Fully Analytical Integration Over the 3D Volume Bounded by the β Sphere in Topological Atoms

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · NOVEMBER 2011

Impact Factor: 2.69 · DOI: 10.1021/jp206223g · Source: PubMed

1 AUTHOR:

Paul L A Popelier
The University of Manchester
190 PUBLICATIONS **7,101** CITATIONS

SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY A

ARTICLE

pubs.acs.org/JPCA

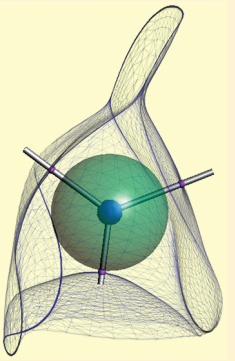
Fully Analytical Integration Over the 3D Volume Bounded by the β Sphere in Topological Atoms

Paul L. A. Popelier

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, University of Manchester, Manchester M1 7DN, Great Britain, and School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, Great Britain

Supporting Information

ABSTRACT: Atomic properties of a topological atom are obtained by 3D integration over the volume of its atomic basin. Algorithms that compute atomic properties typically integrate over two subspaces: the volume bounded by the so-called β sphere, which is centered at the nucleus and completely contained within the atomic basin, and the volume of the remaining part of the basin. Here we show how the usual quadrature over the β sphere volume can be replaced by a fully analytical 3D integration leading to the atomic charge (monopole moment) for s, p, and d functions. Spherical tensor multipole moments have also been implemented and tested up to hexadecupole for s functions only, and up to quadrupole for s and p functions. The new algorithm is illustrated by operating on capped glycine (HF/6-31G, 35 molecular orbitals (MOs), 322 Gaussian primitives, 19 nuclei), the protein crambin (HF/3-21G, 1260 MOs, 5922 primitives and 642 nuclei), and tin (Z = 50) in $Sn_2(CH_3)_2$ (B3LYP/cc-pVTZ and LANL2DZ, 59 MOs, 1352 primitives).



1. INTRODUCTION

The concept of a virial fragment matured throughout the 1970s and culminated into what was coined Quantum Topology[1] in 1979. In that year, the electron density ρ was for the first time described by the language of dynamical systems theory, with early references to catastrophe theory. Terms such as critical points and their classification by means of the Hessian joined the glossary of what then became the "Quantum Theory of Atoms in Molecules" (QTAIM).[2] The concept of an atomic basin, although not yet called so in the 1970s, preceded that of critical points. The bond critical point would later be interpreted by many as a conveniently simple signature of the computability of a bond, a view subject to ongoing debate (e.g., refs 3 and 4 and references therein). In any event, the topological notion of an atom was enunciated before that of a chemical bond, in the early QTAIM sense, through the idea of a bond path.[5] This order of discovery events can be explained by realizing that the roots of QTAIM lie in the quest for spatially defined molecular fragments that are transferable between different systems.[6] Already in 1972, a portion of space bounded by so-called surfaces of zero-flux, was proposed[7] for that purpose. The same virial relationship between the average kinetic and potential energies observed for a *total* molecular system was found to also hold for the topological molecular fragments, that is, QTAIM atoms.

It is fascinating to witness the surge of QTAIM from the initially false attempt of "natural partitioning"[8] (by means of simple planes through points of minimum electron density along the internuclear axis) all the way to a subspace quantum theory.[9] Although regularly used by dozens of laboratories worldwide, one could argue that the uptake of QTAIM has been relatively slow since its inception, about thirty years ago. Perhaps this is partially due to its pronounced novelty, originality, and unique character. A more important reason is most likely the enormous challenge of integration of property densities over topological basins, both in terms of algorithmic robustness and complexity and in terms of CPU time. Detailed reports[10,11] surveying the approximately 150−200 QTAIM papers appearing annually from 1996 to 2001, showed that most applications were confined to a (bond) critical point or bond path analysis. Atomic integrations were far rarer, surely due to the algorithmic challenge mentioned above. Now, more than a decade later, computers have become more than 250 times faster. This development, combined with improved algorithms, makes the rapid calculation of atomic properties much more feasible for molecules of up to 40 atoms or so. Far beyond that rough limit, topologies can become very complex, especially in the presence of large rings. Using Condor high throughout computing at the University of Manchester, this laboratory routinely integrates thousands of topological atoms overnight in the development of a topological force field.[12−15]

Atomic integrations over topological atoms were already being carried out[6] in 1973, be it only for linear molecules (hydrides). The first computer program[16] able to integrate topological atoms in any polyatomic molecule appeared eight years later. This

Special Issue: Richard F. W. Bader Festschrift

Received: July 1, 2011
Revised: September 29, 2011
Published: October 06, 2011

algorithm was based on the mathematically elegant mapping from natural coordinates (describing a topological atom as a complete universe), to Cartesian coordinates (which describe full space). Due to numerical difficulties in the Jacobian when evaluated near a bond critical point, this algorithm (called OMEGA) was abandoned in favor of a fundamentally different one[17] (called PROAIM), which constructs explicit atomic boundaries and shoots integration rays from the nucleus out to where they intersect these boundaries. About two decades later, the former algorithm was resurrected[18] by the same author, recognizing that this algorithm had been abandoned prematurely for infelicities that were then overcome. This renewed effort led to the AIM2000 code[19] and its update.[20]

Meanwhile, in the mid-1990s the program MORPHY[21] appeared as the first integration code[22] written independently from PROAIM or PROMEGA. For the first time, MORPHY also combined in a single program the localization[23] of critical points in the electron density and its Laplacian, the plotting of the electron density and its gradient vector field, as well as the integration of atomic properties. Before, these functionalities had been separated into feebly interfaced programs constituting the AIMPAC suite. However, since 1997, a program called AIMALL[24] emerged from ongoing fine-tuning of the original subroutines of AIMPAC, resulting in a much faster and user-friendly code due to its GUI. From the mid-1990s code, independent from AIMPAC, was developed[25] and implemented in the program GAUSSIAN94 but later withdrawn because of frequent numerical instabilities and even erroneous results. In 1997, the Oviedo group also contributed an independently written topological code called CRITIC,[26] which is devoted to solid state systems, as fully described in a recent publication.[27] In the early 2000s a code called PROMOLDEN was also written by the Oviedo group. This code is specialized in molecular rather than solid state electron densities. It is often used in the context of the energy partitioning[28] and implements the Interacting Quantum Atom (IQA) formalism.[29] Very recent work[30] proposed a new method called QTREE, which is well equipped to cope with the complexity of topological atoms appearing in the solid state. Its close and preceding relative, the OCTREE algorithm,[31] which is also based on a recursive subdivision scheme, was developed to cope with the complexity of basins in the Laplacian of the electron density. This algorithm enabled,[32] for the first time, the calculation of volumes and electron populations of Laplacian basins, as well as their visualization. The Electron Localization Function, a popular tool[33] to extract chemical information from wave functions, triggered the independent development of the topological code TopMoD.[34] Finally, the solid state community embraced the topology early on. Gatti's TOPOND code[35] determines the topology of an analytically expressed Hartree−Fock or DFT wave function obtained by the program CRYSTAL.[36]

The high-resolution X-ray crystallography community also created its own algorithms early on. TOPXD, adapted by Gatti and Volkov, operates on experimental densities expressed by the analytical Hansen−Coppens model, which is fitted to reproduce the experimental diffraction intensities. The Hansen−Coppens model is the most common approach to the analysis of experimental electron densities in codes such as NEWPROP[37] or WINXPRO.[38] However, a topological analysis can also be carried out using grids. The InteGriTy package[39] analyses experimentally obtained electron densities sampled on 3D grids. The accuracy of the topological analysis of gridded electron densities has also been investigated.[40] In 1997, Aray et al. proposed[41] a numerical method to analyze the topology of the electronic density regardless of how it was obtained, that is, analytically or numerically. This approach depends on a careful interpolation of the electron density for the whole molecule or crystal. Grid methods, now in the context of computed electron densities, recently witnessed a resurgence in the work of Henkelman et al.,[42,43] in that of Rodríguez and Köster et al.,[44] and in a very recent contribution by Yu and Trinkle.[45] Two other algorithms should be mentioned because they are based on principles quite different from those that the aforementioned codes are based on. They are the "elastic sheet" method,[46] which involves forces acting on a swarm of fictitious particles, and a finite element method,[47] which makes contact with the world of engineering, structural mechanics, fluid dynamics, and weather prediction.

The brief review above is not intended to be exhaustive but merely aims at demonstrating the activity in algorithm development serving the topological analysis. The development of so many different algorithms and programs is a sign of the need of several communities (including that of material science, geochemistry, and crystallography) to obtain more accurate basin properties and also faster. This is not enough, however, if QTAIM ever wants to reach the status of a mainstream method. The inclusion of topological software in large and well-known software packages, which are typically commercially released such as ADF, will help making Quantum Chemical Topology (QCT) more widely used. The acronym QCT[32,48,49] embraces all work that shares the central idea of the gradient vector field being the partitioning engine of a quantum system. A gradient vector field is a collection of gradient paths, or paths of steepest ascent through a 3D function, such as the electron density, its Laplacian, the electron localization function, the electrostatic potential, the intracule density or a magnetically induced current distribution. QCT is further enriched by new measures such as the local single particle momentum, which was recently proposed[50] as a localized-electrons detector (LED). Although a full topological treatment with the calculation of its own basins still needs to be developed, LED already provides a direct three-dimensional representation of bonding interactions in molecules.[50,51] Noncovalent interactions can be revealed[52] by the reduced density gradient, a fundamental dimensionless quantity in DFT. Although not presented as such, this method is actually part of QCT, due to its dependence on concepts of dynamical systems theory, also appearing in QTAIM.

In this paper we present a fully analytical 3D volume integration over the so-called β sphere.[53] The β sphere is, in principle, the largest sphere that is completely contained within the topological atom at whose nucleus it is centered. In practice, the β sphere is just a sphere with an adequately large radius, typically about 90% of the distance between the nucleus and the nearest bond critical point. In our experience, this percentage has been safe for molecules and molecular clusters. Indeed, MORPHY's in-built test never detected that the β sphere exceeded the basin boundaries for these systems. However, the 90% choice may cause this problem in solid state systems, where this percentage needs to be reduced. Originally, the β sphere was introduced (in the OMEGA[53] algorithm) as a natural set of initial points for tracing (in the backward sense) the gradient paths that constitute a topological atom. However, the β sphere survived as an object in integration algorithms even with an underlying philosophy different from that of OMEGA. This is due to the fact that the electron density is very high near the nucleus and because quadrature permeates through all algorithms discussed

above. Indeed, when one is interested in the energy of an atom, steep variations in the integrand require dense quadrature grids. This requirement is less stringent in the computation of atomic charge. For atomic energy, an extra sphere with its own quadrature grid and placed within the β sphere, is known to improve the integration accuracy for elements of the fourth period and beyond. Ironically, the main reason for the use of quadrature grids is the potentially highly irregular shape of a topological basin. A topologically tame object such as a sphere is most suitable for a fully analytical treatment. This paper aims at setting a very first step toward a topological energy partitioning that is more accurate and faster in the future. The main focus, however, has been on topological multipole moments, which will continue to play a role in the evaluation of interatomic electrostatic energy, provided the convergence condition is obeyed.

2. METHOD AND DISCUSSION

The force field project mentioned in the Introduction warrants the evaluation of spherical tensor multipole moments[54,55] of a topological atom denoted Q_{LM}, which were first computed[22] in 1996. Earlier, Cartesian multipole moments had been calculated[56] in the context of QTAIM but QTAIM spherical multipole moments appeared only in 1996. The spherical tensor multipole moments are less intuitive than their Cartesian counterparts but they are more compact because of their irreducibility. For example, there are only five quadrupole moments instead of the six possible Cartesian ones. The discrepancy between the different numbers of components becomes more dramatic with increasing multipole rank L. In the spherical tensor formalism there are only $2L+1$ multipole moments (i.e., components). Thus, the full atomic electron density can be compressed into only a handful of numbers (i.e., multipole components). For our purpose, octupole moments (Q_{3M}, 7 components) and even hexadecupole moments (Q_{4M}, 9 components) are routinely evaluated.

To obtain the contribution of the electron density $\rho(\mathbf{r})$ inside the β sphere toward multipole moment Q_{LM}, one has to evaluate the following 3D volume integral,

$$Q_{LM}(\beta) = \int_{\beta\text{sphere}} \mathrm{d}\tau\ \rho(\mathbf{r})\ R_{LM}(\mathbf{r}) \tag{1}$$

where the radius β operates as a parameter and where the regular spherical tensor R_{LM}, is defined by eq 2, introducing the spherical harmonic $Y_{LM}(\theta,\varphi)$,

$$R_{LM}(\mathbf{r}) \equiv R_{LM}(r,\theta,\varphi) = \sqrt{\frac{4\pi}{2L+1}}\, r^L Y_{LM}(\theta,\varphi) \tag{2}$$

and the electron density is obtained within the usual SCF-LCAO-MO scheme using Gaussian (primitive functions),

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{i=1}^{n_{\mathrm{MO}}} n_i {\psi_i}^2(\mathbf{r}) = \sum_{i=1}^{n_{\mathrm{MO}}} n_i \Big(\sum_{j=1}^{n_{\mathrm{G}}} c_{ji} G_j(\mathbf{r})\Big)\Big(\sum_{k=1}^{n_{\mathrm{G}}} c_{ki} G_k(\mathbf{r})\Big) \\ &= \sum_{i=1}^{n_{\mathrm{MO}}} \sum_{j=1}^{n_{\mathrm{G}}} \sum_{k=1}^{n_{\mathrm{G}}} n_i c_{ki} c_{ji} G_j(\mathbf{r})\ G_k(\mathbf{r})\end{aligned} \tag{3}$$

where $G(\mathbf{r})$ represents a Gaussian of arbitrary angular momentum, which is *centered at an arbitrary nucleus*. In eq 3, n_{MO} is the number of molecular orbitals (MO), n_{G} is the number of Gaussians in each MO, and c are the LCAO coefficients. The triple summation on the utmost right-hand side is just written in a visually convenient rather than computationally efficient manner.

It is convenient to write compactly the argument of the electron density and R_{LM} as a position vector $\mathbf{r}$, expressed with respect to the global frame. However, this notation glosses over how the integration in eq 1 is actually solved. Equation 2 hints at the use of spherical polar coordinates in the triple integration procedure. However, this equation conceals the mismatch between the origin of the Gaussians and the origin of the regular spherical harmonic R_{LM}, which is the nucleus of the topological atom. This origin mismatch needs to be addressed when the integral is solved over the (finite) spherical boundary. A possible way forward to solve this problem is that of Kaufmann and Baumeister[57] (see below), which we pursued at first and then abandoned. A second comment concerns the complex character of R_{LM}. Atomic multipole moments are usually expressed as real quantities, which can be obtained from simple linear combinations reminiscent of the way the sine and cosine function are obtained from $\exp(ix)$ and $\exp(-ix)$. Details can be found in Appendix B of Stone's book[55] and in the appendix of ref 58. After this transformation, the components of the dipole moment, denoted z, x, and y, become R_{10}, R_{11c}, and R_{11s}, respectively. A recognizable example of a quadrupole moment is probably $R_{22c} = (3^{1/2}/2)(x^2 - y^2)$. Higher-rank multipole moments consist of increasingly complicated linear combinations of terms of the shape $x^{l1}y^{l2}z^{l3}$, where $l1 + l2 + l3 = L$.

Naively, one would expect the integration problem posed in eq 1 to benefit from the large literature on molecular integrals over Gaussians. When $L = M = 0$, one calculates the monopole moment, which boils down to a (two-center) overlap integral. However, the finite boundary in eq 1 makes this problem different from that of a simple overlap integral, which invokes much easier boundaries at infinity. A possible route forward is offered by the work[59] of Cruz and Soullard. They calculated the ground state energy of a helium atom confined by hard spherical walls, following the approach[57] of Kaufmann and Baumeister. The latter researchers introduced the modified Bessel function $I_l(z)$ and the Rayleigh expansion, which writes an exponential function as a series of spherical harmonics. Application of the addition theorem for surface spherical harmonics then yields the desired single-center expansion for a Gaussian centered at an arbitrary point. We will show below that there is no need for this Rayleigh expansion, and that there is a direct method of integration.

Using the Gaussian Product Rule, the product of any two Gaussians can be written as a single Gaussian $G_{jk}(\mathbf{r})$, centered on a point $\mathbf{R}$, calculated from the original centers $\mathbf{R}_j$ and $\mathbf{R}_k$. The new Gaussian exponent α_{jk} and proportionality factor K_{jk} also simply follow from the two original exponents and position vectors. Formally, we have

$$\begin{aligned}&\exp(-\alpha_j|\mathbf{r}-\mathbf{R}_j|^2)\ \exp(-\alpha_k|\mathbf{r}-\mathbf{R}_k|^2) \\ &= K_{jk}\ \exp(-\alpha_{jk}|\mathbf{r}-\mathbf{R}_{jk}|^2) = G_{jk}(\mathbf{r})\end{aligned} \tag{4}$$

where

$$\alpha_{jk} = \alpha_j + \alpha_k \tag{5a}$$

$$\mathbf{R}_{jk} = \frac{\alpha_j\mathbf{R}_j + \alpha_k\mathbf{R}_k}{\alpha_j + \alpha_k} \tag{5b}$$

dx.doi.org/10.1021/jp206223g |J. Phys. Chem. A 2011, 115, 13169–13179

and

$$K_{jk} = \exp\left(-\frac{\alpha_j \alpha_k}{\alpha_j + \alpha_k}|\mathbf{R}_j - \mathbf{R}_k|^2\right) \quad (5c)$$

We assume that the normalization constants preceding the original Gaussians are implicit in the c-coefficients of eq 3 so that there is no need to keep track of these normalization constants in eq 4.

Substituting eq 3 into eq 1 leads to eqs 6 and 7,

$$Q_{LM}(\beta) = \sum_{i=1}^{n_{MO}} n_i \sum_{j=1}^{n_G}\sum_{k=1}^{n_G} c_{ji}c_{ki}\int_{\beta\text{sphere}} d\tau\, R_{LM}(\mathbf{r})\, G_{jk}(\mathbf{r})$$
$$= \sum_{j=1}^{n_G}\sum_{k=1}^{n_G}\sum_{i=1}^{n_{MO}} n_i c_{ki}c_{ji} Q_{LM,jk}(\beta) \quad (6)$$

$$Q_{LM,jk}(\beta) = q_{LM}(\beta) = \int_{\beta\text{sphere}} d\tau\, R_{LM}(\mathbf{r})\, G_{jk}(\mathbf{r}) \quad (7)$$

Note that the sum over the Gaussian contributions contains n_G^2 terms because *all* pairs of original Gaussians need to be considered, including the application of the Product Rule on two identical Gaussians. In the latter case we obtain $K_{jj} = 1$, $\alpha = 2\alpha_j$, and $\mathbf{R} = \mathbf{R}_j = \mathbf{R}_k$. Of course, when $j \neq k$, the product $G_k(\mathbf{r})\, G_j(\mathbf{r})$ does not have to be recomputed if $G_j(\mathbf{r})\, G_k(\mathbf{r})$ has already been computed but both contributions need to be included in the sum. Second, the reason for putting the double sum over the Gaussians first is due to computational efficiency. The function $G_{jk}(\mathbf{r})$ contains the factor K_{jk}, (see eq 4), which is vanishingly small for many Gaussian pairs occurring in larger systems. In fact, it is computationally advantageous to evaluate the argument of the exponent of K_{ij} (see eq 5c) and then avoid the calculation of integrals of $G_{jk}(\mathbf{r})$ for which the prefactor K_{ij} is negligible. The larger (and more extended) the molecule, the more Gaussian product functions will have a K_{ij} value that is so small (e.g., 10^{-5}) that their contribution to ρ and hence Q_{LM} does not need to be evaluated. The percentage of Gaussian product functions that survive this evaluation threshold will be reported in the test section below. A second matter regards computational efficiency in connection with the computational overhead caused by the mandatory use of the Product Rule in the analytical integration. Appendix A, which can be consulted in the Supporting Information, addresses this issue in some detail.

To bring out the essence of the 3D integration, we focus on the simplest case, in which $R_{LM} = R_{00} = 1$ and $G_{jk}(\mathbf{r})$ is an s function. This function can only arise from two original Gaussian s functions. Because the indices j and k would clutter the derivations below, they are dropped. It should be kept in mind that the following derivation refers to two arbitrarily positioned s functions, both contributing to the monopole moment of a β sphere centered on a third center. Both simplifications will be abandoned later, allowing for higher angular momentum functions p and d, as well as for dipole, quadrupole, octupole, and hexadecupole moments. Because $L = M = 0$, eq 7 can then be specified to eq 8,

$$q_{00}(\beta) = \int_{\beta\text{sphere}} d\tau\, G(\mathbf{r}) = K\int_{\beta\text{sphere}} d\tau \exp(-\alpha|\mathbf{r} - \mathbf{R}|^2) \quad (8)$$

where K, α, and $\mathbf{R}$ are shorthand for the parameters of the product Gaussian (see eqs 5a, 5b, and 5c).

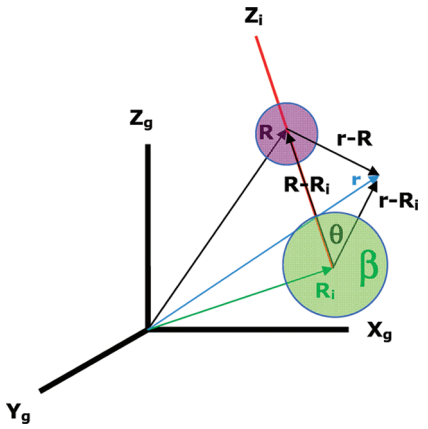


Figure 1. Axis systems (global "g" and local "i") and position vectors ($\mathbf{r}$, $\mathbf{R}$, $\mathbf{R}_i$, $\mathbf{r}-\mathbf{R}$, $\mathbf{r}-\mathbf{R}_i$, and $\mathbf{R}-\mathbf{R}_i$) involved in the analytical integration of a Gaussian s function (isovalue marked by a purple sphere) over the β sphere (green). Detailed explanations in the main text.

The integral in eq 8 can be worked out after inspection of Figure 1, which clarifies the relationship between all vectors involved. The *global* axis system (X_g, Y_g, Z_g) is the overall reference for each topological atom that is to be integrated. However, a *local* axis system (X_i, Y_i, Z_i) serves as a reference for the spherical coordinates that will be used to solve the triple integral at hand. The index i refers to the fact that the *integration* will be formulated with respect to this local axis system, centered at the nucleus, which is the integration center. Only the Z_i axis is shown because there is no need for X_i and Y_i if only an s function is to be integrated. This will change later when higher angular momentum Gaussians are integrated. The integration variables in eqs 1, 6, and 8 have been deliberately summarized using the vague symbol "$d\tau$" to postpone the question as to what these variables are precisely. Below, this question will be answered formally but with an eye on Figure 1, we can already state that "$d\tau$" will stand for spherical (polar) coordinates installed in the local axis system (X_i, Y_i, Z_i).

In Figure 1, the position vector $\mathbf{R}_i$ marks the integration center, that is, the nuclear position of the topological atom to be integrated. Position vector $\mathbf{R}$ marks the center of the Gaussian that contributes to the electron density. Position vector $\mathbf{r}$ refers to the position of an element of electron density contributing to the atom's multipole moment. Note that $\mathbf{r}$ does *not* constitute the three integration variables. The angle between $\mathbf{R}-\mathbf{R}_i$ and $\mathbf{r}-\mathbf{R}_i$ is denoted by θ, which coincides with the familiar azimuthal coordinate θ. *This means that the Z_i axis, which governs the integration, must coincide with $\boldsymbol{R}-\boldsymbol{R}_i$.*

Below, it will become clear that this convenient choice is crucial in the way the integral of eq 8 is solved. This *choice avoids the need to re-express the Gaussian as centered at the integration center $\boldsymbol{R}_i$.* The latter would be achieved by means of a series expansion, which "rebuilt" the Gaussian at its new center. Note that in the current procedure *the Gaussian is integrated from a center that is not its own (original) center.* Although this may appear unnatural at first sight, the following derivation shows that this procedure is perfectly feasible.

The argument of the exponential in eq 8 can be replaced by explicit spherical coordinates using eq 9,

$$\begin{aligned}|\mathbf{r}-\mathbf{R}|^2 &= |(\mathbf{r}-\mathbf{R}_i)-(\mathbf{R}-\mathbf{R}_i)|^2 \\ &= |\mathbf{r}-\mathbf{R}_i|^2 + |\mathbf{R}-\mathbf{R}_i|^2 - 2(\mathbf{r}-\mathbf{R}_i)\cdot(\mathbf{R}-\mathbf{R}_i) \\ &= r'^2 + R'^2 - 2r'R'\cos\theta \end{aligned} \tag{9}$$

The primes can be dropped provided one keeps in mind that the distances r and R are measured with respect to the integration origin and not the global origin. Equation 8 then becomes eq 10,

$$\begin{aligned}q_{00}(\beta) = K\int_0^\beta\int_0^\pi\int_0^{2\pi} \mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi\, r^2\sin\theta \\ \times \exp[-\alpha(r^2 + R^2 - 2Rr\cos\theta)]\end{aligned} \tag{10}$$

where the Jacobian determinant $r^2\sin\theta$ has been introduced. The integrand can be conveniently rewritten as

$$\begin{aligned}q_{00}(\beta) = K\mathrm{e}^{-\alpha R^2}\int_0^{2\pi}\mathrm{d}\varphi\int_0^\beta \mathrm{d}r\, r^2\mathrm{e}^{-\alpha r^2}\int_0^\pi \mathrm{d}\theta\sin\theta \\ \times \exp[(2\alpha R)r\cos\theta]\end{aligned} \tag{11}$$

Equation 11 has been slightly rearranged to emphasize the difference between integration variables and parameters such as α and R. Later we will introduce $p = 2\alpha R$. The integral over θ can easily be solved realizing that

$$\begin{aligned}\int_0^\pi \mathrm{d}\theta\sin\theta\exp[2\alpha Rr\cos\theta] &= -\int_0^\pi \mathrm{d}\cos\theta\exp[2\alpha Rr\cos\theta] \\ &= -\frac{1}{2\alpha Rr}[\exp(2\alpha Rr\cos\theta)]_0^\pi = \frac{1}{2\alpha Rr}[\exp(2\alpha Rr) - \exp(-2\alpha Rr)]\end{aligned} \tag{12}$$

such that eq 11 becomes eq 13 after a trivial integration over φ,

$$\begin{aligned}q_{00}(\beta) &= K\frac{2\pi}{2\alpha R}\int_0^\beta \mathrm{d}r\, r\exp[-\alpha(r^2+R^2)] \\ &\quad\times[\exp(2\alpha rR) - \exp(-2\alpha rR)] \\ &= K\frac{\pi}{\alpha R}\int_0^\beta \mathrm{d}r\, r\{\exp[-\alpha(r-R)^2] - \exp[-\alpha(r+R)^2]\}\end{aligned} \tag{13}$$

Using the final result derived in Appendix B (Supporting Information), eq B6, one obtains

$$\begin{aligned}q_{00}(\beta) = K\frac{\pi}{\alpha R}\Bigg\{&-\frac{1}{2\alpha}\{\exp[-\alpha(\beta-R)^2] - \exp[-\alpha R^2]\} \\ &+\frac{1}{2}R\sqrt{\frac{\pi}{\alpha}}\{\mathrm{erf}(\sqrt{\alpha}(r-R)) - \mathrm{erf}(-\sqrt{\alpha}R)\} \\ &+\frac{1}{2\alpha}\{\exp[-\alpha(\beta+R)^2] - \exp[-\alpha R^2]\} \\ &+\frac{1}{2}R\sqrt{\frac{\pi}{\alpha}}\{\mathrm{erf}(\sqrt{\alpha}(r+R)) - \mathrm{erf}(+\sqrt{\alpha}R)\}\Bigg\}\end{aligned} \tag{14}$$

It is convenient to introduce two shorthand variables due to the recurrence of certain expressions in eq 14. It makes sense to define S as a ("α weighted") sum of the radius of the integration sphere β and the distance of the Gaussian center to the integration origin, R. Similarly, D can be defined as the difference, or

$$\begin{aligned}S &= \sqrt{\alpha}(\beta + R) \quad 0 \le S < \infty \\ D &= \sqrt{\alpha}(\beta - R) \quad -\infty \le D < \infty\end{aligned} \tag{15}$$

After substitution of eq 15 into eq 14, and after rearrangement and term cancellation, also using the property that $\mathrm{erf}(-x) = -\mathrm{erf}(x)$, one obtains

$$\begin{aligned}q_{00}(\beta) = K\frac{\pi}{\alpha R}\Bigg\{&\frac{1}{2\alpha}\{\exp[-S^2] - \exp[-D^2]\} \\ &+\frac{1}{2}R\sqrt{\frac{\pi}{\alpha}}[\mathrm{erf}(D) + \mathrm{erf}(S)]\Bigg\}\end{aligned} \tag{16}$$

which can be reorganized leading to the key result of eq 17,

$$\begin{aligned}q_{00}(\beta) = K\frac{\pi}{2\alpha^2}\Bigg\{&\frac{\exp[-S^2] - \exp[-D^2]}{R} \\ &+\sqrt{\alpha\pi}[\mathrm{erf}(S) + \mathrm{erf}(D)]\Bigg\}\end{aligned} \tag{17}$$

It is important to inspect the limiting cases possibly occurring in eq 17. Four possibilities can be distinguished: $\beta \to \infty$, $\beta \to 0$, $R \to \infty$, or $R \to 0$. First, what happens if the integration volume covers all space? This situation corresponds to $\beta \to \infty$, such that $S \to \infty$ and $D \to \infty$ from eq 15, resulting in

$$q_{00}(\beta) = K\frac{\pi}{2\alpha^2}\left\{\frac{0-0}{R} + \sqrt{\alpha\pi}[1+1]\right\} = K\left(\frac{\pi}{\alpha}\right)^{3/2} \tag{18}$$

This outcome indeed coincides with that of the familiar Gaussian overlap integral, which is of course obtained by integration over whole space, or

$$\left[\frac{\pi}{(\alpha_j+\alpha_k)}\right]^{3/2}\exp\left[\frac{-\alpha_j\alpha_k}{(\alpha_j+\alpha_k)}|\mathbf{R}_j-\mathbf{R}_k|^2\right] \tag{19}$$

as is clear from invoking eqs 5a and 5c.

The second limiting case occurs when β decreases to zero. Intuitively, one expects q_{00} to reduce to zero as well. Indeed, if $\beta \to 0$ then $S \to \alpha^{1/2}R$ and $D \to -\alpha^{1/2}R$. When inserted into the master equation (eq 17) and using $\mathrm{erf}(-x) = -\mathrm{erf}(x)$, it is easy to show that $q_{00}(\beta)$ vanishes, or

$$\begin{aligned}q_{00}(\beta) = K\frac{\pi}{2\alpha^2}\Bigg\{&\frac{\exp[-\alpha R^2] - \exp[-\alpha R^2]}{R} \\ &+\sqrt{\alpha\pi}[\mathrm{erf}(\sqrt{\alpha}R) + \mathrm{erf}(-\sqrt{\alpha}R)]\Bigg\} = 0\end{aligned} \tag{20}$$

As a third limiting case, we investigate what happens if a Gaussian function is far removed from the integration center, that is, when the magnitude of $\mathbf{R}-\mathbf{R}_i$ in Figure 1 becomes very large. This situation corresponds to a large R' value, or $R \to \infty$ because the primes have been dropped, as explained above. From eq 15 it follows that, if $R \to \infty$, then $S \to \infty$ and $D \to -\infty$. Clearly, the first term in the master equation (eq 17) vanishes, as does the second term because $\mathrm{erf}(\infty) = 1$ and $\mathrm{erf}(-\infty) = -1$. Hence, q_{00} again vanishes, as expected, because an infinitely remote Gaussian does not contribute to the β sphere's monopole moment.

dx.doi.org/10.1021/jp206223g | J. Phys. Chem. A 2011, 115, 13169–13179

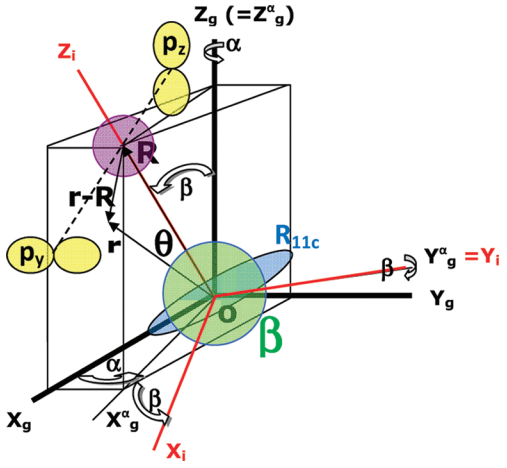


Figure 2. Generalization of Figure 1, enabling integration over a product Gaussian function of arbitrary angular momentum centered at **R.** The integration center is at the global origin, where an example of a regular spherical harmonic (R_{11c}, blue) is also centered. Note that Y^{α}_{g} is orthogonal to the $X^{\alpha}_{g}Z_g$ plane. Detailed explanations are in the main text.

The fourth and final case to be considered is when $R = 0$ in the limit, leading to $S = D = \alpha^{1/2}\beta$, again from eq 15. The second term in the master equation (eq 17) then simply reduces to $K(\pi/\alpha)^{3/2}\exp(\alpha^{1/2}\beta)$. However, the first term is troublesome because it involves a division by zero. Fortunately, the limiting case of $R = 0$ can be treated with l'Hôpital's rule, which for our purpose states that, if $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$, then $\lim_{x\to c} f(x)/g(x) = \lim_{x\to c} f'(x)/g'(x)$ (if the right-hand side exists), where the prime denotes differentiation with respect to x. Indeed, the numerator $[f(x) = \exp(-S^2) - \exp(-D^2) = \exp(-\alpha\beta^2) - \exp(-\alpha\beta^2) = 0]$ and denominator $[g(x) = R]$ of the first term in eq 17 both tend to zero as $R \to 0$. Hence, the first term in eq 17 is calculated to be

$$\begin{aligned}
&\frac{K\pi}{2\alpha^2}\lim_{R\to 0}\frac{\{\exp[-S^2] - \exp[-D^2]\}}{R}\\
&= \frac{K\pi}{2\alpha^2}\lim_{R\to 0}\{(-2\alpha)(\beta + R)\exp[-\alpha(\beta + R)^2]\\
&\quad - (-2\alpha)(\beta + R)(-1)\exp[-\alpha(\beta - R)^2]\}^{-1}\\
&= \frac{K\pi}{2\alpha^2}(2)(-2\alpha\beta)\exp[-\alpha\beta^2] = -\frac{2\pi K\beta}{\alpha}\exp[-\alpha\beta^2]
\end{aligned} \tag{21}$$

Adding this result (the right-hand side of eq 21) to the second term in eq 17 (with $S = D = \alpha^{1/2}\beta$) yields exactly the same result of a calculation where $q_{00}(\beta)$ is directly calculated as a volume integral in which R is set to zero already in the integrand. This calculation is given in Appendix C (Supporting Information) and the final result for comparison is the last branch of eq C3. The discussion of this limiting case is not only important for checking purpose or obtaining a deeper understanding of the behavior of the result in the master equation (eq 17). The most important implication of the above analysis is for the software implementation, where a separate formula needs to be coded for the common case where $R = 0$, corresponding to a Gaussian being centered on the nucleus.

In the wake of the four limiting cases discussed above a fifth case needs to be briefly discussed, namely that of a very "peaked" Gaussian, resembling a Dirac delta function. In this case, α's value is very large, which affects both S and D in eq 14 but not R itself. If $\alpha \to \infty$, then $S \to \infty$ always, but D can reach either $+\infty$ or $-\infty$ depending on whether $R < \beta$ or $R > \beta$, respectively. For $R < \beta$, we find that $q_{00}(\beta) = K(\pi/\alpha)^{3/2}$ following the logic of the first limiting case. For $R > \beta$ we recover the third limiting case (because $S \to \infty$ and $D \to -\infty$) and q_{00} vanishes.

We now extend the analytical integration to higher angular momentum Gaussians and to regular spherical tensors R_{LM} where L and M are both no longer set to zero. Without loss of generality, we can make the nuclear position coincide with the global origin, or $\mathbf{R}_i = 0$. As a result, there is no need any more to introduce the primed variables appearing in eq 9. Figure 2 illustrates the new arrangement of vectors and symbols used from here onward. It is very important *to retain the choice of assigning as the local Z axis (Z_i), the line through the integration center (**o**), and the center **R*** of the (product) Gaussian. However, because of the nonsphericity of the Gaussian, assigning only a Z_i axis is not sufficient (but it was for just s functions). There is now a need to also define an X_i and a Y_i axis. A natural way of achieving this task is to apply successive rotations to the global axis system such that Z_g is made to coincide with Z_i. This transformation automatically puts the remaining axes X_i and Y_i in place and completes the installation of a local axis system for integration. This rotation is described in detail in an instant.

As before, the integration variables are defined with respect to the local axis system $\{X_i, Y_i, Z_i\}$. The angular part of a Gaussian is defined with respect to the global frame $\{X_g, Y_g, Z_g\}$. For example, a p_z function centered at $\mathbf{R}_k$ is represented by $(z_g - Z_{k,g})\exp[-\alpha|\mathbf{r} - \mathbf{R}_k|^2]$, where the subscript g emphasizes the reference to the global frame and $\mathbf{R}_k$ has global coordinates $(X_{k,g}, Y_{k,g}, Z_{k,g})$. Figure 2 shows this p_z function in yellow, together with another example of a nonspherical Gaussian (p_y), which is centered at $\mathbf{R}_j$. In this example, the product of p_y and p_z creates the radial part of the product Gaussian (s-like, purple) centered at $\mathbf{R}$, which is really $\mathbf{R}_{jk}$, as defined by eq 5b. Note that the p_z function is parallel to the global Z axis. The integration center is at the origin (compare with Figure 1: $\mathbf{R}_i = 0$). The position vector $\mathbf{r}$ marks the integration variable, which describes the electron density. As before (Figure 1), the angle between $\mathbf{R}$ and $\mathbf{r}$ is denoted θ, an angle that will again serve as integration variable. The spherical harmonics R_{LM} are centered at the integration center, which is the global origin (i.e., the nucleus of the topological atom). R_{LM} adopts the orientation of the global axis system. Figure 2 shows the example of R_{11c} in blue; this is simply the function x.

Now we explain the details of the Euler rotation that expresses the mapping between the local coordinates $\{x_i, y_i, z_i\}$ describing the integration frame and the global coordinates describing the global frame $\{x_g, y_g, z_g\}$. The rotation of the global frame into the local frame is realized by means of only two Euler angles, referred to as α and β. There is no need for the usual third Euler angle γ, which is therefore conveniently set to zero. Note that in the formulas below, the Euler angle β should not be confused with the radius of the β sphere. We follow one of the twelve possible conventions, called ZYZ, which is used in the area of group theory and the quantum theory of angular momentum. We first rotate counterclockwise by an angle α, around the Z_g axis, such that the X_g axis becomes the X^{α}_{g} axis and the Y_g axis the Y^{α}_{g} axis. This process is illustrated in Figure 2. Normally, angles are defined in such a way that they are positive when they rotate counterclockwise. Note that the Y^{α}_{g} axis is orthogonal to the

$X^{\alpha}{}_{g}Z$ plane. The second rotation is also counterclockwise, over an angle β, but now around the $Y^{\alpha}{}_{g}$ axis. Because the goal of making Z_g coincide with Z_i is then achieved, there is no need for a third Euler rotation (hence $\gamma = 0$). The local axis system is marked in red in Figure 2.

The rotation operator $\hat{R}$ that transforms the local axis system $\{X_i, Y_i, Z_i\}$ (in which the integration is carried out) to the global axis system $\{X_g, Y_g, Z_g\}$, is given by eq 22,

$$\hat{R} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos\beta\cos\alpha & \cos\beta\sin\alpha & -\sin\beta \\ -\sin\alpha & \cos\alpha & 0 \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix} \tag{22}$$

The operator $\hat{R}$ transforms coordinates expressed with respect to the global axis system to coordinates expressed with respect to the integration axis system. To verify the coordinate transformation, we let $\hat{R}$ operate on a unit vector $\mathbf{u}$ lying on the Z_i axis, which has coordinates $(u_x, u_y, u_z) = (\sin\beta\cos\alpha,\ \sin\beta\sin\alpha,\ \cos\beta)$ when expressed in the global axis system, or

$$\hat{R}\begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}_{\text{global}} = \begin{pmatrix} \cos\beta\cos\alpha & \cos\beta\sin\alpha & -\sin\beta \\ -\sin\alpha & \cos\alpha & 0 \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix} \begin{pmatrix} \sin\beta\cos\alpha \\ \sin\beta\sin\alpha \\ \cos\beta \end{pmatrix}_{\text{global}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}_{\text{local}} \tag{23}$$

One indeed recovers that this vector has coordinates (0, 0, 1) in the local axis system (of integration).

The spherical harmonics tensors R_{LM} are expressed in terms of global coordinates $\{x_g, y_g, z_g\}$. To be integrated as part of the integrand in eq 1, their formulas need to be transformed in terms of the integration variables $\{x_i, y_i, z_i\}$, or simply $\{x, y, z\}$. The latter coordinates will then be turned into spherical polar coordinates. The inverse operator $\hat{R}^{-1}$ transforms coordinates with respect to the local axis system to coordinates with respect to the global axis system. Because the rotation matrix R is orthogonal, its inverse is equal to its transpose $\hat{R}^{\tau}$. To verify the coordinate transformation, we let $\hat{R}^{\tau}$ operate on a unit vector lying on the local Z axis or Z_i. This unit vector has coordinates (0, 0, 1) in the local axis system,

$$\hat{R}^{\tau}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}_{\text{local}} = \begin{pmatrix} \cos\beta\cos\alpha & -\sin\alpha & \sin\beta\cos\alpha \\ \cos\beta\sin\alpha & \cos\alpha & \sin\beta\sin\alpha \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}_{\text{local}} = \begin{pmatrix} \sin\beta\cos\alpha \\ \sin\beta\sin\alpha \\ \cos\beta \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix}_{\text{global}} \tag{24}$$

One indeed recovers that this vector has coordinates (u_x, u_y, u_z) in the global axis system.

We introduce the shorthand c for cos and s for sin and add the angles as α and β as subscripts. For example, $\sin(\beta)$ becomes s_β.

Equation 25 then expresses the global coordinates $\{x_g, y_g, z_g\}$ as a linear combination of the local coordinates $\{x, y, z\}$ involved in the integration,

$$\begin{pmatrix} c_\beta c_\alpha & -s_\alpha & s_\beta c_\alpha \\ c_\beta s_\alpha & c_\alpha & s_\beta s_\alpha \\ -s_\beta & 0 & c_\beta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}_{\text{integration}} = \begin{pmatrix} x_g \\ y_g \\ z_g \end{pmatrix} \tag{25a}$$

$$\begin{aligned} x_g &= c_\alpha c_\beta x - s_\alpha y + c_\alpha s_\beta z \\ y_g &= s_\alpha c_\beta x + c_\alpha y + s_\alpha s_\beta z \\ z_g &= -s_\beta x + c_\beta z \end{aligned} \tag{25b}$$

Equation 25 gives an example of a transformed R_{LM} tensor, obtained by using the substitution prescribed by eq 25b,

$$\begin{aligned} R_{21c} &= \sqrt{3}x_g z_g \\ &= \sqrt{3}(-c_\alpha c_\beta s_\beta x^2 + s_\alpha s_\beta xy \\ &\quad + c_\alpha c_\beta{}^2 xz - c_\alpha s_\beta{}^2 xz - s_\alpha c_\beta yz + c_\alpha c_\beta s_\beta z^2) \end{aligned} \tag{26}$$

This example shows that the overall (i.e., summed) power of each term remains the same as L in R_{LM}, which in this case is equal to 2. For example, $xy = x^1y^1$ and $1 + 1 = 2$.

For each Gaussian appearing as the result of the product rule, there is a new value for α and for β, which is why they act as parameters. Effectively, they are constants in the integration of the Gaussian. No pathological cases can arise in the transformation given in eq 25, even when $\alpha = \beta = \pi/2$. In that case, all matrix entries vanish except three off-diagonal elements but the determinant remains 1, as expected.

A general regular spherical harmonic R_{LM} is always a linear combination of powers of x_g, y_g, and/or z_g. Each term is a homogeneous function of degree L, which is preserved after the transformation of eq 25. After conversion of x, y, and z to spherical coordinates (with respect to the local axis system of integration), R_{LM} ends up as a linear combination of cosines and sines of θ and φ,

$$\begin{aligned} R_{LM} &= \sum_i C_i x^{l_i} y^{m_i} z^{n_i} \\ &= \sum_i C_i (r\sin\theta\cos\varphi)^{l_i} (r\sin\theta\sin\varphi)^{m_i} (r\cos\theta)^{n_i} \\ &= \sum_i C_i r^{l_i+m_i+n_i} \sin^{l_i+m_i}\theta\cos^{n_i}\theta\cos^{l_i}\varphi\sin^{m_i}\varphi \end{aligned} \tag{27}$$

where C_i consists of a product of sines and cosines of α and β and, hence, varies parametrically with each Gaussian product center. Although the result of eq 27 is rather involved, the alternative path of rotating spherical harmonics $Y_{lm}(\theta,\varphi)$ directly entails other complexities.[60]

We are now in a position to return to eq 11 and generalize it for any spherical tensor, which leads to eq 28,

$$\begin{aligned} q_{LM}(\beta) = K e^{-\alpha R^2} \sum_i C_i &\int_0^{2\pi} d\varphi \sin^{m_i}\varphi\cos^{l_i}\varphi \\ &\times \int_0^{\beta} dr\, r^{l_i+m_i+n_i+2} \exp(-\alpha r^2) \\ &\times \int_0^{\pi} d\theta \sin^{l_i+m_i+1}\theta\cos^{n_i}\theta \exp[(2\alpha R) r\cos\theta] \end{aligned} \tag{28}$$

dx.doi.org/10.1021/jp206223g | J. Phys. Chem. A 2011, 115, 13169–13179

The integral over φ always vanishes unless $l + m$ is even; evaluating this integral first can save much CPU time because, if it vanishes, there is no need to evaluate the integrals in θ and r. Before moving on to a general expression for a pair of Gaussians of arbitrary angular momentum and an arbitrary R_{LM}, it is instructive to work out the case of R_{10}, which is shown in Appendix D (Supporting Information). It is shown there that the general solution may involve the evaluation of exponential functions of very large arguments. Appendix E (Supporting Information) shows that manual algebraic manipulation can avoid this evaluation.

The methodology of molecular integrals re-expresses the product of Gaussian angular parts in terms of the Gaussian product center **R**. This center is the integration center for the corresponding overlap integral, and because the integration boundaries are infinite, each Gaussian product function can be integrated from its own center. In the current problem of β sphere integration the integration center is fixed and placed at the global origin (Figure 2). Hence, we cannot use the formulas of the molecular integrals literature.

The proposed algorithm will be illustrated on three systems: a small wave function (capped glycine, 322 Gaussian primitives, 19 nuclei), a large wave function (crambin, 5922 Gaussian primitives, 642 nuclei), and a system containing a heavy element ($Sn_2(CH_3)_2$, 1352 Gaussian primitives, 10 nuclei). The implementation of the analytical integration is part of a local version of MORPHY consisting of a dozen subroutines (excluding functions from Numerical Recipes) and more than a thousand lines.

The Gaussian product of two Gaussians G_j and G_k each of arbitrary angular momentum is expressed by eq 29,

$$G_{jk}(\mathbf{r}) = K_{jk}\exp(-\alpha_{jk}|\mathbf{r}-\mathbf{R}_{jk}|^2)(x_g - X_j)^{\bar{l}_j} \times (x_g - X_k)^{\bar{l}_k}(y_g - Y_j)^{\bar{m}_j}(y_g - Y_k)^{\bar{m}_k}(z_g - Z_j)^{\bar{n}_j}(z_g - Z_k)^{\bar{n}_k} \quad (29)$$

Note that the radial part, captured in the s-like product part, still determines the overall procedure of integration over the β sphere, as explained early on and illustrated in Figure 1. Substitution of eq 25b into eq 29 introduces the desired integration variables x, y, and z. Working out the resulting products leads to a number of terms, which are products of powers of x, y, and z. After grouping, they can be written, in general, as

$$G_{jk}(r) = K_{jk}\exp(-\alpha_{jk}|\mathbf{r}-\mathbf{R}_{jk}|^2)\sum_h C_h x^{\bar{l}_h} y^{\bar{m}_h} z^{\bar{n}_h} \times K_{jk}\exp(-\alpha_{jk}|\mathbf{r}-\mathbf{R}_{jk}|^2) \times \sum_h C_h r^{\bar{l}_h+\bar{m}_h+\bar{n}_h}\sin^{\bar{l}_h+\bar{m}_h}\theta\cos^{\bar{n}_h}\theta\cos^{\bar{l}_h}\varphi\sin^{\bar{m}_h}\varphi \quad (30)$$

where the (parametric) coefficients C_h are functions of c_α, c_β, s_α, s_β, X_j, Y_j, Z_j, X_k, Y_k, and Z_k, and where the conversion to spherical coordinates has been carried out, as in eq 27. Finally, the master expression for $q_{LM}(\beta)$ (i.e., contribution of a single Gaussian product) appears in eq 31, after substitution of eq 27 and 30 into eq 7,

$$q_{LM}(\beta) = Ke^{-\alpha R^2}\sum_i\sum_h C_i C_h\left[\int_0^{2\pi} d\varphi \sin^s\varphi\cos^t\varphi\right] \times \left[\int_0^\beta dr\, r^u e^{-\alpha r^2}\int_0^\pi d\theta \sin^v\theta\cos^w\theta\exp[pr\cos\theta]\right] \quad (31)$$

where $p = 2\alpha R$ and

$$s = m_i + \bar{m}_h$$
$$t = l_i + \bar{l}_h$$
$$u = l_i + m_i + n_i + \bar{l}_h + \bar{m}_h + \bar{n}_h + 2$$
$$v = l_i + m_i + \bar{l}_h + \bar{m}_h + 1$$
$$w = n_i + \bar{n}_h \quad (32)$$

The large brackets in eq 31 highlight the independence of the integral in φ of the integral in r and θ. It is straightforward to show that eq 11 follows from eq 31 as a special case in which both Gaussians are s functions and R_{LM} is restricted to R_{00}. In this case, eq 32 returns that $s = t = w = 0$, $u = 2$, and $v = 1$. Also, because there is then no need for frame rotation, we have $\alpha = \beta = 0$ and hence $c_\alpha = c_\beta = 1$ and $s_\alpha = s_\beta = 0$. The detailed software implementation of eq 31, which invoked Mathematica[61] as well as tedious manual calculation and coding, will not be reported here due to space limitation. Note that the combination of high angular momentum (e.g., d_{xz} and d_{yz} being multiplied) and high R_{LM} moments (e.g., fourth powers occurring in hexadecupole moments) leads to high values of s, t, u, v, and w. This warranted well-organized coding in FORTRAN to avoid unmanageably long explicit expressions, as well as a degree of open-ended code that generated the necessary coefficients implicitly, for any power.

Finally, we report preliminary work on the calculation of energy, in particular the integration of the kinetic energy densities $K(\mathbf{r})$ and $G(\mathbf{r})$ over the volume of a β sphere. The quantity $K(\beta)$ is calculated via eq 33,

$$K_{kin}(\beta) = -\frac{1}{2}\sum_{i=1}^{n_{MO}}\sum_{j=1}^{n_G}\sum_{k=1}^{n_G} n_i c_{ki} c_{ji}\int_{\beta\text{sphere}} d\tau\; G_j(r)\;\nabla^2 G_k(\mathbf{r}) \quad (33)$$

and $G(\beta)$ is calculated via eq 34,

$$G(\beta) = \frac{1}{2}\sum_{i=1}^{n_{MO}}\sum_{j=1}^{n_G}\sum_{k=1}^{n_G} n_i c_{ki} c_{ji}\int_{\beta\text{sphere}} d\tau\;\nabla G_j(r)\cdot\nabla G_k(\mathbf{r}) \quad (34)$$

Standard differentiation of the Gaussian primitives in eq 34 leads to eq 35, which yields for the *integrand* of a single product of two s functions s_j and s_k the following result

$$G_{jk} = 2\alpha_j\alpha_k K_{jk}\exp(-\alpha_{jk}|\mathbf{r}-\mathbf{R}_{jk}|^2)[(x-X_j)(x-X_k) + (y-Y_j)(y-Y_k) + (z-Z_j)(z-Z_k)] \quad (35)$$

Not surprisingly, the implementation of eq 34 can benefit from the integration framework as laid out above for multipole moments because eq 35 is similar to eq 29. Equally, no new issues arise in the implementation of eq 33, which leads to an integrand given in eq 36, again for two s functions,

$$K_{kin,jk} = K_{jk}\exp(-\alpha_j|\mathbf{r}-\mathbf{R}_{jk}|^2)\{3\alpha_k - 2\alpha_k^2[(x-X_k)^2 + (y-Y_k)^2 + (z-Z_k)^2]\} \quad (36)$$

K_{jk} on the right-hand side has the usual meaning (see eq 5c) and should not be confused with $K_{kin,kj}$. The actual equation that is implemented is slightly more complicated because the double differentiation must be symmetrically carried out, which means that eq 36 has a counterpart in which indices j and k are swapped.

Table 1. Effect of a Varying Threshold of Gaussian Pairs in the Integration of the Electron Density over the Volume of a β Sphere (Radius 1.1 Bohr) of a Carbon Atom in the Protein Crambin

run	threshold	% surviving	CPU (s)	ΔQ_{00} (e)
1	10^{-2}	1.0	319	2×10^{-3}
2	10^{-3}	1.5	466	3×10^{-4}
3	10^{-4}	2.0	630	7×10^{-7}
4	10^{-5}	2.6	812	2×10^{-7}
5	10^{-6}	3.3	1032	3×10^{-8}
6	10^{-7}	4.0	1260	0

The final integrand is then simply the average of eq 36 and its mirror counterpart.

3. EXAMPLE CALCULATIONS

A first simple test is setting the β sphere radius to a large value and observing whether an expected number of electrons is recovered. This is indeed the case for a β sphere of radius 8 Bohr centered on the oxygen in water. Because a total electronic charge of 9.999995e is obtained, the algorithm is able to cope with off-nuclear peaks of electron density, i.e., those on the hydrogen atoms.

A second concern is how close to zero the parameter R is allowed to be before the master equation (eq 17) can no longer be used. This was tested on a nitrogen atom's β sphere (radius again 1 Bohr) appearing in a HF/6-31G(d,p) wave function of capped glycine, including all possible pairs of Gaussian primitives. If $R \leq 10^{-15}$ Bohr, then an instability appears in the evaluation of eq 17, which calls for the alternative integration path, that is, setting R to zero and factorizing the integral of eq 10. On the basis of further testing, we decided to set the threshold for switching to the "$R = 0$ integration path" at 10^{-13} Bohr.

A third area of testing concerns the "Gaussian primitive cutoff" threshold. As discussed before, the prefactor K_{ij} (see eq 5c) should be prescreened such that integral evaluation can be avoided if K_{ij} is very small. As discussed in Appendix A (Supporting Information), the number of function evaluations increases quadratically in the analytical method. This increased number of function evaluations (that is, of the resulting integral) jeopardizes the intrinsic speed that the analytical method brings. We first demonstrate the algorithm on the small system of capped glycine computed at the HF/6-31G(d,p) level. The monopole of a β sphere with a radius of 1 Bohr inside the oxygen nearest to the α carbon (two bonds away) is 4.7905258e without prescreening; i.e., all Gaussian pairs contribute. Setting K_{ij} to values that increase by an order of magnitude each step, ranging from 10^{-7} to 10^{-1}, reduces the number of surviving Gaussian pairs from 100% to 55%, 51%, 46%, 39%, 33%, 26%, and 14%, respectively. The β sphere monopole is identical to the "unscreened result" (100% survival) of 4.7905258e when K_{ij} is set to 10^{-7}, and while it changes gradually as K_{ij} increases, it differs only by 0.00007e when K_{ij} is set to 10^{-3}. Tests on the nitrogen adjacent to the α carbon show that a K_{ij} value of 10^{-4} (39% survival) preserves the exact β sphere monopole of 3.840246 up to six digits. Rather coarse Gauss−Legendre quadrature grids, such as 35 × 15 × 15 points (radial, polar and azimuthal, respectively) already reproduce this target value with an error of 10^{-6} e. This makes the quadrature competitive in terms of CPU time usage compared to the analytical method. More work is needed to reduce the CPU time required in the current analytical implementation.

The effect of K_{ij} prescreening was also investigated in the case of a large system, where it is more critical for an integration to finish in a reasonable time. We adopted an experimentally determined conformation for the 642-nuclei and 2520-electron protein crambin. Even a small HF/3-21G wave function for this molecule already contained 5922 Gaussian primitives, which would lead to more than 17.5 million integral evaluations at 100% inclusion of Gaussian pairs. Fortunately, millions of integrals do not need to be evaluated, as is clear from Table 1. This table shows the effect of varying the threshold for the prefactor K_{ij} on the integration accuracy of the monopole moment of a carbon atom's β sphere in crambin. For every decrease by an order of magnitude of this threshold, starting at 10^{-2} and terminating at 10^{-7}, the percentage of "surviving" Gaussian primitives increases roughly by 0.5%. The CPU time to evaluate the surviving integrals increases by a factor of 1.2−1.4 for every decrease in the threshold. Times are quoted for an Intel Core 2 Due CPU E6750 @2.66 GHz and 3.25 Gb of RAM with a code developed with Compaq Visual FORTRAN Professional Edition 6.6.0 (2000). It is obvious that a reasonable monopole can be obtained already for a K_{ij} threshold of 10^{-3}. Clearly, the analytical code needs to be optimized in terms of execution time because a value of 466 CPU seconds (run 2) is not competitive with a Gauss−Legendre quadrature, which requires actually slightly less CPU time. In particular, the abundance of "if statements" in the current version of the code may cause delays, as well as inefficient evaluation of the error and/or gamma function. More work is planned on this important point. Further testing with alternative quadrature grids is also planned but in the context of a future study on atomic energies and atom−atom interaction energies. In particular, it has been shown[30] that the Lebedev grid[62] is more efficient than the Gauss−Legendre quadrature in the case of the β sphere.

Finally, tests on $Sn_2(CH_3)_2$ showed that coarse Gauss−Legendre quadrature grids are already producing accurate results for the monopole moment of Sn. For example, a grid of 40, 20, and 40 points in the radial variable, the polar angle and the azimuthal angle, respectively, yields a value of 46.4237851 e, which is only 7×10^{-6} e off the exact analytical value of 46.4237919 e. However, if the grid is further pruned to 25 × 15 × 30 points (radial, polar, and azimuthal, respectively), then the discrepancy is 0.003e.

There is evidence of more pronounced differences in favor of the analytical integration occur for the evaluation of kinetic energies associated with the β sphere. For example, the exact (100% survival) values for $K(\beta)$ and $G(\beta)$ for the same nitrogen in glycine as commented on above are 78.71901 and 87.22506 au, respectively. These results have been computed for a wave function in which all types of Gaussian primitives have been artificially set to s, due to a partial implementation so far. When K_{jk} is set to 10^{-6}, these values are reproduced to all digits at the cost of 2.2 CPU seconds. A grid of 35 × 15 × 15 points takes about 6 times as long and generates an error of 0.4 kJ/mol in $G(\beta)$ and 19 kJ/mol in $K(\beta)$. Only for a grid of 50 × 15 × 15 points do the discrepancies drop to 0.02 and 0.1 kJ/mol, respectively. This grid, which is now 8 times more expensive than the analytical method, introduces an error of 0.03 kJ/mol in $-1/4$ times the Laplacian of the electron density. Finally, a grid of 35 × 15 × 15 points for the β sphere of a carbon atom in crambin gives an error of 0.02 kJ/mol in $G(\beta)$ and 0.07 kJ/mol in $K(\beta)$ while taking about 10 times as long. Quadrature times are

quoted without introducing any exponential cutoffs nor dynamic thresholding.

4. CONCLUSIONS

We present a detailed derivation of a fully analytical 3D integration over the volume bounded by the β sphere inside a topological atom. The integrand is confined to the electron density multiplied by a real regular spherical tensor, which appears in the calculation of atomic multipole moments. We show that closed expressions can be obtained involving the exponential, error function, and gamma functions (and the beta function). In the current implementation, the analytical algorithm does not produce significant CPU savings compared to a Gauss−Legendre quadrature for monopole moments (i.e., "charges"). It is hoped that this first step toward a fully analytical integration can help in making topological energy partitioning, involving both kinetic and potential energy, more accurate and possibly faster.

■ ASSOCIATED CONTENT

Supporting Information. Appendices A−E are available free of charge via the Internet at http://pubs.acs.org.

■ ACKNOWLEDGMENT

Gratitude is due to Professor Peter Gill for pointing out the direct integration method that replaces our initial (and then abandoned) approach along the lines of the single-center expansion of Kauffmann and Baumeister. Thanks to Dr. Yongna Yuan for providing the crambin wave function.

■ REFERENCES

(1) Bader, R. F. W.; Anderson, S. G.; Duke, A. J. *J. Am. Chem. Soc.* **1979**, *101*, 1389.
(2) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, Great Britain, 1990.
(3) Bader, R. F. W. *J. Phys. Chem. A* **2009**, *113*, 10391.
(4) García-Revilla, M.; Popelier, P. L. A.; Francisco, E.; Martín-Pendás, A. *J. Chem. Theory Comput.* **2011**, *7*, 1704.
(5) Runtz, G. R.; Bader, R. F. W.; Messer, R. R. *Can. J. Chem.* **1977**, *55*, 3040.
(6) Bader, R. F. W.; Beddall, P. M. *J. Am. Chem. Soc.* **1973**, *95*, 305.
(7) Bader, R. F. W.; Beddall, P. M. *J. Chem. Phys.* **1972**, *56*, 3320.
(8) Bader, R. F. W.; Beddall, P. M.; Cade, P. E. *J. Am. Chem. Soc.* **1971**, *93*, 3095.
(9) Bader, R. F. W. *Acc. Chem. Res.* **1985**, *18*, 9.
(10) Popelier, P. L. A.; Aicken, F. M.; O'Brien, S. E. *Atoms in Molecules*. In *Chemical Modelling: Applications and Theory*; Hinchliffe, A., Ed.; Vol. 1 RSC Specialist Periodical Report; Royal Society of Chemistry: London, Great Britain, 2000; Chapter 3, pp 143−198.
(11) Popelier, P. L. A.; Smith, P. J. Quantum Topological Atoms. In *Chemical Modelling: Applications and Theory*; Hinchliffe, A., Ed.; Vol. 2, RSC Specialist Periodical Report; Royal Society of Chemistry: London, Great Britain, 2002; Chapter 8, pp. 391−448.
(12) Mills, M. J. L.; Popelier, P. L. A. *Comput. Theo. Chem.* **2011**, *975*, 42.
(13) Mills, M. J. L.; Popelier, P. L. A. *Theor. Chem. Acc.* **2011**in press.
(14) Handley, C. M.; Hawe, G. I.; Kell, D. B.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6365.
(15) Handley, C. M.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2009**, *5*, 1474.
(16) Biegler-Koenig, F. W.; Nguyen-Dang, T. T.; Tal, Y.; Bader, R. F. W.; Duke, A. J. *J. Phys. B* **1981**, *14*, 2739.
(17) Biegler-Koenig, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *3*, 317.
(18) Biegler-Koenig, F. *J. Comput. Chem.* **2000**, *21*, 1040.
(19) Biegler-Koenig, F. W.; Schoenbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545.
(20) Biegler-Koenig, F. W.; Schoenbohm, J. *J. Comput. Chem.* **2002**, *23*, 1489.
(21) Popelier, P. L. A. *Comput. Phys. Commun.* **1996**, *93*, 212.
(22) Popelier, P. L. A. *Mol. Phys.* **1996**, *87*, 1169.
(23) Popelier, P. L. A. *Chem. Phys. Lett.* **1994**, *228*, 160.
(24) *AIMAll*, Version 11.05.16; Todd A. Keith, http://aim.tkgristmill.com, 1997−2010.
(25) Stefanov, B. B.; Cioslowski, J. *J. Comput. Chem.* **1995**, *16*, 1394.
(26) Pendas, A. M.; Costales, A.; Luana, V. *Phys. Rev. B* **1997**, *55*, 4275.
(27) Otero-de-la-Roza, A.; Blanco, M. A.; Martín Pendás, A.; Luaña, V. 2009, *180*, 157.
(28) Pendas, A. M.; Blanco, M. A.; Francisco, E. *J. Chem. Phys.* **2004**, *120*, 4581.
(29) Blanco, M. A.; Pendas, A. M.; Francisco, E. *J. Chem. Theor. Comput.* **2005**, *1*, 1096.
(30) Otero-de-la-Roza, A.; Luana, V. *J. Comput. Chem.* **2011**, *32*, 291.
(31) Malcolm, N. O. J.; Popelier, P. L. A. *J. Comput. Chem.* **2003**, *24*, 1276.
(32) Malcolm, N. O. J.; Popelier, P. L. A. *Faraday Discuss.* **2003**, *124*, 353.
(33) Silvi, B.; Savin, A. *Nature (London)* **1994**, *371*, 683.
(34) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. *Comput. Chem.* **1999**, *23*, 597.
(35) Gatti, C.; Saunders, V. R.; Roetti, C. *J. Chem. Phys.* **1994**, *101*, 10686.
(36) Dovesi, R.; Orlando, R.; Civalleri, B.; Roetti, C.; Saunders, V. R.; Zicovich-Wilson, C. M. *Z. Kristallogr.* **2005**, *220*, 571.
(37) Souhassou, M.; Blessing, R. H. *J. Appl. Crystallogr.* **1999**, *32*, 210.
(38) Stash, A.; Tsirelson, V. *J. Appl. Crystallogr.* **2002**, *35*, 371.
(39) Katan, C.; Rabiller, P.; Lecomte, C.; Guezo, M.; Oisona, V.; Souhassou, M. *J. Appl. Crystallogr.* **2003**, *36*, 65.
(40) Rabiller, P.; Souhassou, M.; Katan, C.; Gatti, C.; Lecomte, C. *J. Phys. Chem. Solids* **2004**, *65*, 1951.
(41) Aray, Y.; Rodriguez, J.; Lopez-Boada, R. *J. Phys. Chem. A* **1997**, *101*, 2178.
(42) Henkelman, G.; Arnaldsson, A.; Jonsson, H. *Comput. Mater. Sci.* **2006**, *36*, 354.
(43) Tang, W.; Sanville, E.; Henkelman, G. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.
(44) Rodriguez, J. I.; Köster, A. M.; Ayers, P. W.; Santos-Valle, A.; Vela, A.; Merino, G. *J. Comput. Chem.* **2009**, *30*, 1082.
(45) Yu, M.; Trinkle, D. R. *J. Chem. Phys.* **2011**, *134*, 064111.
(46) Uberuaga, B. P.; Batista, E. R.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 10664.
(47) Rafat, M.; Popelier, P. L. A. *J. Comput. Chem.* **2007**, *28*, 2602.
(48) Popelier, P. L. A.; Bremond, E. A. G. *Int. J. Quantum Chem.* **2009**, *109*, 2542.
(49) Popelier, P. L. A. Quantum Chemical Topology: on Bonds and Potentials. In *Structure and Bonding. Intermolecular Forces and Clusters*; Wales, D. J., Ed.; Springer: Heidelberg, Germany, 2005; Vol. 115, p 1.
(50) Bohorquez, H. J.; Boyd, R. J.; Matt, C. F. *Int. J. Quantum Chem.* **2010**.
(51) Bohorquez, H. J.; Boyd, R. J. *Theor. Chem. Acc.* **2010**, *127*, 393.
(52) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, *132*, 6498.
(53) Biegler-Koenig, F. W.; Nguyen-Dang, T. T.; Tal, Y.; Bader, R. F. W.; Duke, A. J. *J. Phys. B* **1981**, *14*, 2739.
(54) Price, S. L.; Stone, A. J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987.
(55) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, Great Britain, 1996.
(56) Laidig, K. E. *J. Phys. Chem.* **1993**, *97*, 12760.

(57) Kaufmann, K.; Baumeister, W. *J. Phys. B* **1989**, *22*, 1.
(58) Solano, C. J. F.; Pendás, A. M.; Francisco, E.; Blanco, M. A.; Popelier, P. L. A. *J. Chem. Phys.* **2010**, *132*, 194110.
(59) Cruz, S. A.; Soullard, J. *Int. J. Quantum Chem.* **2001**, *83*, 271.
(60) Su, Z. W.; Coppens, P. *Acta Crystallogr.* **1994**, *A50*, 636.
(61) *Mathematica*; Wolfram Research, Inc.: Champaign, IL, U.S.A., 2008.
(62) Lebedev, V. I.; Laikov, D. N. *Dokl. Math.* **1999**, *59*, 477.