Contents lists available at ScienceDirect

Computers & Graphics

Special Section on MoIVA 2021

Topological analysis of density fields: An evaluation of segmentation methods

Alexei I. Abrikosov*, Talha Bin Masood, Martin Falk, Ingrid Hotz

Department of Science and Technology (ITN), Linköping University, Norrköping, Sweden

A R T I C L E   I N F O

A B S T R A C T

Topological and geometric segmentation methods provide powerful concepts for detailed field analysis and visualization. However, when it comes to a quantitative analysis that requires highly accurate geometric segmentation, there is a large discrepancy between the promising theory and the available computational approaches. In this paper, we compare and evaluate various segmentation methods with the aim to identify and quantify the extent of these discrepancies. Thereby, we focus on an application from quantum chemistry: the analysis of electron density fields. It is a scalar quantity that can be experimentally measured or theoretically computed. In the evaluation we consider methods originating from the domain of quantum chemistry and computational topology. We apply the methods to the charge density of a set of crystals and molecules. Therefore, we segment the volumes into atomic regions and derive and compare quantitative measures such as total charge and dipole moments from these regions. As a result, we conclude that an accurate geometry determination can be crucial for correctly segmenting and analyzing a scalar field, here demonstrated on the electron density field.

1. Introduction

Segmentation is a fundamental step in many visualization pipelines. When it comes to scalar density fields a common class of approaches build on topological concepts. However, despite the solid mathematical foundation, the performance of these methods varies a lot depending on the characteristic of the scalar fields. During the analysis of electronic charge density fields, we observed large differences in the segmentation results using different implementations of the same topological concepts which can have a severe impact on the visualization and the analysis results. This motivated us to perform a case study evaluating most accessible topological segmentation methods.

The electronic charge density plays a central role in the analysis of molecules and crystals, e.g. to compute atomic volumes and charges. The use of geometric and topological analysis for segmentation and visualization of the properties of the electronic charge density has been gaining popularity not only in the visualization community but among chemists and physicists as well. One of the pioneering works on the application of topology within the scientific domain is *"Atoms in Molecules"* by Bader [1] describing the principles of dividing the charge density between atoms. How-

ever, simple Voronoi segmentation has been used for this purpose as well [2,3]. The expectation placed on such analysis is that it will provide insight into the properties of molecules and materials that are otherwise either difficult or impossible to determine using other analysis and computational methods. So far, the main focus of topological analysis of the charge density has been in the form of atomic charge determination [4,5] and interaction/bond analysis between atoms [6–9]. This work focuses on the determination of atomic charge based on topological and geometric segmentation of the volume.

The goal of our work is to evaluate commonly used algorithms and models for the segmentation of the electronic charge density field which are used to compute atomic charges and dipole moments as well as their use for visualization. This entails at first the evaluation of the geometric accuracy of available algorithms as solutions to the underlying models. This is especially interesting for methods based on combinatorial topology. Secondly, we evaluate the models used for segmentation in comparing the quantitative values for charge and dipole moments to the expectations from Chemistry. Lastly, we inspect the properties of the segmentation concerning symmetry preservation which is essential for the generation of reliable visualizations.

The field at the center of this evaluation is the electronic charge density $\rho(\mathbf{r})$. It is an observable, meaning that it can be measured in an experiment or computed theoretically. The total charge is not

* Corresponding author.
  E-mail address: alexei.abrikossov@liu.se (A.I. Abrikosov).

an observable but it can be extracted by analyzing the field $\rho(\mathbf{r})$. This is done by segmenting the total volume of the system into atomic regions and integrating over these regions to determine the atomic charges [1,4,5]. These charges will then interact with each other by means of electrostatic potential. However, charges determined in this way are bad at reproducing the electrostatic potential [10,11]. A way to improve the description of the electrostatic potential is by utilizing a multipole expansion where a charge distribution is described not as a single point charge but as a sum of several terms including charge (q), dipole moment (μ), and further terms of the expansion[12]. Therefore, we extended the analysis of the electronic charge density to include the dipolar contribution as well. By doing so we are also able to grasp the anisotropy of $\rho(\mathbf{r})$ around individual atoms, introducing directional interactions since the dipole moment is a direct measure of this. Thus, to achieve reliable results, an accurate geometry of atomic segments becomes very important.

Methods for computing segmentation that have been developed in the field of quantum chemistry are largely inspired by the work from Bader [1] using numerical integration. Often they are specifically targeted to the electronic density field as a result of the main solvers used in quantum chemistry. The method has also faced some criticism for not being able to handle complex chemical structures. Therefore, we wanted to see if the methods from computational topology could perform better. The methods that have been developed in the field of computational topology provide generic methods with a focus on a robust extraction of the topological skeleton. The geometric embedding, however, is often not very accurate. This observation has already been made in earlier work and a few approaches tackling this problem have been proposed. At first, Reininghaus et al. [13] and Gyulassy et al. [14] proposed stochastic methods to obtain better geometric embedding. Later Gyulassy et al. introduced a Morse–Smale complex that conforms to both an input scalar field and an additional prior segmentation of the domain [15,16]. With the TopoMS framework, the problem has been targeted in the context of electron density fields by Bhatia et al. [5], which is a method that is also used in our comparison. We compare different approaches for the segmentation that are readily available or easily implementable: combinatorial discrete Morse theory as implemented in TTK [17] referred to as DISCRETEMS; a numerical segmentation proposed by Henkelman et al. [4]; TopoMS [5]; and Voronoi diagrams.

Three types of chemical/physical systems are used for this comparison: a copper (Cu) crystal structure, an ionic crystal of NaCl (table salt), a molecular crystal of CO_2 (dry ice), as well as individual molecules of water, benzene, and *p*-nitroaniline (PNA). All systems are well known by the scientific community, so it is easy to judge whether our results are reasonable. Our test data for the crystals were generated with the VASP package [18] using the `mp-30_Cu`, `mp-22862_NaCl` and `mp-20066_CO2` entries in the materials project database [19] for structure information and computational scheme. The molecular data was generated by using the GAUSSIAN package [20].

2. Background

Electronic charge density $\rho(\mathbf{r})$ is an observable charge distribution in a unit of volume. A common way to compute the density field $\rho(\mathbf{r})$ is by utilizing one of the many density functional theory (DFT) packages with the most popular being VASP [18] and GAUSSIAN [20]. These packages generate a discrete 3D grid to represent the charge density distribution as a scalar field in some arbitrary total volume V_{tot}. In our case, we focus mostly on data-grids generated by VASP. Due to the theory used to perform the DFT calculations, the VASP software can generate two types of density fields. The first and main field type only describes the charge dis-

tribution of the valence electrons (the outermost electrons) since only the valence electrons participate in chemical reactions. A field with both the valence and core electrons taken into account can be generated by post-processing the results of the DFT calculations. The difference between the two is that the valence-only field provides a more accurate description of the density field. However, because it lacks the description of core electrons there are cavities around the atoms (Fig. 1(a,b)). This leads to a more complex topological structure of the density field and makes the direct determination of the correct maximum, associated with the atomic position, impossible without utilizing additional algorithms. This in turn hinders a correct segmentation. On the other hand, the grid that considers both, core and valence electrons, has very distinct maxima (Fig. 1(c,d)) at the atomic positions which helps in computing better segmentation. However, the description of the electronic charge density field is less accurate and leads to errors when computing charges; a fact statement that will be evident in the results section.

Basic algorithm of the analysis. The basics of the topological analysis for atomic charge determination are straightforward.

- Determine the atomic volume for each atom in the molecule.
- Integrate over the volume to determine the atomic charge.

The main challenge for this approach from a chemical point of view is how to draw borders that separate atoms. The most mathematically sound suggestion comes from Bader [1] who, based on the Smale theory, proposed to draw the border between atoms along a surface of zero flux in the gradient of $\rho(\mathbf{r})$ satisfying

$$\nabla \rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0 \tag{1}$$

where $\mathbf{n}(\mathbf{r})$ is the normal vector to the surface. One can observe these surfaces appearing in Fig. 1.

Multipole expansion When it comes to dealing with a charge distribution one can rewrite the effective charge of such distribution as a point property by utilizing the multipole expansion. The total charge q_{tot} is the first term of this expansion and is defined as:

$$q_{tot} = \sum q_i \tag{2}$$

with q_i being the partial charge of index *i*. The second term of the expansion is the total dipole moment of the charge distribution μ_{tot} defined as:

$$\mu_{tot} = \sum q_i \mathbf{r}_i \tag{3}$$

with $\mathbf{r}_i$ being the directional vector to q_i. In principle, the multipole expansion has an infinite amount of terms but in practice, one rarely needs to go beyond the second term of the expansion. For the purposes of our study, the dipole moments are calculated with the atom positions as the origin for the directional vector $\mathbf{r}_i$.

Data sets The choice of the data sets that are used in this work is motivated either by the use of the same systems in the work by Henkelman et al. [4] or being so well studied within the chemistry community that most of the properties of these systems can be considered common knowledge. Using these simple examples allows one to have a clear understanding of what was the expected result should be and how this result differs from the results obtained with the help of the different segmentation methods.

In this section, we give a short introduction to the expected properties of the different systems, although most of them are primarily used to compare the resulting atomic segments and not so much the chemical properties. In order to have a broad range of test data sets, we chose to test both crystals and single molecules.

- NaCl is an ionic crystal of a monovalent salt. That is, the Na atom will donate an electron to the Cl atom leading to the formation of Na^+ and Cl^- ions with the formal charge of 1*e* and

(a) Gradient lines in valence $\rho(\mathbf{r})$ (b) A slice in valence $\rho(\mathbf{r})$ (c) Gradient lines in full $\rho(\mathbf{r})$ (d) A slice in full $\rho(\mathbf{r})$
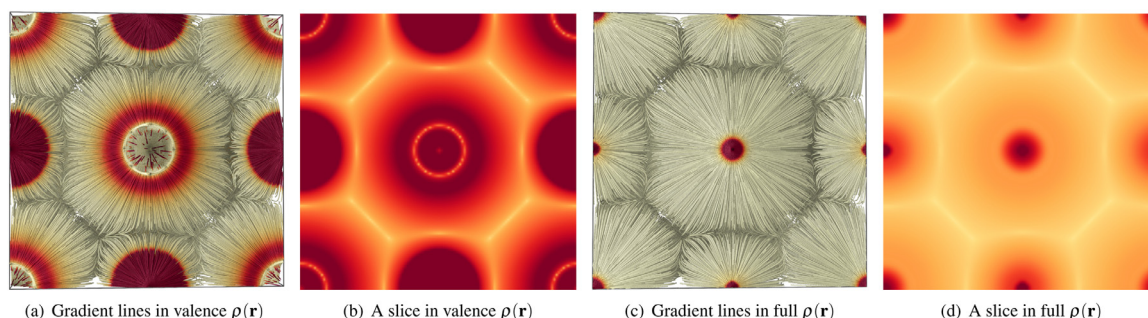
Fig. 1. These images demonstrate the structure of the charge density field $\rho(\mathbf{r})$ of NaCl by (a,c) showing gradient field lines and (b,d) the gradient magnitude (yellow for low and red for high values). It can be seen that the field provides a natural segmentation of the domain in regions that can be associated with the atoms. The images (a,b) show the valence only and (c,d) the full electronic charge density field. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$-1e$ respectively, where e is the elementary charge of an electron. In contrast, the paper uses [e] to mean the number of electrons.

- CO_2 molecular crystal: In CO_2, the O atoms will cause a polarization of the electronic structure along the C=O bonds. This local polarization was the main reason to choose this system. But because CO_2 is a linear molecule the symmetry of the molecule will cancel all electrostatic interactions that are present within the molecule. Thus, the second reason for the choice of this system was to test if the different segmentation methods could capture this behavior.
- Cu is a typical metal with a simple crystal structure.
- H_2O molecule is probably one of the most recognizable molecules to non-chemists and was chosen for its structural simplicity to ease the visualization of atomic segments.
- The benzene molecule is chosen for its highly symmetric planar ring structure and in the scope of this work mostly used as an example of segmentation.
- p-nitroaniline (PNA) is a derivative of the benzene molecule also used as an example of segmentation.

Additionally, it is expected that all the atoms of the same type (e.g. all Na atoms in NaCl) feature the same segmentation shape due to being indistinguishable. Thus, if one would get shapes that geometrically differ for the same atom type, it is a sign that something is not correct. Even if the inaccurate segmentation does not always have a big impact on the derived total charges nor the total dipole moments, it will generally not be accepted by the domain scientists due to the missing symmetries.

3. Segmentation

In recent years, there have been numerous ways proposed to compute the atomic volumes both by numerical determination of the gradient [4,21–23] and by computing the Morse–Smale complex [24]. We use the numerical code that is provided by the group of Henkelman and also the visualization software Inviwo [25] which integrates the Topology Toolkit (TTK) [17]. Here, we utilize TTK for computing the Morse–Smale segmentation. Another tool developed in the visualization community is TopoMS [5], which combines a numerical approach for volume segmentation with a Morse–Smale analysis for determining the molecular graph. Finally, for the sake of comparison, a geometry based segmentation utilizing Voronoi diagrams is also considered as a part of our evaluation since Voronoi diagrams have also previously been used for the computation of atomic charges by chemists [2,3]. Another key feature that most of these methods, except for weighted Voronoi, have is that they do not require any preexisting knowledge about the data which is crucial when exploring novel materials.

3.1. Numerical gradient based approaches

Numerical segmentation by Henkelman. This approach is based on a numerical analysis also known as the Bader analysis [1]. The idea behind this analysis is that one can draw natural borders between atoms along the surfaces with zero gradient cross flow, where $\rho(\mathbf{r})$ satisfies (1). The second property of $\rho(\mathbf{r})$ is that it exhibits a maximum at the atomic positions. Based on these assumptions, the original algorithm proposed by Henkelman et al. [4] starts at the vertices of the grid and follows the numerically determined gradient between the grid vertices until reaching a maximum. All points visited along the way are saved. When all grid points have been visited, they are assigned to segments that correspond to the corresponding maximum. However, this method was quickly deemed unsatisfactory due to the "grid bias" that the determined volumes were displaying. To remove this bias a near-grid method, the current standard method, has been developed [21,22]. It is still based on the principles of the original idea of going from grid point to grid point by utilizing the central finite difference scheme. However, a correction vector is introduced to keep track of the accumulated error when traversing from one point to another. Once the error, that is the vector's magnitude, exceeds a threshold a correction step is made toward a grid point in the direction of the vector and then resetting the vector. For a more detailed explanation of the methods please refer to the original publications especially the one by Tang et al. [21]. We used Version 0.95a of the code provided by the Henkelman group for this study.

TopoMS is a hybrid method combining a numerical segmentation with the Morse–Smale determination of the molecular graph. The method utilizes central differences for the gradient determination with a tri-linear interpolation and an adaptive Euler integrator to trace the integral lines. This method also utilizes an adaptive step by estimating the error of the integration. The method overcomes the problem of having multiple maxima for an atom by assigning each maximum to the closest atom. The process of finding the nearest atom is accelerated using a kd-tree data structure.

Since the scope of this paper is to compare the segmentation of the atomic volumes determined by the different methods, we will treat TopoMS as a numerical method in this case. It does stand to point out that unlike the Henkelman software, TopoMS is well integrated with VTK [26] thus making it a superior software in terms of the topological visualization of molecules.

3.2. Discrete gradient approach

Morse–Smale complex. Given a smooth scalar field $f : M \to \mathbb{R}$ defined on a manifold M, a point $p_c \in M$ is called a *critical point* of f if the gradient ∇f_{p_c} is zero.

An *integral line* is a path in M which follows the gradient direction. The set of all integral lines originating at a critical point p_c along with p_c is called the *ascending manifold* of p_c. Similarly, the set of all integral lines with destination at a critical point p_c along with p_c is called the *descending manifold* of p_c. The *Morse–Smale complex* is the decomposition of M into regions with uniform gradient flow behaviour, that is, it is a partition such that within each cell the set of integral lines share a common origin and destination. This partition can be obtained as the intersection of the ascending and descending manifolds of the critical points. For more details about the computation of Morse–Smale complex, we refer to [27,28], and [29].

In the context of analyzing charge density fields, the features of interest are maxima as they correspond to the atomic positions. The descending manifolds of the maxima provide the atomic segmentation. For computing the Morse–Smale complex and its persistence driven simplification [30], we utilize the Topology Toolkit [17]. We refer to this method of segmenting charge density field as DISCRETEMS in this paper.

3.3. Geometric approaches

Since the position of the atoms are available for the charge density fields, we also consider a purely geometric approach for the segmentation. We use the *Voronoi diagram* and its weighted version which is also referred to as *power diagram* [31] for this purpose.

Voronoi diagram. Given a set of seed points $S = \{p_i\} \in \mathbb{R}^d$, the Voronoi diagram partitions the space based on proximity to the seed points. The Voronoi cell of point p_i consists of the points $p \in \mathbb{R}^d$ which are closer to point p_i than to any other seed points $p_j \in S, j \neq i$.

Usually, the distance measure used is the Euclidean distance. In case the seed points have different weights (consider balls with different radii), the definition of Voronoi diagram extends naturally to weighted seed points with the *power distance* measure [31]. The power distance between a point $p \in \mathbb{R}^d$ and a weighted input seed point p_i with radius r_i is defined as $\mathrm{pd}(p_i, p) = \|p - p_i\|^2 - r_i^2$. We compute the segmentation of the density fields using Voronoi diagrams, the Voronoi cell being the segment corresponding to an atom, both using the Euclidean and power distance measure. For the weighted Voronoi approach, the weight was assigned based on the atomic radii.

4. Results

Here, we present the results for the different approaches that we tested concerning visual appearance and geometric robustness. In Section 4.2, we perform a simple visual analysis of the segmented volumes. This is followed by a test on the methods' sensitivity to the change of resolution of the data in Section 4.3. Section 4.4 explores how well the different methods represent the local anisotropy that can arise in some chemical systems. We also perform a quantitative comparative analysis in Section 4.5. Finally, we visually summarize the overall relation between segmentation and physical properties in Section 4.6.

4.1. Data

As described earlier in Section 2, we use six different representative data sets for this study: three molecules and three periodic crystal structures. They are simple in terms of number of atoms; nonetheless they are well suited to demonstrate the general behavior. The molecular data sets are for H_2O (water), C_6H_6 (benzene)

and $C_6H_6N_2O_2$ (p-nitroaniline or PNA), while the crystals considered in this study are NaCl (table salt), CO_2 (dry ice) and Cu (Copper metal) crystals.

For the three molecular data sets, $\rho(\mathbf{r})$ is generated using the GAUSSIAN software [20], while VASP [18] is used for generating $\rho(\mathbf{r})$ for the three crystal data sets. A major difference between the fields $\rho(\mathbf{r})$ generated by GAUSSIAN and VASP is that GAUSSIAN generates only the valence electron field $\rho(\mathbf{r})$, while VASP generates a valence $\rho(\mathbf{r})$ and a core electron $\rho(\mathbf{r})$ which then can be summed up to generate a full electron density field $\rho(\mathbf{r})$ through post processing.

4.2. Visual comparison

We start the comparison of the different segmentation methods by comparing them visually. We use our molecular data sets for this study since they are small and simple unlike the periodic crystal data sets, where the periodicity impedes understanding purely by visual observation. The resulting segmentation can be seen in Fig. 2. The three molecules on display are water Fig. 2(a–e), benzene (f–j), and PNA (k–o).

We start with water, a molecule that consists of only three atoms. A first intuitive approach is to draw a border in the geometric middle between the atoms resulting in the Voronoi segmentation (Fig. 2(d)). However, this approach does not take into account that atoms can differ in size. Adding this domain knowledge leads to the solution proposed by the weighted Voronoi method (Fig. 2(e)) giving a more realistic solution of the location of the borders between the atoms. While the Voronoi segmentation is conceptually simple and easily implementable, both versions draw borders as straight planes which are not necessarily representative of a real system. The weighted Voronoi segmentation requires in addition prior knowledge of the system to provide good weights. It is worth mentioning that finding good weights is not trivial since the atomic radius of an atom can change drastically depending on its surroundings; atomic radii found in the literature provide only a crude approximation for the weight. Thus, a method that does not require much prior knowledge of the system yet that can draw realistic borders between atoms would be desirable.

This leads to the use of topological segmentation methods that conceptually are purely based on the data without prior knowledge of the system. However, depending on the chosen implementation, this is not always the case. Numerical segmentation as implemented in TOPOMS and Henkelman's Bader analysis software requires the knowledge of the atomic positions (which is not a problem for computationally produced $\rho(\mathbf{r})$). Combinatorial methods based on Forman's discrete Morse theory [32] as implemented in TTK [17] takes only the density field as an input.

Fig. 2 (b,c) shows the resulting segmentation of the two numerical methods. Here, one can see that the borders of the hydrogen atoms become more spherical protruding into the volume occupied by the oxygen atom.

Lastly, DISCRETEMS draws the border between the atoms in a way that the symmetry of the two indistinguishable hydrogen atoms is lost (Fig. 2(a)). Their volumes have a unique shape and differ in size. The other two data sets, benzene and PNA, included in Fig. 2 also illustrate the different unintuitive shapes of atomic volumes inside the molecule. Looking closely at the PNA molecule in Fig. 2(l,m) one can notice that the volume associated with the hydrogen atoms (white) is different depending on whether they are connected to a carbon atom (gray) or nitrogen atom (blue). Since atomic radii are influenced by their surroundings this can be another problem for the weighted Voronoi method as this introduces even more parameters when determining weights.
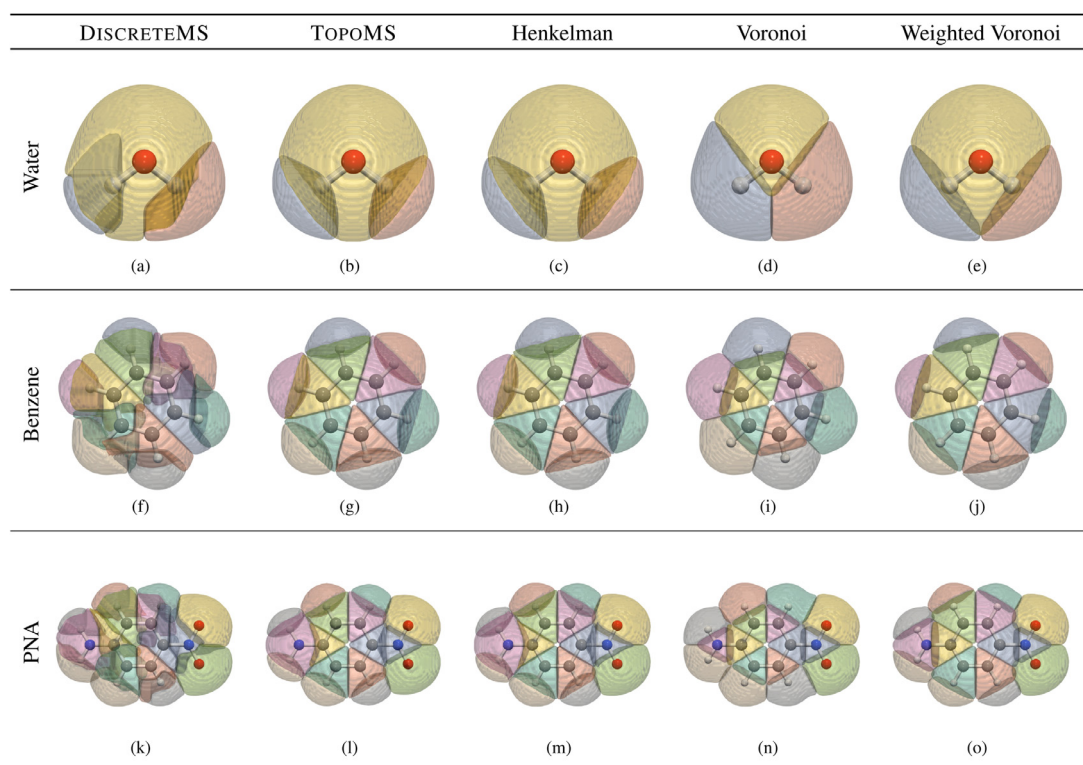
Fig. 2. Comparison of the segmentation for three molecules (top: H_2O, middle: Benzene, bottom: PNA) using the different segmentation methods. The atom colors follow the CPK coloring scheme: hydrogen (white), oxygen (red), carbon (gray), and nitrogen (blue) while the different segments are colored differently to aid the visualization of the segmentation borders between atoms. Please note the unsymmetrical nature of the DISCRETEMS segmentation (a,f,k) and planar cuts between the atoms in Voronoi segmentation (c,d,h,i,m,n). In contrast, numerical segmentations (b,c,g,h,l,m) preserve both the symmetry and have the more intuitively understandable atomic borders. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Charge determination of NaCl

In the following we evaluate the segmentation methods with respect to the accuracy of the aggregated total charges that are associated with each atom. We start with the analysis of NaCl (table salt). NaCl is an ionic crystal meaning that in this test case one electron from the Na atom will jump over to the Cl atom.

At first we compare the partial charges of the NaCl crystal for different grid resolutions of density field $\rho(\mathbf{r})$ to evaluate the stability of the methods.

The resulting atomic volume and partial charge for the Cl atom can be seen in Fig. 3. It can be seen that both the mean charge and mean volume using DISCRETEMS and the Henkelman methods are more stable as a function of the grid size than the other methods for both the mean charge and mean volume. Also note that both the Voronoi and weighted Voronoi substantially underestimate both the charge and volume compared to the other methods. For the weighted Voronoi the ionic radius for Cl in crystals was used as the weight.

Secondly, we compare the partial charges with the theoretically expected values. Overall, the DISCRETEMS and Henkelman method give similar results. The number of electrons on the Cl atom is around 7.86 giving the Cl atom a net charge of about $-0.86e$. This is about what one would expect for a crystal since in such systems electrons are shared between atoms to a larger degree than in for example an aqueous solution where the expected charge would be closer to $-1e$. It is important to point out that in all the above methods we used a combination of the full and the valence field to determine the charge. The combination consists of determining the segments by utilizing the full charge distribution field and using the data from the valence field to compute the charge based on those segments.

On the other hand, TOPOMS uses only one charge density field as input for its analysis and it displayed some interesting results. Firstly, if we look at the results from the full charge density data set we see little difference from DISCRETEMS and Henkelman when it comes to the determined volumes. We can, however, clearly see that the accuracy of the electronic charge data is far off from the expected value of somewhere close to $-1e$, especially for coarser grid sizes. The results seem to converge to the expected result at finer grid sizes but it is clear that the full charge density fields are unfit to be used to determine the partial charges of a system. On the other hand when using only the valence $\rho(\mathbf{r})$ to determine the charge, everything seems to be fine except for one of the grid resolutions for which the segments are not correctly determined. Even if the errors for the coarse grid sizes can be understood, the sudden loss of accuracy at the grid size of 200^3 is concerning.

The multipole expansion for NaCl does not yield any dipole moment on the atoms which is expected for such a system, meaning that the electrons are uniformly distributed around the atoms.

4.4. Multipole determination of CO_2

For the evaluation of the anisotropy of the $\rho(\mathbf{r})$ in the segments, the CO_2 crystal is well suited since it exhibits a dipole moment along the bonds from the oxygen to the carbon atom. The data set consists of a periodic structure of a CO_2 crystal with the periodic cell containing four CO_2 molecules at a grid resolution of 150^3. An example of the segmented atomic volumes and the resulting dipole moments for a single CO_2 molecule can be seen in the supplementary material. Since each cell contains four indistinguishable CO_2 molecules, we expect to see no difference in the results between the different individual molecules.
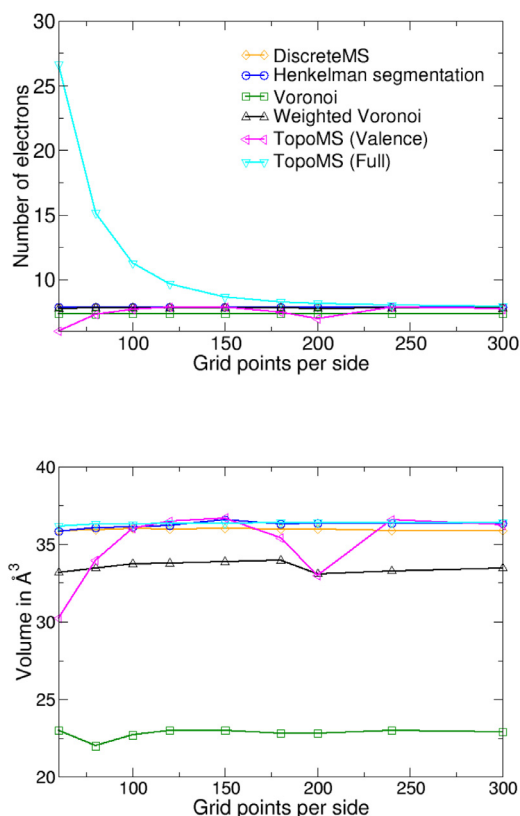
Fig. 3. NaCl comparison of the difference in atomic volume and charge of the Cl atom (number of electrons) between the numerical, DISCRETEMS, and Voronoi segmentation methods as a function of the grid size.

Fig. 4 illustrates the dipole moments of the data set computed using all approaches. We also expanded on the TOPOMS analysis by combining the TOPOMS full density field segmentation with the data from the valence density field mimicking the procedure used

in the other methods. Comparing the results, one can notice that there is a noticeable discrepancy between the methods. First, there is a difference in the size of the arrows between the segmentation methods which in itself was not unexpected. However, the fact that for the DISCRETEMS method, the strength of the dipoles, represented as the size of the arrows, varies a lot for identical molecules reveals a severe issue with the accuracy of the segmentation.

We now further compare the numerical values of the total charge and dipole moments in Tables 1 and 2. The charge on the individual atoms is given by the number of electrons per atom [e] and the dipole moment is given in electron Ångström [eÅ]. Starting with the value for the total charge we see that the values from the Henkelman method are uniform across the same atoms (C-atom 1.87e, O-atom 7.07e). The values for the DISCRETEMS method are, except for C3, O5, O6, quite consistent with values of 1.74e for C and 7.14e for O. Both Voronoi methods, though mostly uniform, clearly underestimate the amount of electrons that the carbon donates to the oxygen in comparison to other methods with C 3.49e and O 6.26e for Voronoi and C 3.82e and O 6.09e for weighted Voronoi, respectively. In this case, the weight for weighted Voronoi was the radius of C and O when the atoms form a covalent double bond. The TOPOMS (valence) has an issue with detecting carbon atoms and assigning the full charge to the oxygen atoms. The TOPOMS (full) is again suffering from the issue of bad data in the full $\rho(\mathbf{r})$ data set. Since we need to post-process the segmentation for the multipole expansion, we add an analysis scheme that is not included by default in the TOPOMS software. That is a segmentation on full density field $\rho(\mathbf{r})$ and electronic data from the valence data set. By doing so, the TOPOMS produced segmentation leads to results similar to those achieved by the Henkelman segmentation yielding 1.88e for C and 7.06e for O, which is expected since now the two analysis methods become virtually identical.

By looking at the charge determination results alone, one might conclude that the numerical and DISCRETEMS methods are quite similar and there are plenty of ionic salt data sets like NaCl leading to this conclusion. There are however outliers, namely the CO_2 molecule consisting of the aforementioned C3, O5, O6. Clearly, something did not go well for this segmentation. The differences
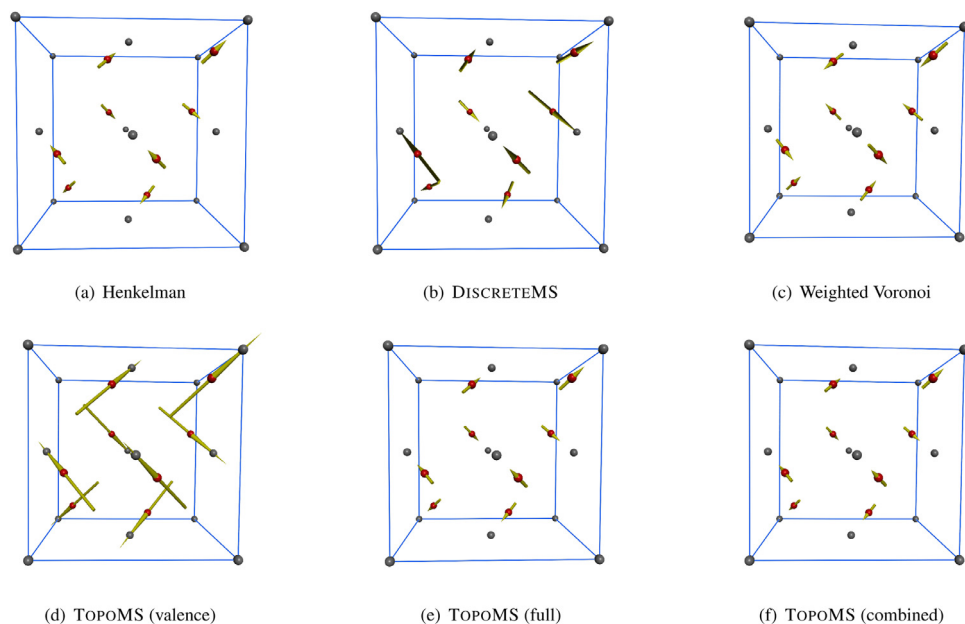

Fig. 4. Calculated dipole moments for all the atoms in CO_2 carbon (black) and oxygen (red) using the (a) Henkelman, (b) DISCRETEMS (TTK), (c) weighted Voronoi, (d) topoMS valence $\rho(\mathbf{r})$, (e) topoMS full $\rho(\mathbf{r})$, and (f) topoMS (combined) segmented data sets. The size of the arrow indicates the strength of the dipole moment and the arrow points along the direction of the dipole vector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Resulting charge of the CO_2 crystal. The charge is given in electrons [e].

Atom	Henkelman	DISCRETEMS	Voronoi	Weighted Voronoi	TOPOMS (Val)	TOPOMS (Full)	TOPOMS (Combined)
C1	1.87	1.74	3.49	3.82	0	3.68	1.88
C2	1.87	1.74	3.49	3.82	0	3.68	1.88
C3	1.87	0.76	3.49	3.82	0	3.68	1.88
C4	1.87	1.74	3.49	3.82	0	3.68	1.88
O1	7.07	7.14	6.26	6.09	8.00	9.16	7.06
O2	7.07	7.14	6.26	6.09	8.00	9.16	7.06
O3	7.07	7.14	6.26	6.09	8.00	9.16	7.06
O4	7.07	7.14	6.26	6.09	8.00	9.16	7.06
O5	7.07	7.61	6.26	6.09	8.00	9.16	7.06
O6	7.07	7.61	6.26	6.09	8.00	9.16	7.06
O7	7.07	7.14	6.26	6.09	8.00	9.16	7.06
O8	7.07	7.14	6.26	6.09	8.00	9.16	7.06

Table 2
Resulting dipole of the CO_2 crystal. The strength of the dipole moments are given in electron Ångström [eÅ].

Atom	Henkelman	DISCRETEMS	Voronoi	Weighted Voronoi	TOPOMS (Valence)	TOPOMS (Full)	TOPOMS (Combined)
C1	0	0	0	0	0	0	0
C2	0	0	0	0	0	0	0
C3	0	0	0	0	0	0	0
C4	0	0	0	0	0	0	0
O1	0.31	0.40	0.27	0.36	1.21	0.33	0.28
O2	0.31	0.40	0.27	0.36	1.21	0.33	0.28
O3	0.31	0.40	0.27	0.36	1.21	0.33	0.28
O4	0.31	0.40	0.27	0.36	1.21	0.33	0.28
O5	0.31	0.85	0.27	0.36	1.21	0.33	0.28
O6	0.31	0.85	0.27	0.36	1.21	0.33	0.28
O8	0.31	0.40	0.27	0.36	1.21	0.33	0.28

become even clearer when looking at the dipole moments of the individual atoms, for a visual representation see Fig. 4. All methods yield a 0 dipole moment for the C atom. The dipole on the O atom is significantly larger for the DISCRETEMS segmentation (Fig. 4(b)) at 0.40 compared to 0.31 to the Henkelman segmentation (Fig. 4(a)), which given the information presented above can be considered the most robust method. The discrepancy in the computed dipole moments for O5 and O6 is even larger: 0.85 based on DISCRETEMS compared to 0.31 based on Henkelman segmentation. Another aspect of the dipole moment is that it has a direction. In the case of the CO_2 molecule, two dipole moments point from the O atom towards the C atom. Thus, we can make an easy check on how good the direction of our computed dipole moments is. For the numerical method, the difference in angle between the computed dipole vector and the vector between O and C atom is $0°$. On the other hand for the DISCRETEMS method, the difference is $13°$, except for the O5, O6 case where the difference is $0°$. Thus, as a quick summary, the DISCRETEMS method overestimates the strength of the dipole and does not obtain the correct directionality. If we combine the definition of the dipole moment in Eq. (3) with the visual comparison from Section 4.2 the reason why DISCRETEMS struggles with determining uniform results for the dipole moments and charges in the case of the CO_2 data set should become clear.

On the other hand, both the Voronoi and weighted Voronoi methods (Fig. 4(c)) yield results that are similar to that of the numerical methods in the strength of the dipole moment, however it seems that the mistake on the directionality of the dipole moment is $180°$. This is a very surprising number since one would not expect such a large shift of the direction if the geometry resulting from a segmentation is off. But if one considers that the electrons are not uniformly distributed within the total volume, and that the bond between the C and O is highly polarized (the electrons from the O shift over to C) a small error in computing the segment geometry can lead to a significant error in higher order moments. For a more detailed explanation, please refer to Fig. 2 in the supplementary material.

Another example of this can be seen in the results for TOPOMS valence segmentation (Fig. 4(d)). Here we have a case where the O atom segments completely swallow the C atom segments giving rise to an extremely bloated dipole moment. On the other hand, the TOPOMS results for the full segmentation (Fig. 4(e)) are comparable to Henkelman. This indicates that the correct geometry of the volume is more important for the determination of the dipole moment than the electronic values of individual voxels. Finally, the combined TOPOMS results (Fig. 4(f)) are comparable to the Henkelman segmentation. Again, this is expected since the two methods become virtually identical. Also, the direction of the dipole moment for all the TOPOMS cases is in accordance with the vector between the C atom and O atom indicating an overall good direction determination in TOPOMS.

4.5. Quantitative comparison

To compare the segmentation methods more quantitatively, we use the *Jaccard similarity coefficient* [33] for the analysis. The Jaccard coefficient between two sets A and B is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

We use this measure to compare the similarity between the segments S_a^i and S_b^i computed for an atom i using the methods a and b. Then the *volume similarity* J_v between these segments is computed as:

$$J_v(S_a^i, S_b^i) = \frac{Volume(S_a^i \cap S_b^i)}{Volume(S_a^i \cup S_b^i)}$$

Since for our application the total charge within each segment is of higher importance, we also use a weighted version of the Jaccard similarity coefficient to quantify the charge similarity between two segments. The *charge similarity* J_c between the segments S_a^i and S_b^i is computed as:

$$J_c(S_a^i, S_b^i) = \frac{\int_{S_a^i \cap S_b^i} \rho \, dv}{\int_{S_a^i \cup S_b^i} \rho \, dv}$$

Table 3

Average J_v and J_c scores for all pairs of segmentation methods. The volume similarity J_v is reported above the diagonal of this table, while charge similarity J_c is displayed below the diagonal. The colored dots indicate scores exceeding 90 % (●) and lower than 60 % (●).

	Molecule	DiscreteMS (full)	TopoMS (full)	Henkelman	Weighted Voronoi	Voronoi	DiscreteMS (val)	TopoMS (val)
DiscreteMS (full)	NaCl		92.08 %	92.24 %	86.46 %	53.81 %	98.44 %	92.07 %
	CO$_2$		70.93 %	68.68 %	58.65 %	61.11 %	61.88 %	57.67 %
	Cu		96.38 %	97.54 %	96.44 %	96.44 %	99.61 %	95.93 %
	H$_2$O		-	-	-	-	-	-
	Benzene		-	-	-	-	-	-
	PNA		-	-	-	-	-	-
	Average		**82.22** %	**81.35** %	**74.22** %	**64.56** %	**80.36** %	**75.51** %
TopoMS (full)	NaCl	99.42 %		98.51 %	85.21 %	52.79 %	91.21 %	97.13 %
	CO$_2$	84.27 %		93.73 %	67.56 %	71.33 %	55.74 %	59.37 %
	Cu	99.41 %		97.49 %	98.65 %	98.65 %	96.11 %	98.58 %
	H$_2$O	-		-	-	-	-	-
	Benzene	-		-	-	-	-	-
	PNA	-		-	-	-	-	-
	Average	**91.85** %		**95.95** %	**78.63** %	**69.70** %	**74.29** %	**78.49** %
Henkelman	NaCl	99.44 %	99.86 %		85.62 %	53.17 %	91.29 %	96.07 %
	CO$_2$	83.86 %	98.77 %		65.18 %	68.02 %	54.05 %	56.25 %
	Cu	99.61 %	99.59 %		97.56 %	97.56 %	97.26 %	97.05 %
	H$_2$O	-	-		74.96 %	42.55 %	67.15 %	97.77 %
	Benzene	-	-		75.27 %	63.73 %	66.47 %	96.73 %
	PNA	-	-		74.43 %	64.42 %	66.50 %	91.84 %
	Average	**91.68** %	**99.27** %		**75.93** %	**64.63** %	**69.66** %	**86.46** %
Weighted Voronoi	NaCl	98.99 %	98.92 %	98.95 %		61.10 %	85.17 %	83.20 %
	CO$_2$	67.03 %	70.79 %	70.48 %		93.25 %	46.42 %	47.76 %
	Cu	99.42 %	99.78 %	99.60 %		100.00 %	96.17 %	98.23 %
	H$_2$O	-	-	70.31 %		52.61 %	65.87 %	76.15 %
	Benzene	-	-	54.66 %		49.13 %	57.42 %	74.13 %
	PNA	-	-	63.56 %		59.18 %	58.76 %	70.95 %
	Average	**83.08** %	**85.00** %	**71.26** %		**67.31** %	**62.73** %	**70.63** %
Voronoi	NaCl	93.34 %	93.23 %	93.29 %	94.22 %		53.07 %	51.70 %
	CO$_2$	70.52 %	74.79 %	74.36 %	93.67 %		48.03 %	49.43 %
	Cu	99.42 %	99.78 %	99.60 %	100.00 %		96.17 %	98.23 %
	H$_2$O	-	-	45.87 %	38.47 %		47.55 %	43.45 %
	Benzene	-	-	74.66 %	45.11 %		58.18 %	65.35 %
	PNA	-	-	71.70 %	57.10 %		57.17 %	64.65 %
	Average	**82.94** %	**85.10** %	**76.69** %	**69.97** %		**57.11** %	**60.88** %
DiscreteMS(val)	NaCl	99.87 %	99.34 %	99.36 %	98.86 %	93.22 %		91.51 %
	CO$_2$	60.93 %	59.48 %	59.36 %	50.42 %	52.27 %		94.23 %
	Cu	99.94 %	99.37 %	99.56 %	99.38 %	99.38 %		95.65 %
	H$_2$O	-	-	77.99 %	62.14 %	49.02 %		67.46 %
	Benzene	-	-	81.32 %	48.91 %	72.40 %		66.27 %
	PNA	-	-	80.58 %	56.35 %	69.94 %		63.70 %
	Average	**80.41** %	**79.42** %	**80.08** %	**63.06** %	**71.01** %		**77.50** %
TopoMS (val)	NaCl	99.20 %	99.55 %	99.44 %	98.50 %	92.85 %	99.16 %	
	CO$_2$	60.61 %	59.71 %	59.49 %	50.48 %	52.34 %	99.04 %	
	Cu	99.33 %	99.77 %	99.51 %	99.71 %	99.71 %	99.29 %	
	H$_2$O	-	-	98.05 %	69.88 %	46.55 %	78.48 %	
	Benzene	-	-	97.55 %	54.68 %	75.61 %	81.02 %	
	PNA	-	-	97.40 %	63.38 %	72.53 %	80.82 %	
	Average	**79.93** %	**79.67** %	**89.65** %	**66.77** %	**72.31** %	**88.72** %	

For each chemical system in our data sets and every combination of methods (a, b), we can now compute the average of J_v scores for the atoms. These average J_v scores are reported for all six chemical systems in Table 3 corresponding to the entries above the diagonal. Similarly, we compute the average J_c scores which are reported in the bottom left triangle. We further compute the overall average J_v and J_c scores over all the atoms in every system in an attempt to quantitatively capture the agreement and disagreement between a pair of methods (a, b) using a single summary score. Although this average score hides a lot of system and atomic level intricacies, it does provide a useful measure of overall similarity and differences between the methods. Note that J_v and J_c scores reported here allow one to quantify how similar or different the segmentations obtained from two methods are. These scores are not a measure of the accuracy of the methods against some ground-truth segmentation as such a segmentation is not known for these data sets.

For a comprehensive atomic level comparison using J_v and J_c scores, refer to the Fig. 8-14 in the supplement. From this quantitative comparative analysis, the following points are worth mentioning:

- In general, $J_c > J_v$ applies to any pair of methods. That is, even if the geometric volumes determined by the two methods differ, the charge within that volume may not differ significantly. This is largely due to the fact that the segments differ in the regions where $\rho(\mathbf{r})$ is very low due to drawing different separating boundaries within the flat plateau region where $\rho(\mathbf{r}) \approx 0$. As long as the two methods are correctly separating the regions with high $\rho(\mathbf{r})$, they will have a high J_c score. This is particularly evident in the case of the Voronoi segmentation of the NaCl crystal. Even though $J_v \approx 53\%$ for the Voronoi based approach compared to other methods, the J_c is consistently above 92%.
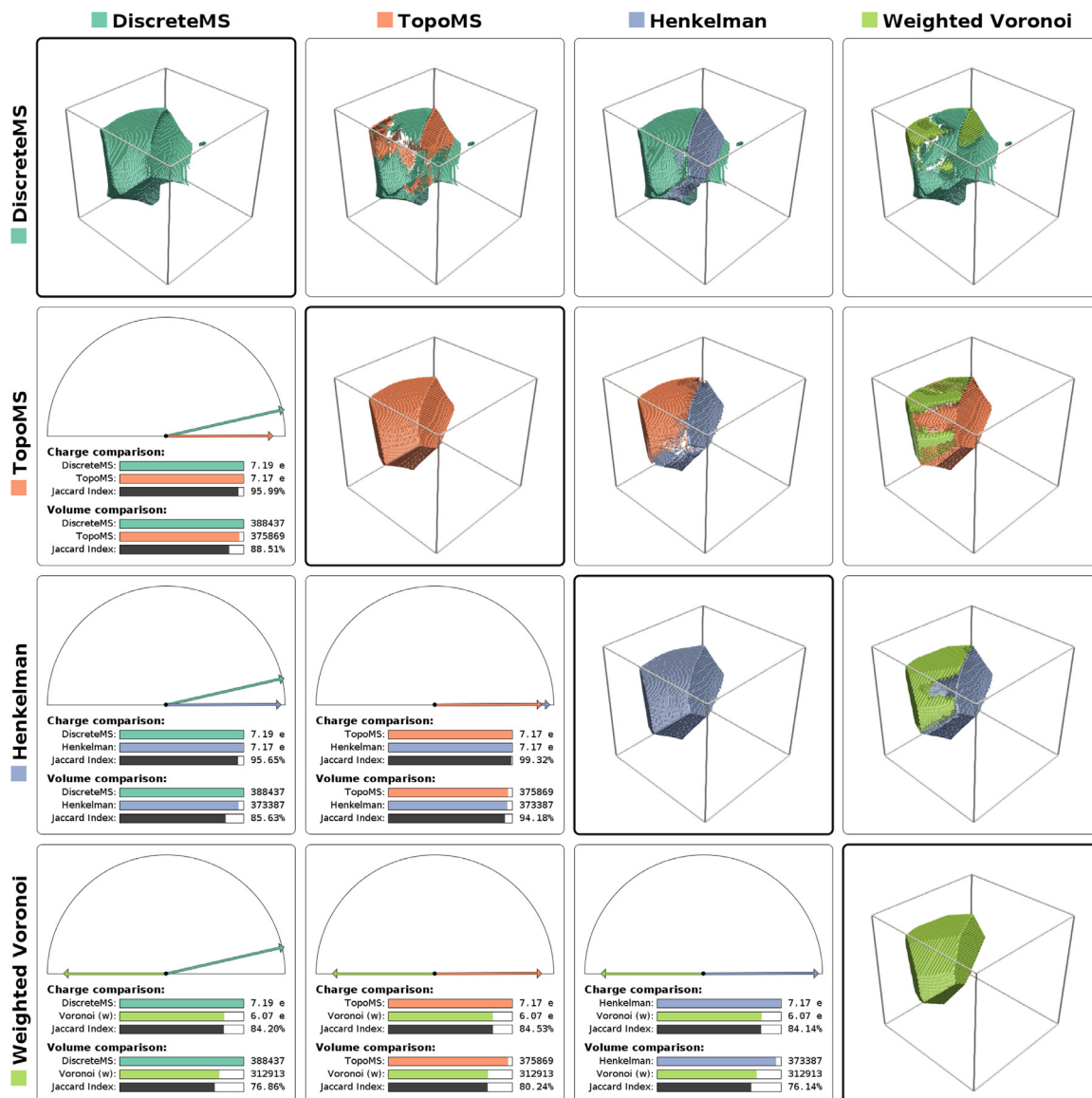
Fig. 5. Comparison of segmentation for an oxygen atom in a CO_2 crystal lattice. The diagonal shows the segments generated by the different methods. Above the diagonal, we can observe the symmetric difference of the segments computed by a two different methods. Below the diagonal we display the difference in the determined physical properties.

- The best scores of $J_v = 95.95\%$ and $J_c = 99.27\%$ were obtained for the (Henkelman, TopoMS) combination using full $\rho(\mathbf{r})$. This is expected because both methods use a numerical gradient approach for the segmentation. However, it is still worth pointing out that there is not a 100% agreement between these approaches. Furthermore, it also matters whether the full or valence $\rho(\mathbf{r})$ is used for the segmentation. The agreement between these two approaches goes down to ($J_v = 86.46\%$, $J_c = 88.72\%$) if a valence $\rho(\mathbf{r})$ field is used for segmentation. This drop in score is largely due to TopoMS being unable to correctly segment the carbon atoms within the CO_2 system.
- All methods agree very well on the segmentation of metallic crystals as evident from the very high J_c and J_v scores for the Cu crystal. This can be explained by the uniform periodic nature of this system containing only one type of atom. Geometric approaches based on the Voronoi diagram work very well for such systems.
- All gradient based approaches, namely the Henkelman method, TopoMS, and DISCRETEMS do a good job of segmenting ionic

crystal systems as demonstrated by the results for NaCl system. The purely geometric Voronoi based approach of segmenting by drawing boundaries exactly in the middle of the atoms fails for this system as unlike the Cu crystal, because it contains two atoms of different sizes. As a result, the Voronoi method has $J_v \approx 53\%$ when compared to other methods. The segmentation can be improved using a weighted version of the Voronoi diagram with J_v improving to $\approx 85\%$ as a consequence. This, however, is still below the $\approx 92\%$ agreement between gradient based approaches.
- The segmentation obtained using DISCRETEMS, the discrete gradient approach as implemented in TTK, differs significantly from the numerical approaches as used in the Henkelman method and TopoMS. This is especially the case for molecular systems with covalent bonds, namely CO_2, H_2O, Benzene, and PNA. The J_v score between DISCRETEMS and TopoMS is $\approx 67\%$ for these molecules while it is $\approx 80\%$ between DISCRETEMS and the Henkelman method. The agreement between DISCRETEMS and numerical approaches improves to $J_v > 90\%$

and $J_c \approx 99\%$ when only non-covalently bonded systems are considered.

4.6. Detailed atomic-level comparative visualization

We provide a visualization approach which allows a comprehensive comparison of the segmentation obtained for an atom using different approaches. Refer to Fig. 5 as an example. We use a matrix to display various results. Each column or a row in the matrix corresponds to one of the segmentation approaches. The segment obtained for the atom using approach a is displayed on the diagonal of the matrix within the panel (a, a) of the matrix. The symmetric difference between the segments is displayed in the upper right triangle of the matrix. The dipole vector comparison and other quantitative measures are depicted in the lower left triangle of the same matrix. This visualization approach supports a deeper look at the differences and similarities between the methods.

As an example, we compare the results from one oxygen atom of the CO_2 data set in Fig. 5. Also for the sake of readability we only compare the DiscreteMS, TopoMS, Henkelman and weighted Voronoi methods in this figure. For the complete comparison showing all the approaches, we would like to refer the reader to Fig. 3-7 in the supplement. All methods use the full $\rho(\mathbf{r})$ for the segmentation but the data from valence $\rho(\mathbf{r})$ for charge and dipole moment computation. From Fig. 5, it is clear that small differences in the segmentation can lead to very drastic changes in the physical properties that one would like to determine. For example, although the volumes as determined by the Voronoi and numerical gradient approaches do not differ much with a similarity score of $\approx 84\%$, the dipole vectors are reversed. This is explained when we look at the volume difference between the Voronoi and Henkelman approaches or the Voronoi versus TopoMS volume difference. We can observe the whole segment as determined by Voronoi approach is shifted to the top left which is in the direction away from the carbon atom in CO_2. Due to this shift in the segment, the dipole moment shifts from pointing towards the carbon atom to pointing away. Refer to Fig. 2 in the supplement for more detailed explanation of this observed change in the orientation of the dipole vector. The second interesting and relevant observation one can make from Fig. 5 is the fact that the dipole vector as computed using DiscreteMS is not aligned to the direction vector of the C=O bond between carbon and oxygen.

5. Conclusion

While it is commonly known that the geometric embedding of segmentation based on combinatorial methods is not very good, the advantage of providing a robust and accurate topological structure is rated of higher importance. In this work, we showcase the implications a bad geometric embedding can have with respect to the analysis of a scalar field. Therefore, we have utilized the example of electronic density distribution fields and we have compared a set of commonly used segmentation methods in this domain. The methods evaluated include (i) combinatorial topology, (ii) numerical segmentation based on the gradient flow, and (iii) purely geometric methods based on Voronoi segmentation which requires domain knowledge for the segmentation while ignoring properties of the density field. Since the dependence on domain knowledge during the analysis is increasing, the methods are becoming less generic.

Our results have confirmed that the geometry of topological segmentation can play a significant role in the determination of properties of scalar field data sets. This is particularly the case for data sets where integral measures become important in terms of volume or anisotropy analysis of the segments.

In more detail, we can confirm that the boundaries of the segmentation for the combinatorial approach, here DiscreteMS, can be far off from the expected location. This is especially serious in areas where the field exhibits a plateau-like behavior. This behavior can be seen in many applications; in our case this concerns the regions in between the atoms where the field takes low values close to zero. One can argue that those are regions that are not of the highest interest and that the corresponding segmentation is not stable and, thus, do not affect the topological structure, which is represented accurately. However, one can also argue that these artifacts are severe and even make the use of discrete topological segmentation methods unacceptable in some applications. As soon as the segmentation represents a physical property, both volume and shape become important. Our case is one example but there are also other applications with similar demands, for example the segmentation of CT scans to determine the physical properties of a material or the analysis of the shape and integrals over burning cells in combustion simulation [34]. In our specific application, we can observe that the computed volume associated with an atom for the different segmentation methods varies significantly. But the total charge associated with an atom is not as much affected due to the low field values in the miss-segmented area and, thus, does not contribute much to the total charge. The impact on the dipole moment that measures the anisotropy of the charge distribution in a segment is stronger. The Voronoi segmentation, which is independent of the electronic charge density, often introduces a strong bias in the segmentation while maintaining the underlying symmetries well. In some cases, the weighted Voronoi segmentation can provide a reasonable approximation for the total charge, compare Table 1.

Besides the implications of an accurate quantification, there is a second argument that is directly related to the visualization of the results. In our experience, domain scientists tend to reject a visualization that does not respect the inherent symmetries of the field. This can be that atoms playing exactly the same role, for example the atoms in H_2O, as shown in Fig. 2(a), are assigned asymmetric volumes. Or in the case of NaCl, the crystal symmetry of each segment is not preserved. In conclusion, we note that in the case of electronic charge density there have been some efforts in the domain to achieve accurate segmentation of the volume. Those methods are, however, often very specific to the application and make use of the domain knowledge, like the location of the atoms. These methods do not scale very well with increasing complexity of the atomic or molecular structure and are also not applicable to other domains. There are also some methods proposed in context with combinatorial topology that are not yet generic enough, thus, leaving significant avenues for future research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Alexei I. Abrikosov: Conceptualization, Investigation, Formal analysis, Writing - original draft, Visualization. **Talha Bin Masood:** Formal analysis, Investigation, Writing - original draft, Visualization. **Martin Falk:** Writing - original draft, Visualization. **Ingrid Hotz:** Conceptualization, Writing - original draft, Supervision, Funding acquisition.

Acknowledgements

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cag.2021.05.015

References

[1] Bader RF. Atoms in molecules. Accounts Chem Res 1985;18(1):9–15.

[2] Guerra CF, Handgraaf J-W, Baerends EJ, Bickelhaupt FM. Voronoi deformation density (VCC) charges: assessment of the Mulliken, Bader, Hirshfeld, Weinhold and VDDD methods for charge analysis. J Comput Chem 2004;25:189–210.

[3] Cho M, Sylvetsky N, Eshafi S, Santra G, Efremenko I, Martin JML. The atomic partial charges arboretum: trying to see the forest for the trees. ChemPhysChem 2020;21:688–96.

[4] Henkelman G, Arnaldsson A, Jónsson H. A fast and robust algorithm for Bader decomposition of charge density. Comput Mater Sci 2006;36(3):354–60.

[5] Bhatia H, Gyulassy AG, Lordi V, Pask JE, Pascucci V, Bremer P-T. TopoMS: comprehensive topological exploration for molecular and condensed-matter systems. J Comput Chem 2018;39(16):936–52.

[6] Günther D, Boto RA, Contreras-Garcia J, Piquemal J-P, Tierny J. Characterizing molecular interactions in chemical systems. IEEE Trans Vis Comput Graph 2014;20(12):2476–85.

[7] Boto RA, Contreras-García J, Tierny J, Piquemal J-P. Interpretation of the reduced density gradient. Mol Phys 2016;114(7-8):1406–14.

[8] Morgenstern A, Wilson T, Miorelli J, Jones T, Eberhart M. In search of an intrinsic chemical bond. Comput Theor Chem 2015;1053:31–7.

[9] Wilson TR, Rajivmoorthy M, Goss J, Riddle S, Eberhart ME. Observing the 3D chemical bond and its energy distribution in a projected space. ChemPhysChem 2019;20(24):3289–305.

[10] Wiberg KB, Rablen PR. Comparison of atomic charges derived via different procedures. J Comput Chem 1993;14(12):1504–18.

[11] Kramer C, Spinn A, Liedl KR. Charge anisotropy: where atomic multipoles matter most. J Chem Theory Comput 2014;10(10):4488–96.

[12] Böttcher CJF, Bordewijk P. Theory of electric polarization, 2. Elsevier Science Limited; 1978.

[13] Reininghaus J., Günther D., Hotz I., Weinkauf T., Seidel H.P.. Combinatorial gradient fields for 2D images with empirically convergent separatrices. arXiv:1208.65232012.

[14] Gyulassy A, Bremer P-T, Pascucci V. Computing Morse–Smale complexes with accurate geometry. IEEE Trans Vis Comput Graph 2012;18(12):2014–22.

[15] Gyulassy A, Günther D, Levine JA, Tierny J, Pascucci V. Conforming Morse–Smale complexes. IEEE Trans Vis Comput Graph 2014;20(12):2595–603.

[16] Gyulassy A, Bremer P-T, Pascucci V. Shared-memory parallel computation of Morse–Smalecomplexes with improved accuracy. IEEE Trans Vis Comput Graph 2018;25(1):1183–92.

[17] Tierny J, Favelier G, Levine JA, Gueunet C, Michaux M. The topology toolkit. IEEE Trans Vis Comput Graph 2017;24(1):832–42.

[18] Kresse G, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 1996;6(1):15–50.

[19] Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson Ka. The materials project: a materials genome approach to accelerating materials innovation. APL Mater 2013;1(1):011002.

[20] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 16 revision C.01. Wallingford CT: Gaussian Inc; 2016.

[21] Tang W, Sanville E, Henkelman G. A grid-based Bader analysis algorithm without lattice bias. J Phys 2009;21(8):084204.

[22] Sanville E, Kenny SD, Smith R, Henkelman G. Improved grid-based algorithm for Bader charge allocation. J Comput Chem 2007;28(5):899–908.

[23] Yu M, Trinkle DR. Accurate and efficient algorithm for Bader charge integration. J Chem Phys 2011;134(6):064111.

[24] Edelsbrunner H, Harer J, Natarajan V, Pascucci V. Morse–Smale complexes for piecewise linear 3-manifolds. In: Annual symposium on computational geometry (SCG '03); 2003. p. 361–70.

[25] Jönsson D, Steneteg P, Sundén E, Englund R, Kottravel S, Falk M, Ynnerman A, Hotz I, Ropinski T. Inviwo—a visualization system with usage abstraction levels. IEEE Trans Vis Comput Graph 2019;26(11):3241–54.

[26] Schroeder WJ, Lorensen B, Martin K. The visualization toolkit: an object-oriented approach to 3d graphics. 4th ed. Kitware; 2006.

[27] Gyulassy A, Bremer P-T, Hamann B, Pascucci V. A practical approach to Morse–Smale complex computation: scalability and generality. IEEE Trans Vis Comput Graph 2008;14(6):1619–26.

[28] Robins V, Wood PJ, Sheppard AP. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. IEEE Trans Pattern Anal Mach Intell 2011;33(8):1646–58.

[29] Shivashankar N, Senthilnathan M, Natarajan V. Parallel computation of 2D Morse–Smalecomplexes. IEEE Trans Vis Comput Graph 2012;18(10):1757–70.

[30] Zomorodian A, Carlsson G. Computing persistent homology. Discrete Comput Geom 2005;33(2):249–74.

[31] Aurenhammer F, Klein R, Lee D. Voronoi diagrams and delaunay triangulations. World Scientific; 2013.

[32] Forman R. A user's guide to discrete Morse theory. In: International conference on formal power series and algebraic combinatorics. In: Advances in Applied Mathematics; 2001.

[33] Levandowsky M, Winter D. Distance between sets. Nature 1971;234(5323):34–5.

[34] Bremer P-T, Weber GH, Pascucci V, Day MS, Bell JB. Analyzing and tracking burning structures in lean premixed hydrogen flames. IEEE Trans Vis Comput Graph 2010;16(2):248–60.