf{x}(s)}{ds} = -\nabla \varrho(\mathbf{x}(s)); \mathbf{x}(0) = \mathbf{x} \right\} \quad (4)$$

The coordinates of all points of a gradient path can be computed using initial value problem, eq. (4). Each point on the path has a unique value of  $s \in \mathbb{R}$ .

Let  $\mathbf{S}_\beta$  be a sphere of radius  $\beta$  with a center in  $\mathbf{X}_A$  with the following properties:

- $\mathbf{S}_\beta \subset \Omega$ ; and
- each gradient path that terminates in  $\mathbf{X}_A$  intersects  $\mathbf{S}_\beta$  exactly once.

In a polar coordinate system centered in  $\mathbf{X}_A$  with coordinates:

$$r \geq 0, \quad \theta \in [0, \pi], \quad \varphi \in [0, 2\pi]$$

we have, for all points on  $\mathbf{S}_\beta$ :

$$\mathbf{x}_{\theta\varphi} = \begin{pmatrix} \beta \cdot \sin(\theta) \cdot \cos(\varphi) \\ \beta \cdot \sin(\theta) \cdot \sin(\varphi) \\ \beta \cdot \cos(\theta) \end{pmatrix} \quad (5)$$

The atomic basin  $\Omega$  belonging to atom  $A$  is then defined as the set of points belonging to gradient paths intersecting  $\mathbf{S}_\beta$ :

$$\Omega = \{ \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \in g(\mathbf{x}_{\theta\varphi}), \theta \in [0, \pi], \varphi \in [0, 2\pi] \} \quad (6)$$

We can now assign a set of “natural atomic coordinates”  $(s, \theta, \varphi)$  to each point in  $\Omega$ . The coordinate  $s$  runs along a gradient path whose starting value on  $\mathbf{S}_\beta$  is uniquely defined by  $\theta$  and  $\varphi$ . The range of values of  $s$  is  $-\infty$  (corresponding to  $\mathbf{X}_A$ ) and  $+\infty$  (another critical point or infinity). Critical points themselves do not define gradient paths, because there is no unique descent direction. In this coordinate system the interatomic surfaces of  $\mathbf{X}_A$  are mapped onto infinity.

### TRANSFORMATION INTO NATURAL COORDINATES

The transformation between Cartesian and natural atomic coordinates is defined by the mappings  $h: \Omega \rightarrow \mathbb{R}^3$  and  $h^{-1}: \mathbb{R}^3 \rightarrow \Omega$ . With eqs. (4) and (5) we get:

$$h^{-1} \begin{pmatrix} s \\ \theta \\ \varphi \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{x}_{\theta\varphi} - \int_0^s \nabla \varrho(\mathbf{x}(t, \theta, \varphi)) dt \quad (7)$$

Integration in Cartesian coordinates can now be substituted by integration in natural coordinates:

$$F(\Omega) = \int_{\Omega} f(\mathbf{x}) d\mathbf{x} = \int_0^{2\pi} \int_0^{\pi} \int_{-\infty}^{\infty} f(\mathbf{x}(s, \theta, \varphi)) \cdot |J(h^{-1}(s, \theta, \varphi))| ds d\theta d\varphi \quad (8)$$

$|J(h^{-1}(s, \theta, \varphi))|$  denotes the determinant of the Jacobian matrix of  $h^{-1}$ ; that is, the determinant of the matrix of partial derivatives of  $h^{-1}$  with respect to  $s$ ,  $\theta$ , and  $\varphi$ :

$$J = J(h^{-1}(s, \theta, \varphi)) = \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{pmatrix} \\ \equiv (\mathbf{x}_s | \mathbf{x}_\theta | \mathbf{x}_\varphi) \quad (9)$$

To minimize computational effort, we can use polar coordinates inside  $\mathbf{S}_\beta$  [the Jacobian determinant for polar coordinates is  $r^2 \cdot \sin(\theta)$ ]. Eq. (8) then changes

into:

$$\begin{aligned}
 F(\Omega) &= F_\beta + F_\Omega \\
 &= \int_0^{2\pi} \int_0^\pi \sin(\theta) \int_0^\beta r^2 \cdot f(\mathbf{x}(r, \theta, \varphi)) dr d\theta d\varphi \\
 &\quad + \int_0^{2\pi} \int_0^\pi \int_0^\infty f(\mathbf{x}(s, \theta, \varphi)) \\
 &\quad \cdot |J(h^{-1}(s, \theta, \varphi))| ds d\theta d\varphi
 \end{aligned} \tag{10}$$

### JACOBIAN DETERMINANT

Introducing the Hessian matrix of partial second-order derivatives of  $\rho$ :

$$H(\rho) = \begin{pmatrix} \frac{\partial^2 \rho}{\partial x^2} & \frac{\partial^2 \rho}{\partial x \partial y} & \frac{\partial^2 \rho}{\partial x \partial z} \\ \frac{\partial^2 \rho}{\partial y \partial x} & \frac{\partial^2 \rho}{\partial y^2} & \frac{\partial^2 \rho}{\partial y \partial z} \\ \frac{\partial^2 \rho}{\partial z \partial x} & \frac{\partial^2 \rho}{\partial z \partial y} & \frac{\partial^2 \rho}{\partial z^2} \end{pmatrix}$$

and using eq. (4) and the chain rule for partial differentiation, the columns of  $J$  can be written as:

$$\begin{aligned}
 \frac{\partial}{\partial s} \mathbf{x}_s &= -\frac{\partial}{\partial s} \nabla \rho = -H(\rho) \cdot \mathbf{x}_s \\
 \frac{\partial}{\partial s} \mathbf{x}_\theta &= \frac{\partial}{\partial \theta} \mathbf{x}_s = -\frac{\partial}{\partial \theta} \nabla \rho = -H(\rho) \cdot \mathbf{x}_\theta \\
 \frac{\partial}{\partial s} \mathbf{x}_\varphi &= \frac{\partial}{\partial \varphi} \mathbf{x}_s = -\frac{\partial}{\partial \varphi} \nabla \rho = -H(\rho) \cdot \mathbf{x}_\varphi
 \end{aligned} \tag{11}$$

Hence, we get an initial value problem for  $J$ :

$$\frac{\partial}{\partial s} J = -H(\rho) \cdot J \tag{12}$$

with starting values:

$$\begin{aligned}
 &J(h^{-1}(0, \theta, \varphi)) \\
 &= \begin{pmatrix} -\frac{\partial \rho}{\partial x} & \beta \cdot \cos(\theta) \cdot \cos(\varphi) & -\beta \cdot \sin(\theta) \cdot \sin(\varphi) \\ -\frac{\partial \rho}{\partial y} & \beta \cdot \cos(\theta) \cdot \sin(\varphi) & \beta \cdot \sin(\theta) \cdot \cos(\varphi) \\ -\frac{\partial \rho}{\partial z} & -\beta \cdot \sin(\theta) & 0 \end{pmatrix}
 \end{aligned} \tag{13}$$

Because we are interested only in the determinant of  $J$ , we can use Davidenko's trace theorem to get a single differential equation for  $|J|$ :

**Davidenko's Trace Theorem<sup>6</sup>.** Let  $D(\lambda)$  be a quadratic matrix whose elements are continuous-differentiable functions in  $\lambda$ . Let  $P(\lambda)$  be the characteristic polynomial of  $D(\lambda)$ . Then, we have for all  $\lambda$  with  $P(\lambda) \neq 0$ :

$$P'(\lambda) = \text{Tr}(D^{-1}(\lambda) \cdot D'(\lambda)) \cdot P(\lambda)$$

$\text{Tr}$  is the trace of a matrix and  $D'(\lambda)$  the matrix of derivatives of the elements of  $D(\lambda)$  with respect to  $\lambda$ .

Using this theorem we get:

$$\begin{aligned}
 \frac{\partial}{\partial s} |J| &= \text{Tr}\left(J^{-1} \cdot \frac{\partial}{\partial s} J\right) \cdot |J| \\
 &= -\text{Tr}(J^{-1} \cdot H(\rho) \cdot J) \cdot |J| = -\text{Tr}(H(\rho)) \cdot |J| \\
 &= -\nabla^2 \rho \cdot |J|
 \end{aligned} \tag{14}$$

with initial values:

$$|J|(s=0) = |J|(h^{-1}(0, \theta, \varphi)) \tag{15}$$

To evaluate the function under the second integral sign in eq. (10) at a point  $\mathbf{x}$ , we must solve a system of four differential equations [eqs. (4) and (14)].

### INTEGRATION AS INITIAL VALUE PROBLEMS

The parts of eq. (10) can themselves be written as three nested systems of differential equations with  $F_\beta = F_\beta(\varphi = 2\pi)$  and  $F_\Omega = F_\Omega(\varphi = 2\pi)$ . The integration inside the  $\beta$ -sphere transforms into:

$$\begin{aligned}
 \frac{\partial}{\partial \varphi} F_\beta(\varphi) &= F_{\beta\varphi}(\theta = \pi, r = \beta) \\
 &\equiv \int_0^\pi \sin(\theta) \int_0^\beta r^2 \cdot f(\mathbf{x}(r, \theta, \varphi)) dr d\theta, \\
 F_\beta(0) &= 0
 \end{aligned} \tag{16}$$

$$\begin{aligned}
 \frac{\partial}{\partial \theta} F_{\beta\varphi}(\theta, \beta) &= F_{\beta\varphi\theta}(r = \beta) \\
 &\equiv \sin(\theta) \int_0^\beta r^2 \cdot f(\mathbf{x}(r, \theta, \varphi)) dr, \\
 F_{\beta\varphi}(0) &= 0
 \end{aligned} \tag{17}$$

$$\begin{aligned}
 \frac{\partial}{\partial r} F_{\beta\varphi\theta}(r) &= r^2 \cdot f(\mathbf{x}(r, \theta, \varphi)) \\
 F_{\beta\varphi\theta}(0) &= 0
 \end{aligned} \tag{18}$$

The calculation of  $F_\Omega$  [with the use of eqs. (4) and (14)] transforms into:

$$\begin{aligned}
 \frac{\partial}{\partial \varphi} F_\Omega(\varphi) &= F_\varphi(\theta = \pi, s = \infty) \\
 &\equiv \int_0^\pi \int_0^\infty f(\mathbf{x}(s, \theta, \varphi)) \\
 &\quad \cdot |J(h^{-1}(s, \theta, \varphi))| ds d\theta, \\
 F_\Omega(0) &= 0
 \end{aligned} \tag{19}$$

$$\begin{aligned}
 \frac{\partial}{\partial \theta} F_\varphi(\theta, \beta) &= F_{\varphi\theta}(s = \infty) \\
 &\equiv \int_0^\infty f(\mathbf{x}(s, \theta, \varphi)) \cdot |J(h^{-1}(s, \theta, \varphi))| ds, \\
 F_\varphi(0) &= 0
 \end{aligned} \tag{20}$$

$$\begin{aligned}
 \frac{\partial}{\partial s} F_{\varphi\theta}(s) &= f(\mathbf{x}(s, \theta, \varphi)) \cdot |J(h^{-1}(s, \theta, \varphi))|, \\
 F_{\varphi\theta}(0) &= 0
 \end{aligned} \tag{21a}$$

$$\frac{\partial}{\partial s}|J| = -\nabla^2\varphi \cdot |J|, \quad |J|(0) = |J(h^{-1}(0, \theta\varphi))| \quad (21b)$$

$$\frac{\partial \mathbf{x}(s)}{\partial s} = -\nabla\varphi(\mathbf{x}(s), \mathbf{X}), \quad \mathbf{x}(0) = \mathbf{x}_{\theta\varphi} \quad (21c)$$

Eqs. (21) form a coupled system of five differential equations, which must be solved simultaneously. If  $n$  properties are calculated, eqs. (21) consists out of  $n + 4$  coupled equations.

### CVODE FOR SOLUTION OF INITIAL VALUE PROBLEMS

In the program package AIM2000,<sup>9</sup> all integrations are formulated as initial value problems [as in eqs. (16)–(21)] and are solved using the routines from the CVODE package of Cohen and Hindmarsh.<sup>7</sup>

CVODE uses multistep methods of the Adams type and employs an automatic step-size control. Multistep methods are preferred over single-step methods (e.g., Runge–Kutta), because the number of evaluations of the right-hand side of the differential equations is much smaller. These evaluations can be very expensive in terms of computer time: for one evaluation of the right-hand side of eq. (19) an initial value problem of the type in eq. (20) has to be solved. Each evaluation of the right-hand side of eq. (20) involves the solution of a system like eqs. (21).

Another very important aspect of CVODE is its step-size control. In atomic basins we find regions where gradient paths are nearly radial and where a large step size is sufficient. Close to critical points, on the other hand, gradient paths have a high curvature and an uneven distribution. A much smaller step size is necessary here. The lack of an efficient step-size control mechanism was one of the main reasons (beside singularities, see later) why this approach was previously abandoned.

### SINGULARITIES AT BOND CRITICAL POINTS

If eqs. (21) are solved along a gradient path starting at a critical point, it is frequently necessary to integrate over a singularity. Consider the formal solution of initial value problem of eq. (14)/(21):

$$|J|(s) = |J|(0) \cdot e^{-\int_0^s \nabla^2\varphi dt} \quad (22)$$

We get for the limit of the Jacobian determinant:

$$\lim_{s \rightarrow \infty} |J(s)| = \infty, \quad \text{if } |J|(0) \neq 0 \text{ and } \lim_{s \rightarrow \infty} \nabla^2\varphi < 0 \quad (23)$$

If a gradient path starts at infinity, these problems do not occur because all functions considered decay exponentially. If, on the other hand, a gradient path starts at a critical point, especially a bond critical point with two negative and one positive eigenvalue in the Hessian matrix, eqs. (19) and (20) must integrate over a singularity.

Although the value of the integral is defined and finite, integrating over a singularity requires many function evaluations, making the computational effort prohibitive. This effect can be avoided if  $|J|(0) = 0$  is used for the bond path. This is the case if the bond path cuts the  $\beta$ -sphere at the north pole of the coordinate system ( $\theta = \varphi = 0$ ).

To use this property, we define for each bond critical point a polar coordinate system with a north pole on the bond path, which develops out of the atomic coordinate system by rotation. Let:

$$\beta \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{with} \quad \left\| \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\|_2 = 1$$

be the point of intersection of a bond path with the  $\beta$ -sphere. The matrix  $R$  transforms this point into  $\begin{pmatrix} 0 \\ 0 \\ \beta \end{pmatrix}$ :

$$R = \begin{pmatrix} -\frac{xz}{\sqrt{x^2+y^2}} & -\frac{yz}{\sqrt{x^2+y^2}} & \sqrt{x^2+y^2} \\ \frac{y}{\sqrt{x^2+y^2}} & -\frac{x}{\sqrt{x^2+y^2}} & 0 \\ x & y & z \end{pmatrix} \quad (24)$$

$R$  has the properties:

$$R \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad R^T \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad |R| = \det(R) = 1 \quad (25)$$

If integration takes place in the rotated system, no singularity appears along the bond path. In the case of  $m$  bond paths in the atomic basin, the surface of the  $\beta$ -sphere has to be partitioned into  $m$  parts. Each part contains a bond path whose intersection with the  $\beta$ -sphere is the north pole of the corresponding rotated coordinate system. We obtain the solution as a sum:

$$F(\Omega) = F_\beta + F_\Omega \\ = F_\beta + \sum_{i=1}^m \int_0^{2\pi} \int_0^{\theta_i(\varphi)} \int_0^\infty f(\mathbf{x}(s, 0, \varphi)) \cdot |J| ds d\theta d\varphi \quad (26)$$

The function  $\theta_i(\varphi)$  reflects the partitioning of the  $\beta$ -sphere. For  $m = 2$ ,  $\theta_1(\varphi)$  can be calculated in the following way. Consider a coordinate system with north pole on a bond path. The bond path intersects  $S_\beta$  at

$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ 0 \\ \beta \end{pmatrix}$$

Another bond path intersects the  $\beta$ -sphere at:

$$\mathbf{x}_2 = \beta \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{with} \quad \left\| \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\|_2 = 1$$

For given  $\varphi$  we integrate  $\theta$  from 0 to  $\theta_1(\varphi)$ .  $\theta_1(\varphi)$  corresponds to a point on the  $\beta$ -sphere with coordinates  $(\beta, \theta_1(\varphi), \varphi)$ , which is equidistant to  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . For  $\theta_1(\varphi)$ , we get the formula:

$$\begin{aligned} & (x - \sin(\theta_1(\varphi)) \cdot \cos(\varphi))^2 \\ & + (y - \sin(\theta_1(\varphi)) \cdot \sin(\varphi))^2 + (z - \cos(\theta_1(\varphi)))^2 \\ & = \sin^2(\theta_1(\varphi)) \cdot \cos^2(\varphi) + \sin^2(\theta_1(\varphi)) \cdot \sin^2(\varphi) \\ & + (\cos(\theta_1(\varphi)) - 1)^2 \end{aligned} \quad (27)$$

As solution for  $\theta_1(\varphi)$ , we get:

$$\theta_1(\varphi) = \arctan\left(\frac{1 - z}{x \cdot \cos(\varphi) + y \cdot \sin(\varphi)}\right) \quad (28)$$

For more than two bond paths, eq. (28) must be calculated for all bonds.  $\theta_1(\varphi)$  is the minimum of all calculated values.

These formalisms allow integration over atomic basins without integrating over singularities. Of course, gradient paths close to bond paths are sometimes extremely curved. The step-size control of CVODE usually inserts many paths in the vicinity of  $\theta = 0$ , while using larger steps for larger values of  $\theta$ . In fact, most computer time is used in the vicinity of bond paths.

No problems appear, if interatomic surfaces contain ring or cage critical points (most other methods in literature run into difficulties in these cases).

### CHECK OF INTEGRATION ACCURACY

An important aspect of numerical calculations is the *a posteriori* examination of accuracy. CVODE gets the desired relative and absolute accuracy as input parameters. Step-size is controlled such that every *single* step satisfies these conditions. However, to check the global accuracy, these values are not suitable.

Using the Gauss theorem, a relation between volume and surface integrals can be proven, which

enables us to establish a test of the global accuracy of integration results.

Let  $\mathbf{u}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a continuous differentiable and bounded vector field, and  $G \in \mathbb{R}^3$  a region enclosed by a surface  $S$ . The following relation then holds:

$$\int_G \operatorname{div} \mathbf{u}(\mathbf{x}) d\mathbf{x} = \int_S \mathbf{u}(\mathbf{x})^T \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x}$$

where  $\mathbf{n}$  is the outward pointing normal to the surface and  $\operatorname{div} \mathbf{u} = \frac{\partial}{\partial x} u_x + \frac{\partial}{\partial y} u_y + \frac{\partial}{\partial z} u_z$ . Introducing the basin  $\Omega$  for  $G$ , the vector field  $-\frac{1}{4}\nabla\rho$ , and the function:

$$l = -\frac{1}{4}\nabla^2\rho \quad (29)$$

with the Gauss theorem we get:

$$\begin{aligned} \int_\Omega l(\mathbf{x}) d\mathbf{x} &= -\frac{1}{4} \int_\Omega \nabla^2 \rho(\mathbf{x}) d\mathbf{x} \\ &= -\frac{1}{4} \int_S \nabla \rho(\mathbf{x})^T \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (30)$$

Here  $S$  is the total surface enclosing  $\Omega$ . The function under the surface integral vanishes everywhere because of eq (1). Hence:

$$L(\Omega) = \int_\Omega l(\mathbf{x}) d\mathbf{x} = 0 \quad (31)$$

The value for  $L(\Omega)$  is a good test for global accuracy in the volume integration.  $l(\mathbf{x})$  is a difficult function to integrate because of its large fluctuations.

## Integration Over Atomic Basins—Examples

In general, every integrable function  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$  can be integrated over atomic basins. From a numerical point of view, the following functions are of special interest because they allow an assessment of integration accuracy [AIM2000 contains a much larger set of functions]:

- $N_A = \int_\Omega \rho(\mathbf{x}) d\mathbf{x}$ —the number of electrons in basin  $\Omega$  of atom  $A$ ;
- $L_A = \int_\Omega l(\mathbf{x}) d\mathbf{x}$ —the integral over the Laplace density;
- $G_A = \int_\Omega g(\mathbf{x}) d\mathbf{x}$ —the integral over the Lagrange kinetic energy density;
- $K_A = \int_\Omega k(\mathbf{x}) d\mathbf{x}$ —the integral over the Hamiltonian kinetic energy density;
- $V_A^k = \int_\Omega \|\mathbf{x} - \mathbf{X}\|^k \cdot ((k+3) \cdot \rho + (\mathbf{x} - \mathbf{X})^T \cdot \nabla \rho) d\mathbf{x}$ —the integral over the hypervirial functions of degree  $k = -1, 0, 1, 2$ .  $\mathbf{X}$  is the nuclear position of atom  $A$ .



In addition to eq. (31), we have the relationships:

$$N = \sum_A N_A \quad \text{is the total number of electrons} \\ \text{in the molecule} \quad (32)$$

$$l(\mathbf{x}) = g(\mathbf{x}) - k(\mathbf{x}), \quad \text{and from here with eq. (31):}$$

$$G_A = K_A \quad \text{for all atoms } A$$

The hypervirial integrals are needed as a check for the surface integration (see next section). For all calculations presented the following values are pre-set:

- Requested absolute accuracy:  $10^{-6}$ .
- Requested relative accuracy:  $10^{-6}$ .
- Gradient paths will be computed up to  $s = 10^6$ . This corresponds to path lengths of approximately 6–8 atomic units for paths starting at infinity.
- The radius of the  $\beta$ -sphere is 0.5 atomic units, and for hydrogen atoms 0.3 atomic units.

With these preset values the results are usually accurate to five significant digits. The computer time needed is heavily dependent on the requested accuracies. By reducing these values to  $10^{-5}$ , the needed computer time decreases by 60% to 80%.

As a measure of program speed we report the number of wave function evaluations and the computer time required on a PC with a Pentium-3 processor and a frequency of 550 MHz.

### HCN MOLECULE

The HCN molecule is linear, and hence we can use rotational symmetry. The volume integration reduces to a single  $\varphi$ -plane. The charge density is very flat near the bond critical point between C and H. Step-size control is therefore needed. The wave

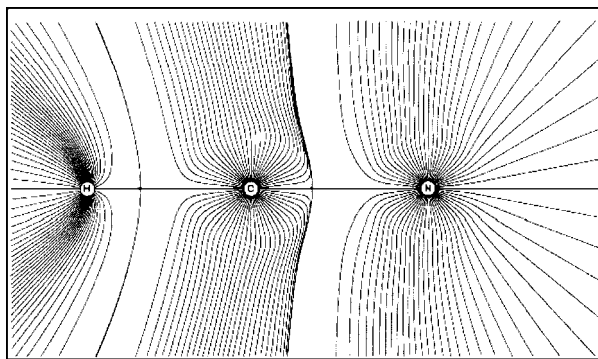


FIGURE 1. Gradient plot of HCN.

function type used was 6-31G\* at 6-31G\* geometry and computed with GAUSSIAN.<sup>8</sup>

Results for H:  $N_H = 0.809745$ ;  $L_H = -0.000005$ ;  $G_H = 0.534193$ ;  $K_H = 0.534188$ ;  $V_H^{-1} = 0.661124$ ;  $V_H^0 = 0.699014$ ;  $V_H^1 = 0.846039$ ;  $V_H^2 = 1.218209$ ; number of evaluations: 48,222; CPU time: 7.7 seconds.

Results for C:  $N_C = 4.835524$ ;  $L_C = -0.000052$ ;  $G_C = 37.159797$ ;  $K_C = 37.159745$ ;  $V_C^{-1} = 1.931136$ ;  $V_C^0 = 2.518463$ ;  $V_C^1 = 3.768075$ ;  $V_C^2 = 6.502861$ ; number of evaluations: 76,494; CPU time: 12.6 seconds.

Results for N:  $N_N = 8.354475$ ;  $L_N = -0.000082$ ;  $G_N = 55.173112$ ;  $K_N = 55.173030$ ;  $V_N^{-1} = 1.520931$ ;  $V_N^0 = 2.896940$ ;  $V_N^1 = 5.826199$ ;  $V_N^2 = 12.538411$ ; number of evaluations: 40,230; CPU time: 7.0 seconds.

Total number of electrons in the molecule:  $N_H + N_C + N_N = 13.999744$ .

### C<sub>4</sub>H<sub>4</sub> MOLECULE

This molecule consists of four equivalent C atoms forming a tetrahedron. Each C atom is bonded to the other three C atoms and to one H atom. There are ring critical points on the tetrahedron surfaces and a cage critical point inside. The interatomic surfaces thus contain rings and cages. Again, we can use symmetry for the integration. Both atoms are integrated in  $\varphi$  from  $0^\circ$  to  $120^\circ$ , and the results are multiplied by three. The wave function used was 6-31G\* at 6-31G\* geometry, and computed with GAUSSIAN.<sup>8</sup>

Results for H:  $N_H = 0.889107$ ;  $L_H = -0.000004$ ;  $G_H = 0.562803$ ;  $K_H = 0.562799$ ;  $V_H^{-1} = 0.690737$ ;  $V_H^0 = 0.766146$ ;  $V_H^1 = 0.963769$ ;  $V_H^2 =$

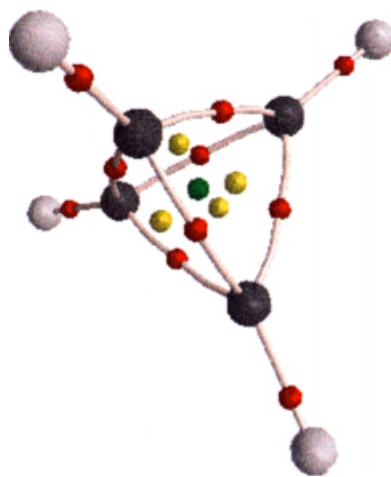


FIGURE 2. C<sub>4</sub>H<sub>4</sub> with critical points and bond paths.

1.416846; number of evaluations: 1,406,491; CPU time: 464.5 seconds.

Results for C:  $N_C = 6.110839$ ;  $L_C = -0.000045$ ;  $G_C = 37.815977$ ;  $K_C = 37.815931$ ;  $V_C^{-1} = 2.920664$ ;  $V_C^0 = 4.989121$ ,  $V_C^1 = 8.923690$ ;  $V_C^2 = 16.966577$ ; number of evaluations: 11,504,248; CPU time: 3832.5 seconds.

Total number of electrons in the molecule:  $4 \cdot (N_H + N_C) = 27.999784$ .

## Integration Over Interatomic Surfaces

### DEFINITION OF INTERATOMIC SURFACES WITH GRADIENT PATHS

Similar to atomic basins we can use gradient paths to define interatomic surfaces.

Let the interatomic surface,  $S_{AB}$ , between atoms  $A$  and  $B$  contain the  $(3, -1)$ -critical point  $\mathbf{x}_c$ . Let  $\lambda_1 \leq \lambda_2 \leq \lambda_3$  be the eigenvalues of  $H(\rho(\mathbf{x}_c))$  and  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  are the corresponding eigenvectors. Starting at  $\mathbf{x}_c$ ,  $\mathbf{v}_3$  defines the direction of the bond path while the linear combinations of  $\mathbf{v}_1$  and  $\mathbf{v}_2$  define the initial directions of all gradient paths in the interatomic surface. To approximate, for small  $\beta > 0$ , the points:

$$\mathbf{x}_{\beta\varphi} = \mathbf{x}_c + \beta \cdot (\mathbf{v}_1 \cdot \cos(\varphi) + \mathbf{v}_2 \cdot \sin(\varphi)) \quad (33)$$

are also located in  $S_{AB}$ . The interatomic surface can then be approximated by:

$$S_{AB} \simeq \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \in g(\mathbf{x}_{\beta\varphi}), \varphi \in [0, 2\pi]\} \quad (34)$$

with

$$g(\mathbf{x}_{\beta\varphi}) = \left\{ \mathbf{x}(s) \in \mathbb{R}^3 \mid \mathbf{x}(s) = \mathbf{x}_{\beta\varphi} - \int_0^s \nabla \varrho(\mathbf{x}(t)) dt \right. \\ \left. \text{and } \mathbf{x}(0) = \mathbf{x}_{\beta\varphi} \right\}$$

### INTEGRATION IN NATURAL COORDINATES

With  $(s, \varphi)$ , a natural coordinate system in the interatomic surface is defined. This coordinate system can be used for integration:

$$F(S_{AB}) = \int_{S_{AB}} f(\mathbf{x}) d\mathbf{x} \\ = \int_0^{2\pi} \int_{-\infty}^{\infty} f(\mathbf{x}(s, \varphi)) \cdot \sqrt{|J^T \cdot J|} ds d\varphi \quad (35)$$

Again:

$$J = \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial \varphi} \end{pmatrix} = (\mathbf{x}_s | \mathbf{x}_\varphi) \quad (36)$$

is the Jacobian matrix of the coordinate transformation. Analogous to eqs. (11) and (12) this matrix can be determined with an initial value problem:

$$\frac{\partial}{\partial s} \mathbf{x}_s = -\frac{\partial}{\partial s} \nabla \rho = -H(\rho) \cdot \mathbf{x}_s \quad (37a)$$

$$\frac{\partial}{\partial s} \mathbf{x}_\varphi = \frac{\partial}{\partial \varphi} \mathbf{x}_s = -\frac{\partial}{\partial \varphi} \nabla \rho = -H(\rho) \cdot \mathbf{x}_\varphi \quad (37b)$$

Together:

$$\frac{\partial}{\partial s} J = -H(\rho) \cdot J \quad (38)$$

with starting values:

$$J_0 = (-\nabla \rho(\mathbf{x}_{\beta\varphi}) | \beta \cdot (\mathbf{v}_2 \cdot \cos(\varphi) - \mathbf{v}_1 \cdot \sin(\varphi))) \quad (39)$$

Because  $J$  is a nonquadratic matrix, the trace theorem is not applicable. The first column of  $J$  is determined by  $\nabla \rho$ , which leaves a system of three differential equations for the second column.

Inside the  $\beta$ -circle we assume that all functions are constant:

$$F_S \simeq f(\mathbf{x}_c) \cdot \pi \cdot \beta^2 + \int_0^{2\pi} \int_0^\infty f(\mathbf{x}(s, \varphi)) \cdot \sqrt{|J^T \cdot J|} ds d\varphi \quad (40)$$

### SURFACE INTEGRATION AS INITIAL VALUE PROBLEM

Again, the calculation of  $F_S = F_S(\varphi = 2\pi)$  can be written as nested initial value problems:

$$\frac{\partial}{\partial \varphi} F_S(\varphi) = F_{S\varphi}(s = \infty) \\ \equiv \int_0^\infty f(\mathbf{x}(s, \varphi)) \cdot \sqrt{|J^T \cdot J|} ds, \\ F_S(0) = f(\mathbf{x}_c) \cdot \pi \cdot \beta^2 \quad (41)$$

$$\frac{\partial}{\partial s} F_{S\varphi}(s) = f(\mathbf{x}(s, \varphi)) \cdot \sqrt{|J^T \cdot J|}, \\ F_{S\varphi}(0) = 0 \quad (42a)$$

$$\frac{\partial}{\partial s} \mathbf{x}_\varphi = -H(\rho(\mathbf{x}(s, \varphi))) \cdot \mathbf{x}_\varphi, \\ \mathbf{x}_\varphi(0) = \beta \cdot (\mathbf{v}_2 \cdot \cos(\varphi) - \mathbf{v}_1 \cdot \sin(\varphi)) \quad (42b)$$

$$\frac{\partial}{\partial s} \mathbf{x}(s) = -\nabla \varrho(\mathbf{x}(s, \varphi)), \\ \mathbf{x}(0) = \mathbf{x}_c + \beta \cdot (\mathbf{v}_1 \cdot \cos(\varphi) + \mathbf{v}_2 \cdot \sin(\varphi)) \quad (42c)$$

Each evaluation of the right-hand side of eq. (41) needs the solution of an initial value problem of type (42). If  $n$  different functions are being integrated in eq. (42a), then eqs. (42) consists of  $(n + 6)$ -coupled differential equations.

## NORMAL SURFACE

Many functions that are integrated over interatomic surfaces contain the surface normal  $\mathbf{n}(\mathbf{x})$ , which should satisfy some relations:

$$\mathbf{x}_s^T \cdot \mathbf{n}(\mathbf{x}) = -\nabla \rho(\mathbf{x})^T \cdot \mathbf{n}(\mathbf{x}) = 0 \quad (43a)$$

$$\mathbf{x}_\phi^T \cdot \mathbf{n}(\mathbf{x}) = 0 \quad (43b)$$

$$\mathbf{n}(\mathbf{x})^T \cdot \mathbf{n}(\mathbf{x}) = 1 \quad (43c)$$

$$\mathbf{n}(\mathbf{x}_{old})^T \cdot \mathbf{n}(\mathbf{x}) > 0 \quad (43d)$$

where eqs. (43a) and (43b) define the main property of a normal—that is, it is perpendicular to the surface. Eq. (43c) is the normalization; eq. (43d) is necessary to ensure that the orientation of the normal remains the same along a gradient path.  $\mathbf{x}_{old}$  is the previously computed point on the path. At the beginning of the integration the orientation of the normal is defined as:

$$\mathbf{n}(\mathbf{x}_c) = \pm \mathbf{v}_3$$

The calculation of the normal for a point  $\mathbf{x}$  on the interatomic surface contains two steps. First, we solve a linear system of three equations:

$$\begin{pmatrix} \mathbf{x}_s^T \\ \mathbf{x}_\phi^T \\ \mathbf{n}(\mathbf{x}_{old})^T \end{pmatrix} \cdot \tilde{\mathbf{n}}(\mathbf{x}) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (44)$$

We then normalize  $\tilde{\mathbf{n}}(\mathbf{x})$ :

$$\mathbf{n}(\mathbf{x}) = \frac{\tilde{\mathbf{n}}(\mathbf{x})}{\|\tilde{\mathbf{n}}(\mathbf{x})\|_2} \quad (45)$$

## CHECK OF INTEGRATION ACCURACY

Because  $\nabla \rho^T \cdot \mathbf{n}$  vanishes at every point on the surface, the integral over this function is not suitable as a global accuracy check. Again, the Gauss theorem provides a more efficient estimate of accuracy.

Consider the hypervirial vector field of degree  $k$ :

$$\text{vir}_A^k : \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad \text{vir}_A^k(\mathbf{x}) = \|\mathbf{x} - \mathbf{X}\|^k \cdot \rho \cdot (\mathbf{x} - \mathbf{X}) \quad (46)$$

with nuclear position  $\mathbf{X}$  of atom  $A$ . The divergence of this vector field is:

$$\text{div } \text{vir}_A^k(\mathbf{x}) = \|\mathbf{x} - \mathbf{X}\|^k \cdot ((k+3) \cdot \rho + (\mathbf{x} - \mathbf{X})^T \cdot \nabla \rho) \quad (47)$$

Hence, we get:

$$\begin{aligned} V_A^k &= \int_{\Omega} \|\mathbf{x} - \mathbf{X}\|^k \cdot ((k+3) \cdot \rho + (\mathbf{x} - \mathbf{X})^T \cdot \nabla \rho) d\mathbf{x} \\ &= \int_S \|\mathbf{x} - \mathbf{X}\|^k \cdot \rho \cdot (\mathbf{x} - \mathbf{X}) \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (48)$$

This relationship with  $k \in \{-1, 0, 1, 2\}$  is an efficient estimate of the accuracy for volume and surface integrals.

## Integration Over Interatomic Surfaces—Examples

The following results are reported for the interatomic surface,  $S_{AB}$ , between atoms  $A$  and  $B$  (AIM2000 contains a larger number of functions):

- $N_{AB} = \int_{S_{AB}} \rho(\mathbf{x}) d\mathbf{x}$ —the integral over the charge density;
- $L_{AB} = \int_{S_{AB}} l(\mathbf{x}) d\mathbf{x}$ —the integral over the Laplace density;
- $G_{AB} = \int_{S_{AB}} g(\mathbf{x}) d\mathbf{x}$ —the integral over the Lagrange kinetic energy density;
- $K_{AB} = \int_{S_{AB}} k(\mathbf{x}) d\mathbf{x}$ —the integral over the Hamiltonian kinetic energy density;
- $GRN_{AB} = \int_{S_{AB}} \nabla \rho^T(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x}$ ;
- $V_{AB}^k = \int_{S_{AB}} \|\mathbf{x} - \mathbf{X}\|^k \cdot \rho \cdot (\mathbf{x} - \mathbf{X}) \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x}$ ;
- $V_{BA}^k = \int_{S_{BA}} \|\mathbf{x} - \mathbf{X}\|^k \cdot \rho \cdot (\mathbf{x} - \mathbf{X}) \cdot \mathbf{n}(\mathbf{x}) d\mathbf{x}$ —the integrals over the hypervirial terms of degree  $k = -1, 0, 1, 2$  for atoms  $A$  and  $B$ , respectively.

Further relationships:

$$\nabla \rho^T(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in S_{AB}$$

hence  $GRN_{AB} = 0$ , and:

$$V_A^k = \sum_B V_{AB}^k$$

The sum runs over all atoms bonded to atom  $A$  [see eq. (48)], and:

$$V_B^k = \sum_A V_{BA}^k$$

by analogy. The same preset values as for the volume integration (see earlier) have been used. In addition, the radius of the  $\beta$ -circle around the bond critical point is set to 0.001 atomic units.

## HCN MOLECULE

Because of rotational symmetry, the integration reduces to a single gradient path.

Results for the interatomic surface between  $H$  and  $C$  are:  $N_{HC} = 0.915528$ ;  $L_{HC} = -0.033343$ ;  $G_{HC} = 0.460394$ ;  $K_{HC} = 0.427048$ ;  $GRN_{HC} = 0.000000$ ;  $V_{HC}^{-1} = 0.661134$ ;  $V_{HC}^0 = 0.698840$ ;  $V_{HC}^1 = 0.844760$ ;  $V_{HC}^2 = 1.210236$ ;  $V_{CH}^{-1} = 0.626417$ ;  $V_{CH}^0 = 1.063377$ ;  $V_{CH}^1 = 1.890534$ ;  $V_{CH}^2 = 3.561976$ ; number of evaluations: 325; CPU time: 0.2 seconds.

Results for the interatomic surface between  $N$  and  $C$  are:  $N_{NC} = 2.079225$ ;  $L_{NC} = -0.063289$ ;  $G_{NC} =$



2.125198;  $K_{NC} = 2.061912$ ;  $GRN_{NC} = 0.000000$ ;  
 $V_{NC}^{-1} = 1.520830$ ;  $V_{NC}^0 = 2.896015$ ;  $V_{NC}^1 = 5.818946$ ;  
 $V_{NC}^2 = 12.484208$ ;  $V_{CN}^{-1} = 1.304693$ ;  $V_{CN}^0 = 1.454703$ ;  
 $V_{CN}^1 = 1.873922$ ;  $V_{CN}^2 = 2.912303$ ; number of evaluations: 345; CPU time: 0.2 seconds.

Estimates of global accuracy according to the Gauss theorem are:  $V_H^{-1} - V_{HC}^{-1} = -0.000010$ ;  $V_C^{-1} - V_{CH}^{-1} - V_{CN}^{-1} = 0.000026$ ;  $V_N^{-1} - V_{NC}^{-1} = 0.000101$ ;  
 $V_H^0 - V_{HC}^0 = 0.000174$ ;  $V_C^0 - V_{CH}^0 - V_{CN}^0 = 0.000383$ ;  
 $V_N^0 - V_{NC}^0 = 0.000925$ ;  $V_H^1 - V_{HC}^1 = 0.001279$ ;  
 $V_C^1 - V_{CH}^1 - V_{CN}^1 = 0.003619$ ;  $V_N^1 - V_{NC}^1 = 0.007253$ ;  
 $V_H^2 - V_{HC}^2 = 0.007973$ ;  $V_C^2 - V_{CH}^2 - V_{CN}^2 = 0.028582$ ;  
 $V_N^2 - V_{NC}^2 = 0.054203$ .

### C<sub>4</sub>H<sub>4</sub> MOLECULE

Both interatomic surfaces are integrated without use of symmetry.

Results for the interatomic surface between H and C are:  $N_{HC} = 0.932348$ ;  $L_{HC} = -0.108307$ ;  $G_{HC} = 0.483587$ ;  $K_{HC} = 0.375279$ ;  $GRN_{HC} = 0.000000$ ;  
 $V_{HC}^{-1} = 0.690779$ ;  $V_{HC}^0 = 0.766187$ ;  $V_{HC}^1 = 0.963560$ ;  
 $V_{HC}^2 = 1.415034$ ;  $V_{CH}^{-1} = 0.605643$ ;  $V_{CH}^0 = 1.021745$ ;  
 $V_{CH}^1 = 1.813801$ ;  $V_{CH}^2 = 3.434086$ ; number of evaluations: 23,636; CPU time: 16.8 seconds.

Results for the interatomic surface between two C atoms:  $N_{CC} = 0.956428$ ;  $L_{CC} = -0.196636$ ;  $G_{CC} = 0.669237$ ;  $K_{CC} = 0.472600$ ;  $GRN_{CC} = 0.000000$ ;  
 $V_{CC}^{-1} = 0.771714$ ;  $V_{CC}^0 = 1.322507$ ;  $V_{CC}^1 = 2.369971$ ;  
 $V_{CC}^2 = 4.510445$ ; number of evaluations: 134,125; CPU time: 92.6 seconds.

Estimated of global accuracy according to the Gauss theorem are:  $V_H^{-1} - V_{HC}^{-1} = -0.000042$ ;  $V_C^{-1} - V_{CH}^{-1} - 3 \cdot V_{CC}^{-1} = -0.000121$ ;  $V_H^0 - V_{HC}^0 = -0.000041$ ;  
 $V_C^0 - V_{CH}^0 - 3 \cdot V_{CC}^0 = -0.000154$ ;  $V_H^1 - V_{HC}^1 = 0.000209$ ;

$$V_C^1 - V_{CH}^1 - 3 \cdot V_{CC}^1 = -0.000024; V_H^2 - V_{HC}^2 = 0.001812; V_C^2 - V_{CH}^2 - 3 \cdot V_{CC}^2 = -0.001156.$$

## Conclusion

In this study, we have presented mathematical formalisms that allow accurate and fast calculations of average properties of atoms and interatomic surfaces in molecules. The proposed methods are part of the AIM2000 program, which consists of many procedures to deal with atoms in molecules in a setting with a comfortable user interface. AIM2000 will be described in a forthcoming study.

It is hoped that these reliable methods in an easy-to-use program will lead to increased use of the theory of atoms in molecules.

## References

1. Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, UK, 1990.
2. Biegler-König, F.; Bader, R. F. W.; Tang, T. H. *J Comput Chem* 1982, 13, 317–328.
3. Cioslowski, J.; Stefanov, B. B. *Molec Phys* 1995, 84, 707–716.
4. Stefanov, B. B.; Cioslowski, J. *J Comput Chem* 1995, 16, 1394–1404.
5. Biegler-König, F.; Nguyen-Dang, T. T.; Tal, Y.; Bader, R. F. W. *J Phys B* 1981, 14, 2739–2751.
6. Davidenko, D. F. *Sov J Math* 1960, 1, 316.
7. Cohen, S. D.; Hindmarsh, A. C. *Comput Phys* 1996, 10 (to download: <ftp://netlib2.cs.utk.edu>).
8. <http://www.gaussian.com>.
9. <http://gauss.fh-bielefeld.de/aim2000>.