LLNL-TR-704318

Understanding Lithium Solvation and Diffusion through Topological Analysis of First-Principles Molecular Dynamics

H. Bhatia, A. Gyulassy, M. Ong, V. Lordi, E. Draeger, J. Pask, V. Pascucci, P. T. Bremer

October 4, 2016

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Understanding Lithium Solvation and Diffusion through Topological Analysis of First-Principles Molecular Dynamics

Harsh Bhatia, Attila Gyulassy, Mitchell Ong, Vincenzo Lordi, Erik Draeger,
John Pask, Valerio Pascucci, and Peer-Timo Bremer

September 27, 2016

Abstract

The performance of lithium-ion batteries is strongly influenced by the ionic conductivity of the electrolyte, which depends on the speed at which Li ions migrate across the cell and relates to their solvation structure. The choice of solvent can greatly impact, both, the solvation and diffusivity of Li ions. In this work, we present our application of the topological techniques to extract and predict such behavior in the data generated by the first-principles molecular dynamics simulation of Li ions in an important organic solvent – ethylene carbonate. More specifically, we use the scalar topology of the electron charge density field to analyze the evolution of the solvation structures. This allows us to derive a parameter-free bond definition for lithium-oxygen bonds, to provide a quantitative measure for bond strength, and to understand the regions of influence of each atom in the simulation. This has provided new insights into how and under what conditions certain bonds may form and break. As a result, we can identify and, more importantly, predict, unstable configurations in solvation structures. This can be very useful in understanding when small changes to the atoms' movements can cause significantly different bond structures to evolve. Ultimately, this promises to allow scientists to explore lithium ion solvation and diffusion more systematically, with the aim of new insights and potentially accelerating the calculations themselves.

1 Introduction

Lithium-ion batteries are one of the most common portable energy sources, which are currently used to power many small hand-held consumer devices as well as larger machinery such as electric vehicles. Typically batteries consist of an anode and a cathode separated by an electrolyte. In the case of Lithium-ion batteries, the electrolyte usually consists of a Lithium (Li) salt dissolved in an organic liquid. It is well known that the choice of electrolyte is one of the major factors influencing the performance of such batteries [20]. Different electrolytes cause different solvation structures of Li-ions and directly impact how the ions diffuse from the anode to the cathode. However, some of the processes involved are not yet well understood, making it difficult to design better electrolytes to improve battery performance [17].

While the battery performance of different electrolytes can be observed experimentally, it is difficult to deduce any information about the underlying molecular physics from these studies. Instead, large-scale computer simulations are typically used to simulate Li diffusion to gain new insights. Of particular interest are first-principles molecular dynamics simulations, so called because they do not assume any physical model of the diffusion process, but act entirely on the quantum-mechanical principles. Assuming such systems can be accurately modeled, the macro-level behaviors of solvation and diffusion evolve naturally from the quantum-mechanical interactions. However, such simulations are computationally expensive, and even using some of the largest super-computers currently operating, can be performed for only hundreds of atoms and on time-scales of tens of picoseconds. For example, the simulations that are the focus of this paper, consisting of 638 atoms with only a single Li-ion simulated for 40 picoseconds, used approximately 550,000 CPU hours and 2 months of wall-clock time using 384 compute cores.

The challenge is that due to the stochastic nature of the process, the small set of atoms, and the short (in real-time) duration, even such a large-scale simulation effort may only yield a small number of reconnection events. These are changes in the solvation structure (the bonds) of the Li-ion which are of greatest interest. As will be discussed in more detail below, the Li-ion predominately exists in a stable solvation structure – bonded to four neighboring oxygen (O) atoms – which yields little information not already known. However, in rare circumstances the Li-ion switches solvation structures by breaking and reforming bonds in a reconnection event. Understanding the causes and mechanics of such reconnections is integral to understanding the underlying physical phenomena. In particular, understanding how often one would expect such reconnections to occur can help determine the macro-scale diffusion coefficients of the simulated materials. Furthermore, this would allow model-based simulations of significantly bigger molecular systems, which could dramatically accelerate our search for better electrolytes and batteries as a whole.

At a high level, a better understanding of reconnection events requires an in-depth understanding of the bonds formed by the Li ion. This includes detecting whether a bond exists as well as understanding when a bond may break and how to potentially predict a change in bond structure. The latter is particularly interesting to domain scientists as it may allow a slight, directed pertubation to a simulation state to alter and create new events in a shorter period of time. For example, there may exist periods of instability in which multiple possible bond structures could occur depending on small variations in atom positions and speeds. If these could be identified, it might allow restarting a simulation with a slightly-altered state to create additional reconnection events. Intuitively, this would correspond to exploring multiple possible futures, as small-scale changes of position and velocity are random in nature.

Existing analysis of such simulations often defines bonds via a distance threshold that is carefully adjusted by hand to different chemical systems. The bond cutoffs may be chosen to correspond to rough measures of chemical bonding, such as minima in pair correlation functions of typical systems, but general transferability of these cutoffs can be problematic. However, there exists no single correct set of thresholds, and different thresholds can produce quite different results. Nonetheless, lengths of bonds between atoms and angles between bonds are often analyzed [17] with limited success.

Here, we propose to use topological analysis techniques developed within the visualization community to: 1) develop a more stable and effectively parameter-free technique to define bonds; and 2) derive various additional quantities such as bond strengths and regions of influence of atoms. In particular, we analyze a simulation of the solvation of a $LiPF_6$ molecule in ethylene carbonate (EC) molecules as electrolyte. Our team includes domain experts as well as analysis and high-performance computing experts to tackle the various challenges this application area presents. We demonstrate how topological analysis and the corresponding visualizations produce new insights and may provide a key to better understand, predict, and potentially alter reconnection events.

Contributions. We propose to use the Morse-Smale complex [5, 8] of the *electron charge density field* to identify and predict the solvation behavior of Li ions. The electron charge density field [1, 15] and its derivatives [4, 10] have previously been used to detect both covalent and non-covalent bonding structures, respectively. Thus, while our choice of this scalar field is not novel itself, it is particularly suitable for the molecular system under investigation. Furthermore, while many of the constructs presented here exist in theory, the lack of a practical realization of the theory, and understanding of the stability of the bonds defined this way, often creates hindrance. One of the contributions of this paper is that it presents, for the first time, an in-depth analysis of the stability of this topologically-driven bond definition. In particular, we show that, while there exist some discretization artifacts and noise that require simplification, the topological definition of bonds is stable over a wide range (orders of magnitude) of parameters, and is, thus, effectively parameter free. This observation is crucial for this bond definition to be accepted in the application domain. Furthermore, we provide a better understanding of various factors that may influence the formation of bonds. For example, the topological analysis of the electron charge density field allows understanding the regions of influence of various atoms around Li – an insight currently not possible using the standard distance-based analysis techniques. Similarly, our analysis allows the scientists to, for the first time, observe the oscillations on the Li–O bond strengths, and understand how these oscillations play an integral role in determining whether or not a new bond may be formed. Finally, we believe that our analysis enables the prediction of a bond switch, at a much coarser temporal resolution, thus creating the possibility to perform guided simulations with the ability to change the configurations in order to create the required conditions for a particular oxygen atom to bond with Li. While the analysis of altered simulations is beyond the scope of this paper, we believe that this can provide a new type of stability analysis of chemical configuration, as well as to enable scientists to understand the behavior in-situ and "branch out" a simulation at a lower cost.

2 Background and Related Work

This section provides the relevant background in molecular systems, and discusses some important techniques for their analysis and visualization. Followed by that, a discussion on the topological constructs used in our analysis is given.

2.1 Molecular Interactions and Their Analysis

The interactions between different atoms are characterized through *bonds*. Different types of chemical bonds exist, including *covalent* (electron-sharing), ionic (electrostatic), metallic (delocalized), dispersive (e.g., Van der Waals), hydrogen bonds, mixed bonds, and others, each associated with different quantum-mechanical features. One common way of defining bonds is via a distance-based approach. The bond cutoffs may be chosen to correspond to rough measures of chemical bonding, such as minima in pair correlation functions of typical systems, but general transferability of these cutoffs can be problematic. Various ideas beyond distance cutoffs have also been proposed to analyze bonding.

In particular, the *electron charge density (ECD)* function [15] has been successfully used in defining covalent interactions. The ECD is a quantum-mechanical quantity that gives the relative probability of finding an electron at a particular point in space. In this context, Bader [1] decomposed the charge density into regions of uniform gradient each associated with one atom, and the bonds are then related to the 2-saddles of the field. Gyulassy et al. [8] used the topology of the ECD to identify the atoms and bonds in a C_4H_4 molecule and orbitals of a hydrogen atom. The topology of the electron localization function [18] has also been shown to correspond to molecular interactions [19]. *Non-covalent* bonds, which do not involve sharing of electrons, typically cannot be detected using the ECD functions. To extract such interactions, derived metrics, such as signed electron density and reduced gradient field have been successfully used [10]. Recently, Günther et al. [4] described a combinatorial algorithm to perform a joint analysis of these two fields to extract both covalent and non-covalent interactions. In this work, we focus our interest on Li–O bonds, which are of mixed character, but which can still be described reliably using only the scalar ECD function.

2.2 Topological Analysis using Morse-Smale Complex

The following provides a brief introduction to the Morse-Smale complex, which is used in this paper to identify bonding structures, and topological simplification, used to study the stability of bonds.

Morse function. Given a smooth function $f : \mathbb{M} \to \mathbb{R}$, defined over a compact d-manifold $\mathbb{M}$, a point $p \in \mathbb{M}$ is called a *critical point* if the gradient of f at p equals zero ($|\nabla f(p)| = 0$). A critical point is *nondegenerate* if the Hessian, i.e., the matrix of second partial derivatives at the point, is invertible. The function f is a *Morse function* if all its critical points are nondegenerate and no two critical points have the same function value. In this case, the *Morse Lemma* [14] states that there exist local coordinates around p such that $f(p)$ can be represented as the following *standard form*: $f_p = \pm x_1^2 \pm x_2^2 \cdots \pm x_d^2$. The number of minus signs in this equation gives the *index* of critical point p. In three-dimensional functions, minima are index-0, 1-saddles are index-1, 2-saddles are index-2, and maxima are index-3.

Morse-Smale (MS) complex. An integral line in f is a path in $\mathbb{M}$ whose tangent vector agrees with the gradient of f at each point along the path. The integral line passing through a point p is the solution to $\frac{\partial}{\partial t}L(t) = \nabla f(L(t)), \forall t \in \mathbb{R}$, with initial value $L(0) = p$. Each integral line has an origin and destination at critical points of f, at $t = \pm\infty$. *Ascending* and *descending* manifolds are obtained as clusters of integral lines having common origin and destination respectively. The descending manifolds of f form a cell complex that partitions $\mathbb{M}$; this partition is called the *Morse* complex. Similarly, the ascending manifolds also partition $\mathbb{M}$ in a cell complex. A Morse function f is a *Morse-Smale function* if ascending and descending manifolds of its critical points only intersect transversally. An index-i critical point has an i-dimensional descending manifold and a $(d-i)$-dimensional ascending manifold. The simply-connected cells formed by the intersections of ascending and descending manifolds form the cells of the *Morse-Smale (MS) complex.* A 3-dimensional MS complex is a cell complex where cells of dimension zero through three are called nodes, arcs, quads, and crystals, respectively. Each arc is a 1-manifold bounded by two nodes, 0-manifolds, each quad is a 2-manifold bounded by arcs, and finally, each crystal of the MS complex is bounded by quads. Cells of the MS complex satisfy several combinatorial properties: end points of arcs are critical points whose indices differ exactly by one; quads contain exactly four arcs on their boundary (although some might be repeated); and the closure of the boundary of a crystal contains a collection of quads, arcs, saddles and exactly one minimum and one maximum. The *1-skeleton* of the MS complex is formed by the nodes and arcs, representing much of the connectivity information of the complex.

Topological simplification. A function f is simplified by repeated cancellation of pairs of critical points connected by an arc in the MS complex. The local change in the MS complex indicates a smoothing of the gradient vector field and hence of the function f. Gyulassy et al. [6] provided a full characterization of cancellation operations in terms of how they affect the connectivity of the complex and the geometry of the ascending/descending manifolds, operating solely on the combinatorial structure of the complex. Each cancellation operation removes a pair of critical points, reconnects arcs of the complex, and merges their ascending and descending manifolds with their neighbor's geometry. Repeated application of cancellations in order of *persistence* [3], the absolute difference in function value of the canceled critical points, results in a hierarchy of MS complexes and a multi-resolution representation of features. Gyulassy et al. [7] described data structures and search algorithms to reconstruct the ascending and descending manifolds of any critical point at any stage of simplification, allowing rapid browsing of the extracted features at multiple scales.

3 First-Principles Molecular Dynamics Simulation

Equilibrium atomic trajectories of one molecule of a Lithium salt $\mathrm{LiPF_6}$ dissolved in 63 molecules of ethylene carbonate (EC) at finite temperature were generated using first-principle molecular dynamics (FPMD). FPMD uses forces derived from quantum mechanics (density functional

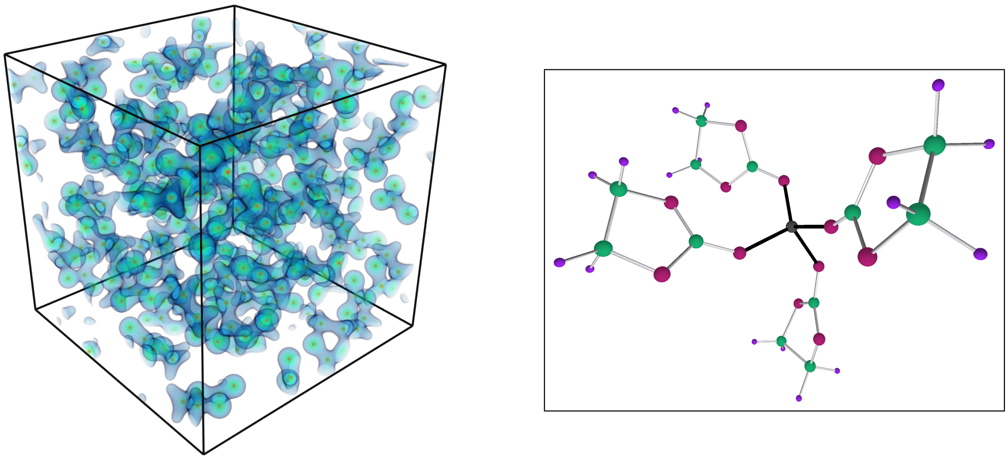


Figure 1: Visualization of the molecular system under investigation. (Left) The electron charge density field has maxima at atom locations and decays fast away from the atoms. (Right) The Lithium ion (dark gray) is shown at the center, and is bonded to four carbonyl Oxygen atoms (magenta), which are part of EC molecules on the other side. The Li–O bonds, which are the focus of this work, are shown in black, whereas the remaining are shown in white.

theory) to propagate the motions of atoms using classical Newtonian dynamics. A fundamental quantity of density functional theory is the electron density function, which is tracked through the simulation and used in the present analysis.

Simulations were performed using the Vienna *ab initio* simulation package (VASP) [11, 12] with the projector augmented wave (PAW) method [2, 13]. The simulation box was a cube 19.283 Ångströms in each dimension. A 450 eV plane-wave cutoff was used with Brillouin zone sampling restricted to the Γ point. A Nosé-Hoover thermostat [16, 9] with Nosé frequency of $\sim$1000 cm^{-1} was used to equilibrate the system at 330 K within the NVT ensemble. A temperature of 330 K was used to ensure the EC remained liquid ($T_{\text{melt}}^{\text{EC}} = 310$ K). The equations of motion were integrated using a time step of 0.5 fs. After 7.5 ps of equilibration, trajectories of 30 ps were run to gather statistics. Total electron densities for use in the MS complex analysis were computed by summing the valence electron density with the PAW-derived frozen core electron density.

The simulated molecular system (refer to Fig. 1) contains 638 atoms (2190 valence electrons) in total, containing 1 Lithium, 1 Phosphorus and 6 Fluorine that constitute a LiPF_6 molecule and 189 Carbon, 189 Oxygen, and 252 Hydrogen atoms that form 63 $(\text{CH}_2\text{O}_2)_2\text{CO}$ (EC) molecules. Each EC molecule contains a 5-ring structure of 3 Carbon atoms and 2 Oxygen atoms, and a double bond between a Carbon and an additional Oxygen (see Fig. 1). The Oxygen atom forming a double bond with a Carbon is called a *carbonyl Oxygen*, whereas the Oxygen atoms present in the ring are called *ether Oxygens*. It must be noted that there is a preference for EC to solvate the lithium ion.

Total wall clock time was approximately 2 months using 384 cores on 2.6 GHz Intel Xeon CPUs. The size of this simulation is close to the practical limit for typical accurate FPMD, since the cost of the calculation scales as the cube of the number of valence electrons (system size). The simulation saves checkpointing information at every time-step, which contains information about the positions, momenta, and forces corresponding to every atom. This is sufficient to recreate the entire state of the simulation at each time step if necessary. However, doing so would be prohibitively expensive. For example, creating an ECD field as well as a number of related fields using a $280 \times 280 \times 280$ regular grid takes about 550,000 CPU hours to generate. In general, sampled fields are available only every 625th time steps. For temporal regions of interest, i.e., the intervals containing reconnection events discussed in this paper ($t = [11.875, 12.187]$ ps and $t = [15, 15.312]$ ps), we refine these to every 25th time-step which represents the current limit in available post-processing resources. Fortunately, as demonstrated below the processes evolve sufficiently slowly for this limited temporal resolution to produce reliable results.

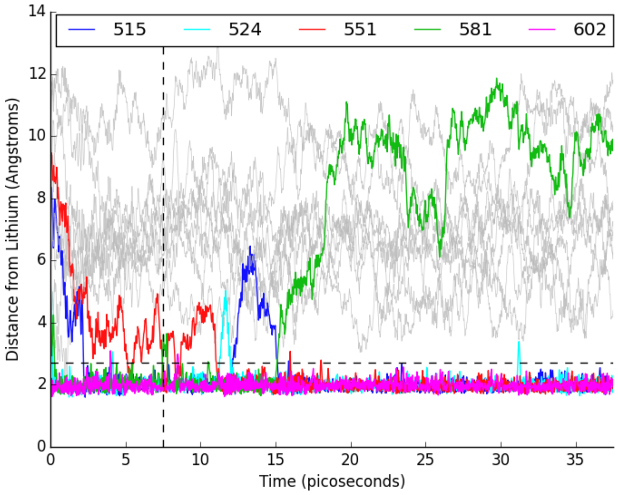


Figure 2: Distances of carbonyl oxygen atoms from Li for the entire simulation time-range. The oxygen atom that bond with Li for at least one time-step are colored by id; rest are shown in gray. The horizonal dashed line shows the typical distance thresholds used for detecting bonds, and the vertical dashed line marks when the simulation achieves equilibrium.

4 Analysis

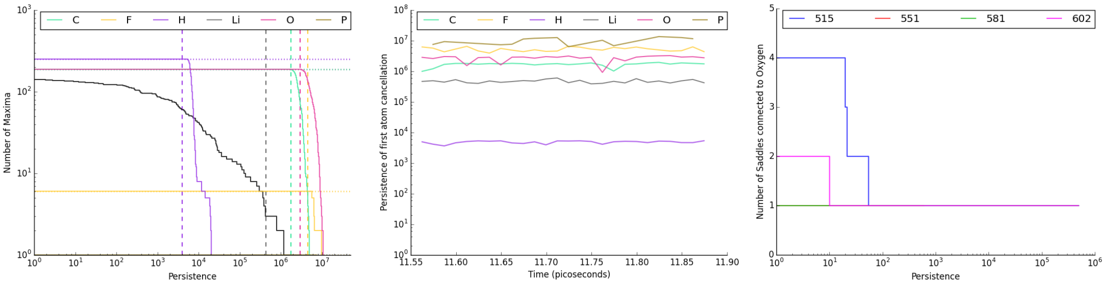
The main scientific goal of the FPMD simulations is to understand the dynamics of the Li-ion solvation by the EC molecules and the relation to ionic transport in the electrolyte. Previous work [17] has demonstrated a direct relationship between the solvation structure and diffusivity of Li ions in this and related systems. However, FPMD simulations are limited in total simulated time, therefore analysis of rare bond switching events is useful to understand longer time behavior without requiring explicit calculations for orders of magnitude longer simulation times.

The primary goal of the analysis presented in this paper is to characterize the Li–O bonds and reconnections events, and to potentially predict the latter. This section discusses the new topology based properties we use to analyze bonds and the solvation structure in general and the corresponding visualizations we use to illustrate our results.

4.1 Parameter-free Bond Detection

To date, the most common and most widely accepted definition of bonds is based upon geometric distances. Such an approach is attractive due to its connection with bond lengths for ideal covalent bonds, which are known for different types of atoms. However, for large and complex molecular systems containing different types of atoms forming different types of bonds, this approach may not be ideal. In particular, it requires a priori knowledge of the type of atoms involved and what type of atom interactions are likely to occur. Additionally, as temperature and pressure are changed, atomic distance are changed correspondingly, making distance based bond- definitions particularly problematic. In practice, it is difficult to pick an appropriate distance, and often obvious misclassifications occur. Fig. 2 shows the distances of all carbonyl oxygen atoms from Li for the entire simulation time-range, and the horizontal dashed line shows the distance threshold typically used for bond detection in such systems. The colored lines indicate all oxygen atoms that at some point are bonded to the Li ion, with all other oxygen atoms shown in gray. After equilibration, i.e., removing effects of initial conditions, (vertical dashed line), the reconnection events are easy to spot as one colored line switches with another. However, there are also multiple instances in which atoms that should be considered bonded are above the distance threshold, e.g., the cyan line at around 31 ps. According to the definition this would indicate a missing bond but typically this is considered an artifact rather than a reconnection event. Furthermore, raising the threshold to avoid this problem would likely result in spurious bonds in other time-steps, thus causing similar issues.

Given the central importance of the bond structures for most subsequent analysis steps, it is not surprising that identifying bonds in a robust manner is among the top priorities of domain scientists. Instead of the distance criterion, which is spatial in nature, our analysis is based



(a) (b) (c)

Figure 3: (a) Only spurious maxima (black line), which contain boundary artifacts, noise, and gridding artifacts, are cancelled during the simplification. For a wide range of persistence values, the maxima corresponding to atoms persist. Vertical lines represent the persistence values where atoms start getting cancelled, and horizontal lines represent the number of atoms in the system. (b) The minimum persistence at which an atom of a certain type gets cancelled stays stable over time. (c) While there may exist multiple high-saddles between two atoms representing the same bond, there must exist at least one for the bond to exist. This definition of bond is stable over a wide range of persistence simplification.

upon the ECD field, which is a functional metric, and provides a robust and parameter-free way of detecting bonds. The intuition is to use the structure of the ECD field independent of the distances involved, since it is ultimately the electronic charge distribution that determines whether a bond exists or not. We start with the concept of a "region of influence" for each atom. In the ECD field, each atom creates a local maximum and we use its descending 3-manifold – the mountain associated with this local peak – to define its region of influence. Effectively, this defines the region where an electron is more attracted to the given atom than to any others. Neighboring regions share surfaces and we say a bond between a Li and a carbonyl oxygen atom exists if and only if they share a boundary surface, see Fig. 4. With respect to a test particle, this means that there exists at least one point between two atoms where a test particle could reach either atom following the gradients of the ECD field. Note that this definition is tailored to the Li bonds in the sense that other atoms, i.e., oxygen and carbon atoms, may share boundaries with neighbors that do not indicate bonds. However, to understand Li diffusion which is the focus of this simulation the other bonds are not of interest.

Mathematically, this bond definition translates directly into concepts from Morse theory. As mentioned above, regions of influence are equivalent to descending 3-manifolds. For a smooth 3-dimensional function, two descending 3-manifolds share a boundary if and only if there exists a shared 2-saddle. For a discrete function, there can exist rare cases in which 3-manifolds share a boundary geometrically without a shared saddle. However, these must be due to discretization artifacts since the underlying true ECD field is certainly smooth. Consequently, we use only the topological structure, i.e., the existence of a 2-saddle, to indicate a bond. Nevertheless, other gridding artifacts can cause problems and must be compensated for. For example, some atoms create two or more local maxima in the sampled ECD field which is clearly unphysical.

Fortunately, standard persistence simplification can address this problem in an L_∞-optimal way and with a stable simplification threshold. In this case, the most obvious artifacts are the multiple maxima per atom, which can be detected easily as the true number of atoms and their location is known exactly. Fig. 3a shows the number of maxima versus persistence with maxima split by their corresponding atom types. The black line indicates the number of local maxima not related to atoms at all. These are local bumps in the ECD field with very low charge, far outside any region of interest and can safely be ignored. Even on the log-log scale of Fig. 3a, the persistence at which the duplicate maxima are removed is too small to see. At the other end of the plot, the colored vertical lines indicate the maximal persistence that preserves the correct maxima. This indicates that at least all atoms are resolved correctly for persistence values ranging from $0+\varepsilon$ to around 5000. In practice, the hydrogen atoms actually play little role in the analysis so this range could be extended to 400,000. Fig. 3b shows the maximum persistence at which each type of atom is correctly represented over time to show that these ranges remain highly stable.

Finally, Fig. 3c shows the number of 2-saddles connecting Li with its four surrounding oxygen

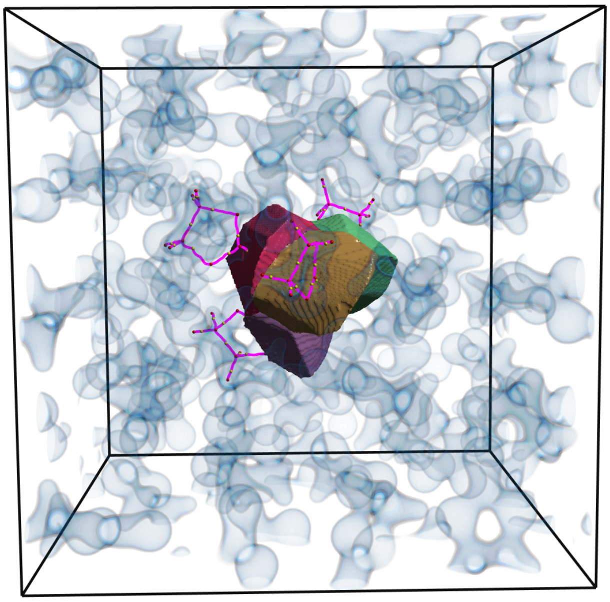
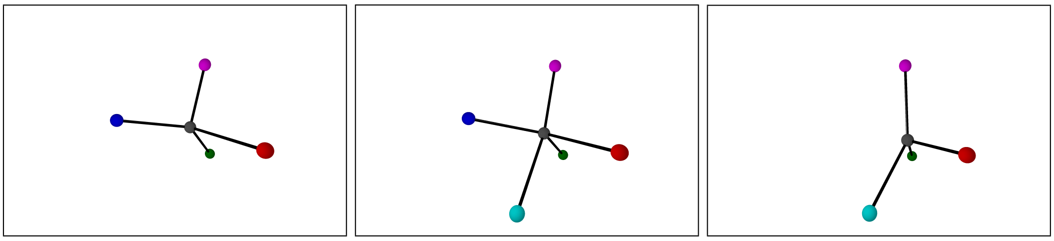


Figure 4: The regions of influence of Li, its bonded oxygen atoms, and their corresponding EC molecules.

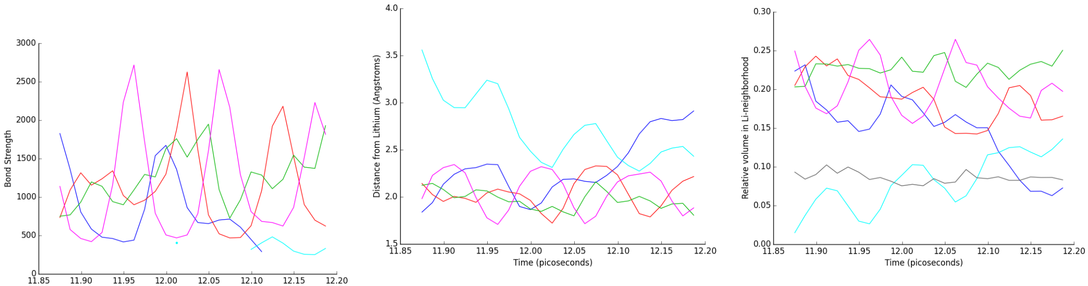


(a) $t = 12.087$.

(b) $t = 12.0995$.

(c) $t = 12.1245$.

Figure 5: In a 4-5-4 bond switch, Li gains an extra bond and exists temporarily in an unstable configuration bonded to five oxygen atoms. As a result of this instability, Li tries to shed a bond, effectively by moving towards one, red in this case.



(a) Strengths of bonds.

(b) Distance to bonded oxygen atoms.

(c) Volumes of the regions of influence.

Figure 6: Topological properties for the 4-5-4 bond switch configuration.

atoms. Recall that the definition uses the existence of at least one 2-saddle to indicate a bond, since the fact that at low persistences more than one may exist is inconsequential. Typically, these are due to well-known artifacts of discretizing a continuous function onto a grid, and computing the MS complex with a discrete approach [5]. Nevertheless, the important information of Fig. 3c is that bonds remain stable at very high persistence thresholds. In practice, we typically use small persistence values of around 10 as the multiple 2-saddles are easily handled. However, as demonstrated above, the analysis results would not change for persistence of 1 or 100. Therefore, even though persistence, in principle, represents a free parameter of our bond detection, its value at least for Li–O bonds can safely be set a priori within a range spanning multiple orders of magnitude. As a result, we consider the simplification to remove gridding artifacts a part of the algorithm with a fixed threshold that is never modified, and thus the subsequent bond detection is effectively parameter free.

4.2 Bond switching

Using the information derived from the topological definition of bonds, we study two different types of reconnection events present in the simulation, called *4-5-4* and *4-3-4* switches, respectively. In a stable configuration, Li is attached to four oxygen atoms [17], which form a tetrahedron-like structure enclosing Li. However, during a reconnection event, there exists a transient unstable configuration where Li is connected to either three or five oxygen atoms, hence the names.

Although distance-based measure can detect such events, it is not well suited to provide more insights into their mechanics. Here, we use topological constructs to understand the two switching configurations better and possibly predict when a switching will happen or not. In particular, the topological properties allow understanding the appropriate balance of conditions that must be met in order for a reconnection to take place.

4-5-4 Switch. An example of this type of reconnection is shown in Fig. 5. Here, Li initially exists in a stable state bonded with four oxygen atoms – colored red, green, blue, and magenta. At $t = 12.0995$ ps, a new oxygen atom (cyan) moves close to and attaches itself to Li before an existing bond breaks. Temporarily, Li is bonded to five oxygen (with respect to the topological bond definition), an unstable configuration that will quickly result in the Li shedding one bond. In this example, the blue bond breaks at $t = 12.112$ ps and the corresponding oxygen moves away.

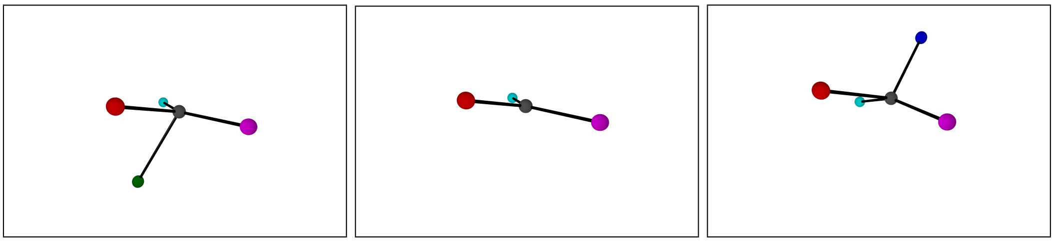
Fig. 6a shows the strengths of Li–O bonds, i.e., the ECD function value of the 2-saddle connecting Li and O, in the chosen time window, and confirms the blue bond weakening substantially before breaking. One interesting feature of bond strengths, which has not been observed previously, is the significant and regular oscillation in their behavior. All bonds appear to have comparable periods but are phase-shifted and have different amplitudes.

This is especially interesting at $t = 12.075$ ps when the cyan oxygen, in fact, forms a temporary bond, which immediately breaks. From the spatial arrangement of the corresponding atoms (see Fig. 5), we find that the cyan oxygen atom approaches Li from between the red and the blue oxygen atoms towards the red-blue edge of the tetrahedron formed by the four bonded oxygen atoms. As a result, it can switch only with either the red or the blue oxygen. However, at $t = 12.075$ ps both of these bonds are strong, and even though the cyan oxygen is close enough to force an intermediate bonding, the existing solvation structure is very stable, causing Li to shed this excess bond immediately.

Considering the actual reconnection event, the atom position suggests that the cyan oxygen approaches Li approximately half way between the red and the blue oxygen (along the red-blue edge of the tetrahedron). This observation matches the distances (see Fig. 6b) around $t = 12.1$ ps, as well as the knowledge that all oxygen atoms will repel each other, and thus will prefer an equal spatial distribution. However, despite this apparent balance, the period of the oscillation of the red bond has its strength increasing at this point in time and thus the blue bond breaks instead.

The final plot of Fig. 6c shows the relative volume of the regions of interest in a ball of radius 2.2 Ångström around Li. While the general behavior matches that of the bond strength, the volume appears to have a longer lead time, which may be used to predict potentially unstable configurations well in advance. In particular, notice how the volume of the red oxygen declines almost monotonically in preparation for the reconnection event, even though its bond strength continues to oscillate normally. Similarly, the blue oxygen atom's region shrinks significantly early in the process, though it recovers intermittently. Furthermore, it is interesting to note that the bond strength of all oxygen atoms combined are relatively low when the reconnection happens (potentially because the charge is now distributed among five bonds), yet the volumes of the uninvolved magenta and green oxygen remain comparatively high. While this analysis is still in the early stages of development, the relationships between the bond strength, their phase of oscillation, and the volume shows significant promise to better understand and predict reconnection events.

4-3-4 Switch. In the second type of reconnection, one of the existing bonds breaks early

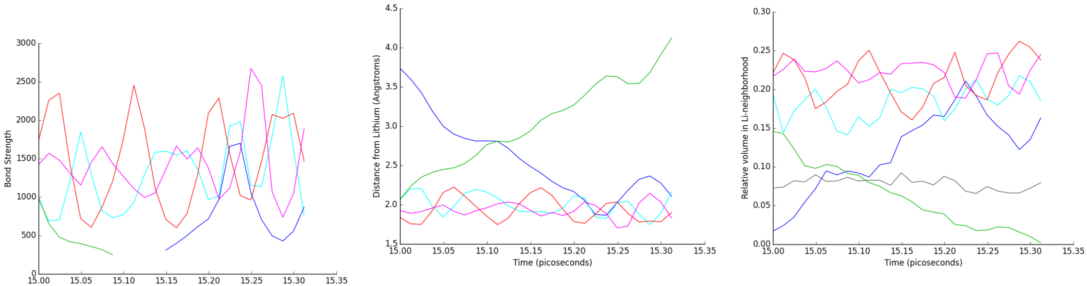


(a) $t = 15.087$.

(b) $t = 15.112$.

(c) $t = 15.1495$.

Figure 7: In a 4-3-4 bond switch, Li loses a bond and exists temporarily in an unstable configuration bonded to only three oxygen atoms. As a result of this instability, Li tries to gain a bond by pulling another oxygen inward, blue in this case.



(a) Strengths of bonds.

(b) Distance to bonded oxygen atoms.

(c) Volumes of the regions of influence.

Figure 8: Topological properties for the 4-3-4 bond switch configuration.

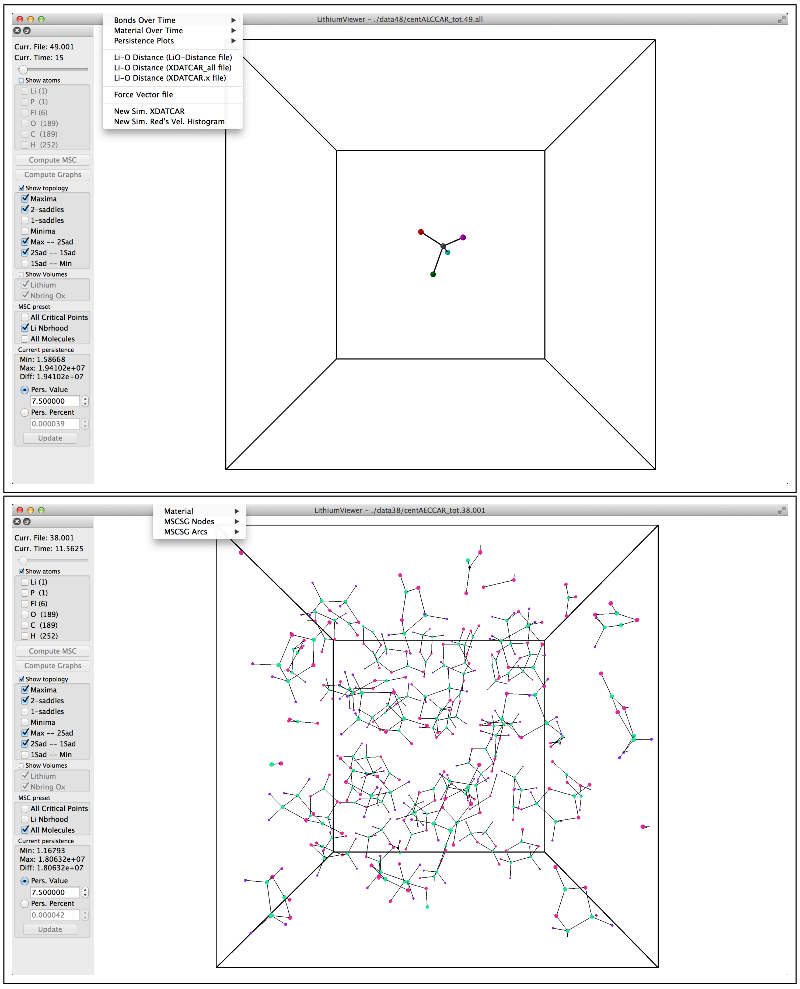
and Li exists in a deficient configuration bonded to only three oxygen atoms. Fig. 8a shows the bond strengths during such a reconnection event. As seen from the figure, the green bond weakens substantially and breaks at $t = 15.087$ ps. 0.0625 ps later, the blue oxygen comes close enough to bond with Li. Notice in Fig. 7 that in this case, the blue and the green oxygen atoms lie on opposite sides of the plane formed by the other three oxygen atoms. We believe that this reconnection is made possible only because the green bond breaks first, and is probably pulled away from Li by its attached EC molecule. To compensate, the Li forms a bond on the other side of the plane and ultimately moves across flipping its solvation tetrahedron. Similar to the 4-5-4 event the volumes of the corresponding regions of influence appear to predict the switch early. As shown in Fig. 8c the volume of the green oxygen declines significantly before breaking. Furthermore, we observe other apparent mismatches in strength and volume. Prior to the bond breaking the red oxygen shows a very low strength connection yet its volume remains comparatively high.

4.3 Analysis and Visualization Tool

We developed a custom tool for the analysis and visualization of the data generated from FPMD simulations (see a screenshot in Fig. 9). The tool is able to load a variety of variables including types and positions of atoms, various scalar fields, forces on atoms, types of oxygen atoms (carbonyl vs. ether), etc. The electron charge density field is passed through the MS complex computation pipeline, which first computes its discrete gradient, and then computes the complex before performing persistence simplification. An optimal persistence level is then chosen (based on the discussion in Section 4.1) for bond detection, which is performed by traversing through the MS complex.

In order to understand the dynamic nature of the evolution of bonds, the tool can handle multiple time-steps of the same data simultaneously, allowing the generation of real-time animation. Processing 50 time-steps of data requires about 7 GB of memory including the scalar field, gradient, complete MS complex hierarchy, and various custom structures extracted for analysis, e.g., the bonds.

Finally, to analyze the data seamlessly, Qwt plotting widget was used within our Qt-based tool to allow interactive plotting of various variables over time, e.g., bond strengths and bond



LithiumViewer – ./data48/centAECCAR_tot.49.all

Curr. File: 49.001
Curr. Time: 15

☐ Show atoms
☐ Li (1)
☐ P (1)
☐ Fl (6)
☐ O (189)
☐ C (189)
☐ H (252)

☑ Show topology
☑ Maxima
☑ 2-saddles
☐ 1-saddles
☐ Minima
☑ Max -- 2Sad
☑ 2Sad -- 1Sad
☐ 1Sad -- Min

☐ Show Volumes
☑ Lithium
☑ Nbring Ox

MSC preset
☐ All Critical Points
☑ Li Nbrhood
☐ All Molecules

Current persistence
Min: 1.58668
Max: 1.94102e+07
Diff: 1.94102e+07
● Pers. Value 7.500000
○ Pers. Percent 0.000039

LithiumViewer – ./data38/centAECCAR_tot.38.001

Curr. File: 38.001
Curr. Time: 11.5625

☑ Show atoms
☐ Li (1)
☐ P (1)
☐ Fl (6)
☐ O (189)
☐ C (189)
☐ H (252)

☑ Show topology
☑ Maxima
☑ 2-saddles
☐ 1-saddles
☐ Minima
☑ Max -- 2Sad
☑ 2Sad -- 1Sad
☐ 1Sad -- Min

☐ Show Volumes
☑ Lithium
☑ Nbring Ox

MSC preset
☐ All Critical Points
☐ Li Nbrhood
☑ All Molecules

Current persistence
Min: 1.16793
Max: 1.80632e+07
Diff: 1.80632e+07
● Pers. Value 7.500000
○ Pers. Percent 0.000042

Figure 9: Screenshots of our custom Qt-based tool for the analysis and visualization of molecular data. Left: Visualization of Li's bonds; the drop-down menu on the top-left shows the types of plots that can be generated interactively. Right: Visualization of all EC molecules; the drop-down menu on the top-left shows different rendering modes. The panel on the left of the tool displays various visualization options available, including the slider to animate in time, and the spin-boxes to choose from the range of available persistences.

volumes, as well the number of critical points over persistence cancellations. Using the GLE tubing and extrusion library, our tool can create 3D visualizations of atoms and bonds, as well as directly visualize the components of the MS complex, in addition to the volume rendering of the entire field and surface rendering of the boundaries of the regions of influence.

Overall, the user-interface and interactivity of this tool have proved very useful during the exploratory analysis of molecular data; with an easy learning curve, our non-visualization-expert collaborators have also found the tool intuitive.

4.4 Expert Evaluation

In the opinion of the domain experts that are a part of our team, the stability of the topological definition of bonds, and its systematic and parameter-independent nature is very useful. In particular, the new definition agrees with the distance-based criterion for most of the simulation. However, in cases of disagreement, e.g., the lost bonds discussed in Fig. 2, the new definition produces results better aligned with the physical intuition. Consequently, during reconnection events, where the bonds are unstable, topologically-defined bonds will likely be more reliable and thus better able to explain the process.

Additionally, the topological definition provides related information on the bonds and the solvation structure that has already proven valuable and cannot be derived from the distances alone. For example, the relative position of the saddle, i.e., the ratio of distances between its two atoms, approximately equals the ratio of the corresponding electron charge densities, which further validates the new bond definition. Furthermore, the charge density value at saddles indicates bond strength, which appears to play an important role during reconnection events, as will be discussed in the next section. Finally, the volumes of the regions of influence may provide another property of interest to understand, e.g., how tightly packed the current solvation structure is.

The new bond definition and the subsequent analysis discussed in this paper have raised significant hope in being able to achieve the ultimate goal of predicting unstable "branch-points" of a simulation at which the outcome could be altered with minor changes. In particular, the analysis of the 4-5-4 reconnection events has convinced the domain scientists to invest significant computational resources to recompute a part of the initial simulations with alterations suggested by the topological analysis. This preliminary analysis is still exploratory, yet interesting enough to open new directions for investigation. For example, a detailed topological analysis of the modifications might explain some of these observed behaviors in terms of bond strengths. Unfortunately, due to the time and costs involved in performing the simulations, the modifications in simulations have a long turn-around time, and a more detailed analysis of the phenomena will be the focus of future work.

Nevertheless, the experts are convinced that this will greatly facilitate both the consistency and ease of carrying out such analysis, especially in the case of multiple species and extreme conditions of pressure and temperature, where interatomic distances can be quite far from distances at ambient-conditions.

5 Impact and Discussion

Understanding the dynamics of first-principles molecular simulations is crucial in many areas of material science. Bonding analysis is used throughout the computational chemistry and physics communities to understand the properties of materials. Nevertheless, most of the commonly used ideas for the detection of bonds are not parameter free, or are known primarily in theory only.

In this paper, we introduce and apply a new topology-based analysis of the bonding behavior of Li ions in ethylene carbonate. In particular, we employ a functional definition of the Li–O bonds, and show that it is completely systematic and parameter free. Furthermore, this topological definition of bonds leads to a number of additional quantities of interest such as bond strengths, relative bond positions, and relative volumes of the regions of influence of each atom, which have not been explored before. In the case of the molecular system under consideration, among the most important features of interest are the reconnection event, which shed light on the mechanics of Li ion solvation. The bonding structures and the additional indicators detected topologically, lead to a number of interesting insights, such as the strong oscillations in bonds, and their influence on the formation and breaking of bonds.

Furthermore, one of the main challenges in understanding molecular systems is the computational cost of simulating such behavior, which requires large-scale efforts, but often yields only a small set of interesting events. The analysis presented in this paper has enabled our team consisting of data analysis and material science experts to locate "branch points" in the simulation, where minor changes in the state of the atoms, e.g., positions and/or velocities, etc., can be used to study the stability of Li–O bonds, and potentially produce a structurally different reconnection event. Nevertheless, more research is needed to further explore the least amount of modifications necessary to effect such changes, when to apply them to maximize the differences, and how likely this would have occurred naturally assuming some amount of randomness in the process.

Going forward, we plan to extend this analysis to other simulations and other types of bonds, as well as to tackle other fundamental scientific goals. We believe that by analyzing the time-lags observed between the changes in bond strength, distances, and volumes it may be possible to gain more insights into these phenomena. Ultimately, the hope is that the new analysis will allow scientists to explore solvation and diffusion more systematically, with the aim of new insights and, with branch point prediction, potentially accelerating the calculations themselves.

References

[1] R. Bader. *Atoms in Molecules: A Quantum Theory.* Oxford University Press, 1994.

[2] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, 1994.

[3] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

[4] D. Günther, R. A. Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing molecular interactions in chemical systems. *IEEE Trans. on Vis. and Comp. Graph.*, 20(12):2476–2485, Dec. 2014.

[5] A. Gyulassy, P. Bremer, and V. Pascucci. Computing Morse-Smale complexes with accurate geometry. *IEEE Trans. on Vis. and Comp. Graph.*, 18(12):2014–2022, Dec. 2012.

[6] A. Gyulassy, P.-T. Bremer, V. Pascucci, and B. Hamann. A practical approach to Morse-Smale complex computation: Scalability and generality. *IEEE Trans. on Vis. and Comp. Graph.*, 14(6):1619–1626, 2008.

[7] A. Gyulassy, N. Kotava, M. Kim, C. D. Hansen, H. Hagen, and V. Pascucci. Direct feature visualization using Morse-Smale complexes. *IEEE Trans. on Vis. and Comp. Graph.*, 18(9):1549–1562, 2012.

[8] A. Gyulassy, V. Natarajan, V. Pascucci, P.-T. Bremer, and B. Hamann. Topology-based simplification for feature extraction from 3D scalar fields. *IEEE Trans. on Vis. and Comp. Graph.*, 12(4):474–484, 2006.

[9] G. H. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[10] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang. Revealing noncovalent interactions. *J. of the Amer. Chem. Soc.*, 132(18):6498–6506, May 2010.

[11] G. Kresse and J. Furthmuller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mat. Sci.*, 6(1):15–50, July 1996.

[12] G. Kresse and J. Furthmuller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct. 1996.

[13] G. Kresse and D. Joubert. From ultrasoft pseudo potentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758–1775, 1999.

[14] Y. Matsumoto. *An Introduction to Morse Theory.* American Mathematical Society, 2002. Translated from Japanese by Kiki Hudson and Masahico Saito.

[15] C. Matta and R. Boyd. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design.* Wiley, 2007.

[16] S. A. Nose. Unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81(1):511, 1984.

[17] M. T. Ong, O. Verners, E. W. Draeger, A. C. T. van Duin, V. Lordi, and J. E. Pask. Lithium ion solvation and diffusion in bulk organic electrolytes from first-principles and classical reactive molecular dynamics. *J. of Phys. Chem. B*, 119(4):1535–1545, Jan. 2015.

[18] A. Savin, R. Nesper, S. Wengert, and T. F. Fässler. ELF: The electron localization function. *Angewandte Chemie International Edition in English*, 36(17):1808–1832, 1997.

[19] A. Savin, B. Silvi, and F. Colonna. Topological analysis of the electron localization function applied to delocalized bonds. *Can. J. Chem.*, 74:1088–1096, 1996.

[20] K. Xu. Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chem. Rev.*, 104(10):4303–4418, Oct. 2004.