Interactive Exploration of Atomic Trajectories Through Relative-Angle Distribution and Associated Uncertainties

Harsh Bhatia*
Lawrence Livermore National Laboratory

Attila G. Gyulassy & Valerio Pascucci
The University of Utah

Martina Bremer
San Jose State University

Mitchell T. Ong, Vincenzo Lordi, Erik W. Draeger, John E. Pask, & Peer-Timo Bremer
Lawrence Livermore National Laboratory

ABSTRACT

Exploration of atomic trajectories is fundamental to understanding and characterizing complex chemical systems important in many applications. For instance, any new insight into the mechanisms of ionic migration in catalytic materials could lead to a substantial increase in battery performance. A new statistical measure, called the *relative-angle distribution*, has been proposed to understand complex motion – whether Brownian, ballistic, or diffusive. The relative-angle distribution can be represented as a collection of 1D histograms, but is currently created in a slow, offline process, making any parameter exploration a tedious and time-consuming task. Furthermore, the resulting plot can hide uncertainty in both the data and the visualization. As a result, once rastered or printed at a fixed resolution, these histograms can be misleading.

We present a new analysis tool for the exploration of atomic trajectories that combines an interactive histogram visualization with uncertainty information for both data and plotting errors, and is also linked to an interactive 3D display of trajectories. Our tool enables a holistic exploration of trajectories previously not feasible, with the potential for significant scientific impact. In collaboration with domain experts, we have deployed our tool to analyze molecular dynamics simulations of lithium-ion diffusion. Users have found that the tool significantly accelerates the exploration process and have used it to validate a number of previously unconfirmed hypotheses.

1 INTRODUCTION

Understanding the stochastic nature of random walks in complex dynamical systems offers insights into the underlying phenomena, such as in molecular dynamics [18, 22], turbulence modeling [3], cell biology [10, 14, 26], and ecology and social biology [21, 24, 25]. One particularly interesting application is the design of batteries, where atomic trajectories, modeled as random walks, may provide insight into the diffusive behavior of lithium ions. Lithium-ion batteries are currently one of the most common portable energy sources, powering everything from hand-held consumer devices to electric vehicles. Typically, such batteries consist of an anode and a cathode separated by an electrolyte, and lithium ions move between the two ends to generate external voltage. The speed at which an ion moves is often quantified by its diffusion coefficient, which determines the cycling rate performance of a battery. As a result, the exploration of atomic trajectories to characterize their diffusive behavior is key to understanding the underlying complex dynamics and ultimately to improve battery performance [18, 27].

The structure of complex atomic trajectories is typically explored manually through 3D spatial visualization tools such as VMD [13]. In addition, various statistical indicators, such as mean-square displacement, are computed to derive quantitative properties, such as the diffusion coefficient [2, 9]. To understand the properties of complex trajectories, a new statistical measure, called the *relative-angle distribution* [5], was proposed recently. The idea is to measure the directional motion of a particle by computing the distribution of relative angles of its positions along its trajectory (see Section 3). The shape of the distribution can highlight different effects, such as the restricted motion due to neighboring molecules vs. consistent forward motion, Brownian vs. non-Brownian motion, and the like. This measure is especially important since the diffusion coefficient alone cannot characterize non-Brownian motion.

The relative-angle distribution for a given time-scale, visualized as a histogram, has proven useful in a variety of applications [3, 5, 14, 23, 26]. However, the application of our interest requires studying these distributions for a wide range of time-scales simultaneously. These distributions can be visualized as a *2D image* (Figure 1a), where the horizontal axis represents time-scales, and each column is a 1D histogram of relative angles for the corresponding time-scale. Although useful in theory, such visualizations pose important challenges, especially for large-scale data.

In particular, the approximation of a relative-angle distribution through histograms with a finite number of bins and a finite sampling of points along a trajectory introduces errors, which are typically disregarded. Even though the combination of bin-size and sample-size can significantly impact the shape of the resulting histogram, current tools, unfortunately, are too cumbersome to allow an exploration of various settings and their consequences. For example, the existing Mathematica-based tool chain required about 5-6 minutes of computation time, and about an hour of human time to produce the result in Figure 1, which shows 160 histograms with 36 bins for a time-scale resolution of 0.025 picoseconds (ps). Since repeating this process sufficiently often to explore bin sizes and sampling uncertainties is impractical, scientists currently use a best-guess single solution for the analysis.

There is a clear need and significant interest to interactively explore the solution space, e.g., to highlight certain time-scales, changing bin sizes, zooming and panning in the 2D image, etc., as well as to understand the uncertainties in data. Furthermore, once an interesting feature in the relative-angle visualization has

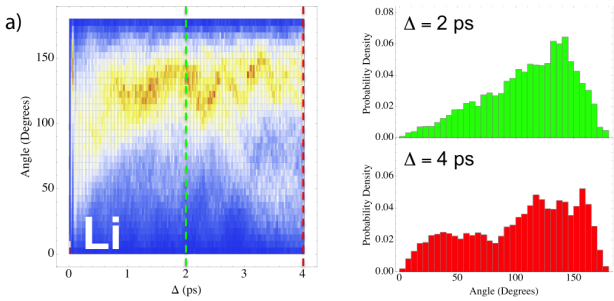


Figure 1: Relative-angle distribution for a lithium ion for different time-scales shown as a 2D image and two histograms corresponding to the highlighted columns.

*e-mail: bhatia4@llnl.gov

IEEE Pacific Visualization Symposium 2016
19–22 April, Taipei, Taiwan
978-1-5090-1451-4/16/$31.00 ©2016 IEEE

been found, the natural next step is to explore the corresponding 3D visualization of the associated atomic trajectories. However, currently there exists no link between the statistical analysis and the 3D visualization, resulting in a tedious and error-prone workflow of manually synchronizing, e.g., the time-scales of interest.

Finally, the visualization itself can cause additional errors, which are rarely evaluated or even acknowledged. Creating a 2D visualization of multiple histograms in a tool such as Mathematica will render it at a fixed resolution, and without special precaution, this resolution is often chosen to be one pixel per data item, i.e., histogram bin. However, if histograms computed at high resolution are downscaled to match a lower image resolution, over-plotting and color blending can occur, which might cause spurious and misleading effects. On the other hand, if the data is low resolution, the final image will be upscaled by the rendering engine, causing the distributions to appear artificially smooth due to blurring, and possibly suffering from false color interpolation. Both situations cause visualization errors that can even outweigh the data uncertainties.

Contributions. We present an interactive tool addressing these issues by creating and rendering 2D visualizations of a relative-angle distribution at desired resolution and granularity, while presenting both the corresponding data and visualization uncertainties. Furthermore, the statistical analysis is tightly integrated with 3D visualization of associated trajectories. In particular, we demonstrate:

- Interactive visualization of a relative-angle distribution for arbitrary time-scale ranges and bin sizes;
- Integrated analysis and illustration of uncertainties in both the data and visualization of the resulting plots;
- A linked-view framework to explore, analyze, and compare trajectories using relative-angle distributions and 3D visualization; and
- Application of our techniques to a first-principles simulation of lithium ion diffusion, which has allowed our collaborators to validate previously unconfirmed hypotheses, as well as observe a number of unexpected phenomena.

2 Related Work

Visualization of relative-angle distributions, proposed in 2013 [5], has not gone far beyond the obvious solution of displaying 1D histograms for a given time-scale using line graphs [3, 5], bar graphs [23], or star plots [14, 26]. So far, only Savage et al. [22] have used 2D visualizations for multiple 1D histograms for different time-scales, which, however, are non-interactive, omit uncertainties, and can introduce visualization errors.

One way to interactively visualize the temporal evolution of histograms is TimeHistograms [15]. However, TimeHistograms cannot be used for our application; since the number of data samples in histograms decreases substantially with increasing time-scale (cf. Section 4), we must treat the data as a collection of 1D distributions (a 2D image) and not as a 2D histogram.

2D image visualization. Specifically, we consider the problem of displaying $N \times B$ data (distributions for N time-scales with B bins each) onto an output resolution of $n \times b$. Depending upon these values, one can run into two important visualization challenges: under-plotting and over-plotting.

Under-plotting occurs when the output resolution is much higher than the input resolution. A small amount of data is mapped to a large image, often performed implicitly by the renderer, which interpolates between adjacent pixels. Depending on the specific rendering, the interpolation might blur the image, implying a nonexistent continuity, and may incorrectly interpolate colors.

Over-plotting, on the other hand, occurs when a large amount of data must be visualized on a relatively smaller output space, and multiple data points end up sharing the same space, such that not all data points are discernible to the viewer. A number of ideas have been proposed to remove or reduce over-plotting [8]. These approaches can be broadly categorized into three classes: (a) changing the appearance of data points; (b) jittering the data points; and (c) reducing the amount of data, e.g., by subsampling, filtering, or aggregating the data, to visualize only representative data points.

For example, embedded plots [11] organize a collection of graphs into a larger graphic, and the Abstract Rendering framework [6] provides an explicit control over the rendering pipeline. Considering the rendering process as binning allows unifying the different strategies discussed above. Continuous scatterplots [1] and continuous parallel coordinates [12] have been developed that use combinations of the above strategies to display discrete data as continuous, thus, indirectly addressing over-plotting problems.

It must be noted that, although similar in the sense of the umbrella definition, our problem of reducing $N \times B$ data to $n \times b$ pixels is not the same as the over-plotting generally considered by the visualization community. As a result, existing techniques do not apply here. In general, strategies (a) and (b) are not helpful, since changing the appearance (colormap) does not reduce over-plotting, and pixel positions cannot be changed. Furthermore, representing discrete 2D histograms as continuous information is not meaningful in this case, as discontinuities across different time-scales represent important information, and removing them can be misleading. As a result, we are left with strategy (c) and have elected to aggregate data in a controlled and meaningful way.

Ensemble visualization [19, 20] is a common technique that employs aggregation of related datasets, and the resulting aggregated data is visualized along with associated uncertainties. For example, surface temperature may be available using a number of climate models or different parameters for the same model, and one may wish to display the mean and the standard deviation only.

The goal in ensemble visualization is to show a representative of the underlying models defining a function on the domain. The representative is typically derived from hundreds or thousands of samples. However, our goal is to compress a few (of the order of tens) 1D histograms into a single one. Since these histograms for different time-scales may or may not show similar behavior, computing statistical measures, such as standard deviation, is not meaningful in general.

3 Application Data and Simulation Details

This paper focuses on applying visualization principles to better understand data generated by large-scale simulations of lithium-ion battery systems. In particular, these are first-principles molecular dynamics (FPMD) simulations that use density functional theory (DFT) with the projector augmented wave (PAW) method as implemented in the Vienna *ab initio* simulation package (VASP) [16, 17].

The simulation was performed in a cube 19.283 Angstroms in each dimension containing one molecule of a lithium salt, $LiPF_6$, dissolved in 63 molecules of ethylene carbonate (EC), i.e., a total of 638 atoms: 1 lithium, 1 phosphorus, 6 fluorine, 189 carbon, 189 oxygen, and 252 hydrogen. The equations of motion were integrated using a time step of 0.0005 picoseconds (ps). After 7.5 ps of equilibration, atomic trajectories of 30 ps were run to gather statistics. Further computational details on the simulation setup and data generation were provided by Ong et al. [18].

Given a particle trajectory, $\mathbf{X}(t)$, and a *time-scale* of interest, Δ, the *relative angle*, $\theta(t)$, with respect to Δ is defined as

$$\theta(t;\,\Delta) = \cos^{-1}\left(\frac{\mathbf{V}(t;\,\Delta)\cdot\mathbf{V}(t+\Delta;\,\Delta)}{|\mathbf{V}(t;\,\Delta)|\,|\mathbf{V}(t+\Delta;\,\Delta)|}\right), \tag{1}$$

where, $\mathbf{V}(t;\,\Delta)$ is the displacement vector for the particle between time t and $t+\Delta$, i.e., $\mathbf{V}(t;\,\Delta) = \mathbf{X}(t+\Delta) - \mathbf{X}(t)$ (see Figure 2).

Essentially, θ measures the change in the particle's directional motion after Δ time units. The distribution of θ for a given Δ

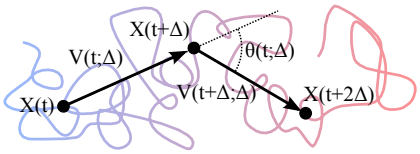


Figure 2: As an atom moves along a trajectory, with the position given by $\mathbf{X}(t)$, a single angle $\theta(t;\,\Delta)$ records the deviation from straight line motion for the positions $\mathbf{X}(t)$, $\mathbf{X}(t+\Delta)$, and $\mathbf{X}(t+2\Delta)$.

is approximated using a histogram of sampled values along $\mathbf{X}(t)$, which helps understand the characteristics of the particle's motion. In particular, an evenly-distributed, "flat" histogram indicates a Brownian-like motion where the relative angles are uncorrelated in time; the particle has an equal probability of moving in all possible directions, representing ideal diffusive behavior. On the other hand, a histogram centered around some angle suggests non-Brownian motion, typically not easily characterized by previously used indicators, such as the mean-square displacement. For example, for bonded atoms, strong peaks around higher angles may indicate "rebounding" as the bond is stretched and compressed.

4 Visualization of Relative-Angle Distribution

Given a discrete and sampled trajectory $\mathbf{X}[t]$ for $0 \leq t < T$, we compute relative angles for each consecutive atom position $\mathbf{X}[t_0]$ using Eq. (1). Since each computation requires two future positions, $\mathbf{X}[t_0+\Delta]$ and $\mathbf{X}[t_0+2\Delta]$, we can use only $T-2\Delta$ samples for computing relative angles. This raw data is then used for the computation of histograms for the required number of bins.

Typically, the relative-angle distribution is visualized as a histogram of angles with respect to a given time-scale Δ, allowing characterization of atomic motion; e.g., Figure 3 indicates non-Brownian motion. Nevertheless, since a histogram aggregates data, all temporal and causality information is lost. We augment the visualization by also showing the distribution of angles vs. time (Figure 4), which helps in two ways. First, this additional plot highlights patterns in the time history of the atom, identifying possibly recurring behavior, e.g., the figure indicates some periodicity where the particle exhibits sharp turns with periods of smaller deviations, and then back again. Second, the span of the time axis gives an indication of the number of samples available for this time-scale to indirectly assess the quality of the corresponding histogram. Finally, we show a histogram of time-scales for a given angle, which indicates whether or not an angle under consideration is significantly more frequent for certain time-scales (Figure 5). Such 1D statistics can be direct indicators for detecting prominent time-scales for different characteristic motions.

Multiple distributions as a 2D image. Next, we compute relative-angle distributions for time-scales ranging from 0.005 ps to 14.995 ps with a sampling of 0.005 ps – 5x finer resolution than Figure 1. For the purpose of exploration, the largest possible range of time-scale was chosen for the experiments. These histograms are visualized as a 2D image, as shown in Figure 6, where the time-scale increases left to right, and the angle increases from bottom to top. For the best perceptual understanding of the relative-angle distributions, a color map that varies a single hue was used. Each column represents a histogram, which must be normalized individually to allow consistent comparison of distributions across time-scales. This normalization is crucial to the analysis, because the number of data samples decrease with increasing Δ. In our experiments, the left-most histogram ($\Delta = 0.005$ ps) contains 59,980 data points, whereas the right-most histogram ($\Delta = 14.995$ ps) has only 20 samples. This large disparity between the number of samples exposes a challenge in applying a color map. Individual samples become relatively more important for larger time-scales, resulting in higher peaks in the distribution. As a result, a global color map applied to individually normalized histograms causes the histograms for

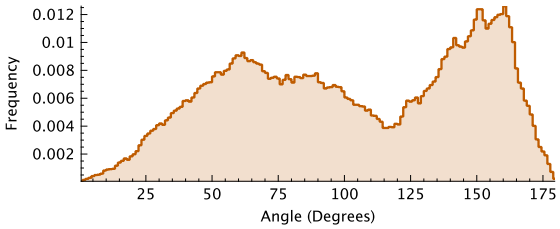


Figure 3: Histogram of angles for $\Delta = 4.95$ ps.

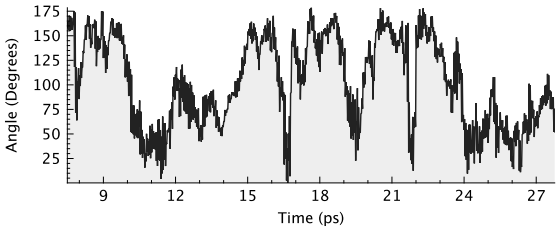


Figure 4: Distribution of angles with time for $\Delta = 4.95$ ps.

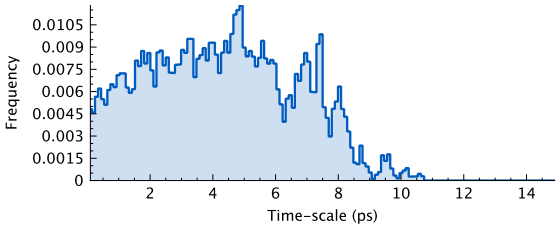


Figure 5: Histogram of time-scales for $\theta = 166°$.

smaller time-scales to be "washed out" and impairs the visualization (Figure 6a). Although the resulting image provides a direct comparison of histograms, which is the most significant aspect of the data, it loses detail for shorter time-scales, and the corresponding histograms appear similar. To remedy this limitation and to allow visualizing regions of interest in greater detail, the color map is applied dynamically only to the visible range of data values interactively as the user zooms or pans inside the image (Figure 6c).

5 Visualization of Uncertainty

Given relative-angle distributions for multiple time-scales, next we describe how to visualize the corresponding uncertainties. We consider two types of uncertainties: *data uncertainty* and *visualization uncertainty*. The uncertainty in data arises from a number of different sources, ranging from the simulation parameters to sampling methods. In particular, scientists are interested in understanding how well a histogram, computed on a given sample of points along a trajectory, estimates the true distribution. The second source of uncertainty is the visualization technique itself. The goal is to show the user exactly how much the data might have been distorted by the visualization in order to create a given plot.

5.1 Data Uncertainty

A histogram of relative angles is defined by a time-scale, Δ, and a number of bins, B. The computation of this histogram requires sampling along a given trajectory and measuring relative angles using Eq. (1). Given S such samples and corresponding angles, the goal is to estimate how well the histogram computed on the given set of samples represents the true distribution. Since the true distribution is unknown, we use statistical techniques to estimate the accuracy of a histogram. In particular, we use a random sampling technique called *jack-knifing*, a variant of *bootstrapping* [7], which allows capturing confidence intervals and error estimates by computing statistics on random subsets of data. Specifically, given a

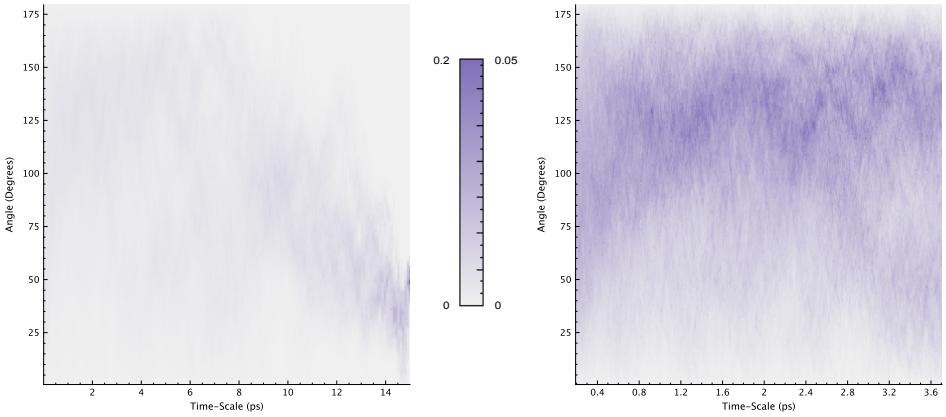


Figure 6: Relative-angle distribution visualized for multiple time-scales visualized as a single 2D image; each column is a 1D histogram. Vast differences in number of samples between histograms cause the left of the image (smaller time-scales) to be "washed out" when the color map is applied to the function range [0, 0.2] of the entire image (a). To remedy this, the same color map (b) must be dynamically mapped only to the function range, i.e., [0, 0.05], in the visible area to improve the visualization (c).

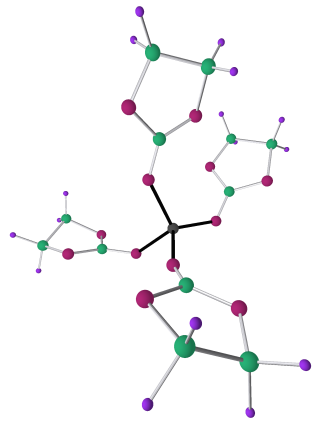


Figure 7: A lithium ion (dark gray) is shown at the center, and is bonded to four carbonyl Oxygen atoms (magenta), which are part of the EC molecules.

sample set of size S, we pick a random subset of size R without replacement and compute the histogram on the subset. Analyzing the differences between different random samples and especially the variance among them provides an indication of the sampling quality of a given histogram. For example, a very low error with $R = 0.5 * S$ would indicate that even using half the samples, the histogram would not change, indicating that the sample size is sufficient. The two important parameters for jack-knifing are the number of resampling steps and the size of the resampled subset R. In order to choose these parameters, we study the convergence of histogram errors with respect to the number of sampling steps for various sizes of the resampled subset. Figure 8 shows a log-log graph of error metrics on the vertical axis with the number of resampling steps on the horizontal axis. Different curves represent the mean L_2 and mean L_∞ errors for subset sizes ranging from 40% to 90% of S. The figure shows that the errors converge after about 500 resampling steps, irrespective of the size of the resampled subset. Similar behavior was observed in the convergence of histograms of other values of Δ and B. Therefore, with the goal of making a moderately conservative choice, we pick $R = 70\%$ of S, and perform 1000 resampling steps for the visualizations of data uncertainty.

We note that uncertainty computation is, unsurprisingly, a computationally intensive process, and currently cannot be performed interactively. In our unoptimized implementation, jack-knifing a single histogram using the above-mentioned parameters takes up to 36 seconds for the shortest time-scales with the largest sample sizes. Therefore, to preserve interactivity, we precompute data uncertainties for various bin sizes and store the results in separate files for each Δ. This data is then read on demand by the analysis tool, and information is updated to the user in real time.

Figure 9 shows the evolution of data uncertainty with time-scale in a log-linear plot. The error metric used here is the mean L_2 error normalized with respect to the number of samples available for each time-scale. It can be concluded that for small time-scales (less than 5 – 6 ps), the computed histogram captures the distribution well: an expected behavior since a large number of samples are available for small time-scales, and therefore, the distribution can be approximated more accurately.

5.2 Visualization Uncertainty

Given N histograms with B bins each, this section discusses how to visualize this information using an output resolution of $n \times b$

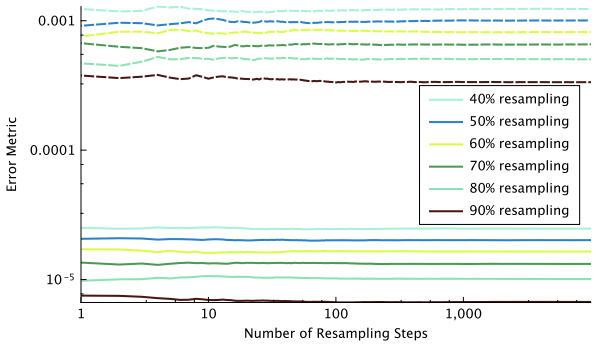


Figure 8: Mean L_2 (solid lines) and mean L_∞ (dashed lines) in the computed PDFs appear to have converged after about 500 steps of bootstrapping using any of the subset sizes which we experimented.

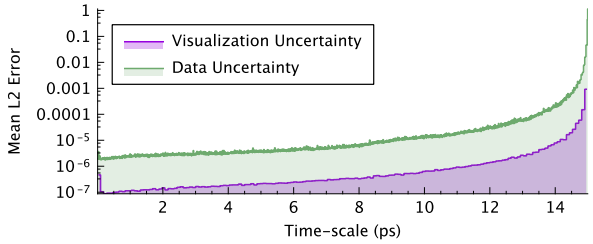


Figure 9: Visualizing data and visualization uncertainties in histograms for all time-scales.

through aggregation. The typical value of B is 180, with the bin-resolution of one degree per bin, whereas a typical N can be much greater. In the data used for our experiments, N equals 2999. On the other hand, the typical resolution used for such 2D plots by standard viewing is of the order of a few hundreds in each dimension.

In order for each aggregated data point to be visually discernible, we show it using $k \times k$ pixels. Therefore, for the given output resolution $n \times b$, we reduce the data from $N \times B$ to $n/k \times b/k$. In the visualizations presented here, we use $k = 3$. However, to keep the discussion simple, we assume $k = 1$ without loss of generality and describe $N \times B$ to $n \times b$ reduction. Depending upon the input and output resolutions, we can have the following cases:

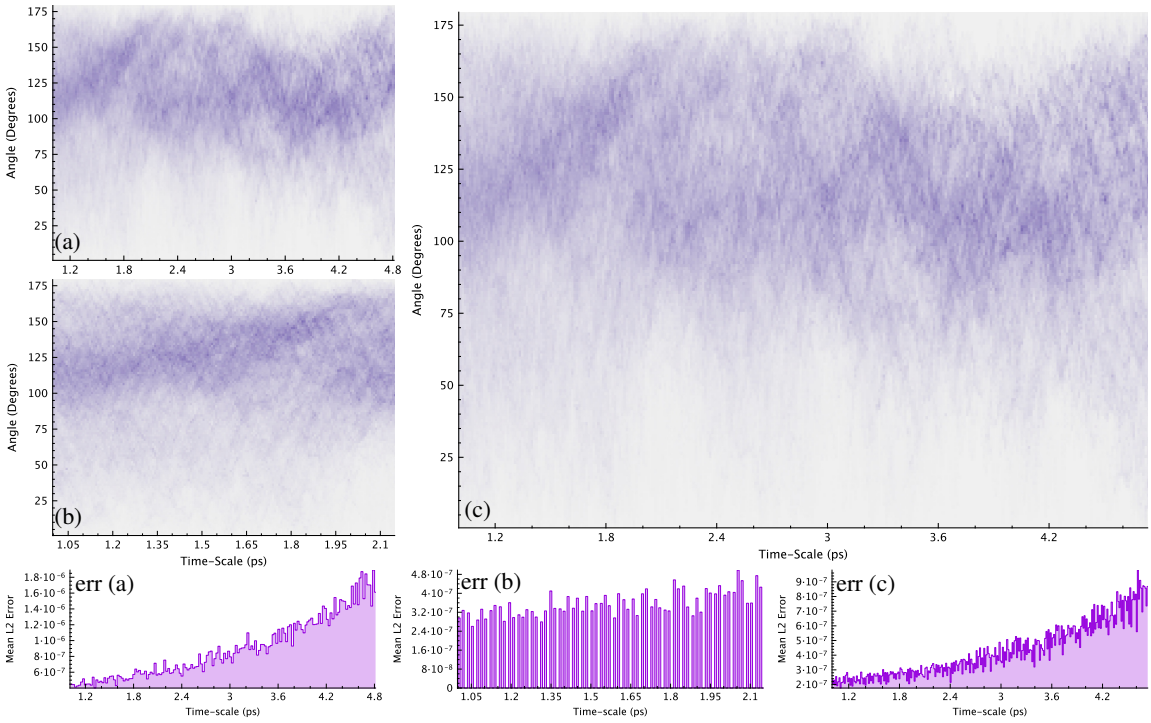


Figure 10: Visualization error in a given 2D image (a) can be reduced by either zooming in with in the limited resolution (i.e., reducing N for a given n) (b) or increasing the total available resolution (c) (i.e., increasing n for a given N).

Case 1: $b < B$**.** When the output resolution allows displaying fewer than B bins, we must recompute 1D histograms for the smaller number of bins. We emphasize that it is not enough to simply aggregate B bins into b, because that can cause round-off errors. For example, consider the simple case of $B = 3$ and $b = 2$. The correct width of output bins is 90°, but collapsing 3 bins into 2 will make the bin widths uneven: one output bin will represent a 60° span whereas the other will represent a 120° span. The uneven bin-widths will skew the histogram and mislead the analysis.

Case 2: $n < N$**.** This case requires aggregating N histograms into n. In particular, each output histogram H_i where $0 \leq i < n$ is computed by merging m input histograms $\{H_j\}$, for all $j \in [\mathcal{M}(i),\ \mathcal{M}(i+1))$. The mapping $\mathcal{M}$ must be injective, monotonic, and ensure that $\mathcal{M}(0) \geq 0$ and $\mathcal{M}(n-1) \leq N-1$. We use Bresenham's rasterization algorithm [4] to define this mapping, i.e., $\mathcal{M}(i) = N \cdot i/n$, rounded to the nearest integer.

Assuming $B = b$, merging of histograms can be performed by averaging corresponding bins. Although one may consider other more-sophisticated kernels to perform this merging, we believe that averaging (rectangular kernel) is the best choice since it weighs input histograms uniformly. Furthermore, we perform this aggregation of (unnormalized) frequencies rather than distributions. In this sense, we are creating a single histogram covering the considered range of time-scales, rather than an average distribution. As before, we normalize the resulting histogram to create a sampled distribution. Finally, we compute the L_2 error for every input histogram with respect to the output histogram and use the maximum value as the visualization error for the output histogram.

As n decreases, more histograms must be aggregated, leading to a higher visualization error. Therefore, unsurprisingly, in order to reduce the visualization error, one requires a higher output resolution, as illustrated in Figure 10.

Case 3: $b > B$ **and/or** $n > N$**.** When either dimension of the output texture is greater than the corresponding dimension of the input data, we must be careful to repeat data values to create an $n \times b$ texture. If only an $N \times B$ texture is created, the underlying graphics renderer usually stretches the texture to output size, e.g., OpenGL uses linear interpolation to map the texture size to data size. As a result, the visualization gives a false sense of continuity in the data. Furthermore, this interpolation is performed on the color map, rather than on the function values. The final color further depends on the color map used, making the effect of such color interpolation even more unpredictable. As a result, the duplication of values is important, and we use the mapping $\mathcal{M}$ for this purpose.

6 Design Choices and Visualization Tool

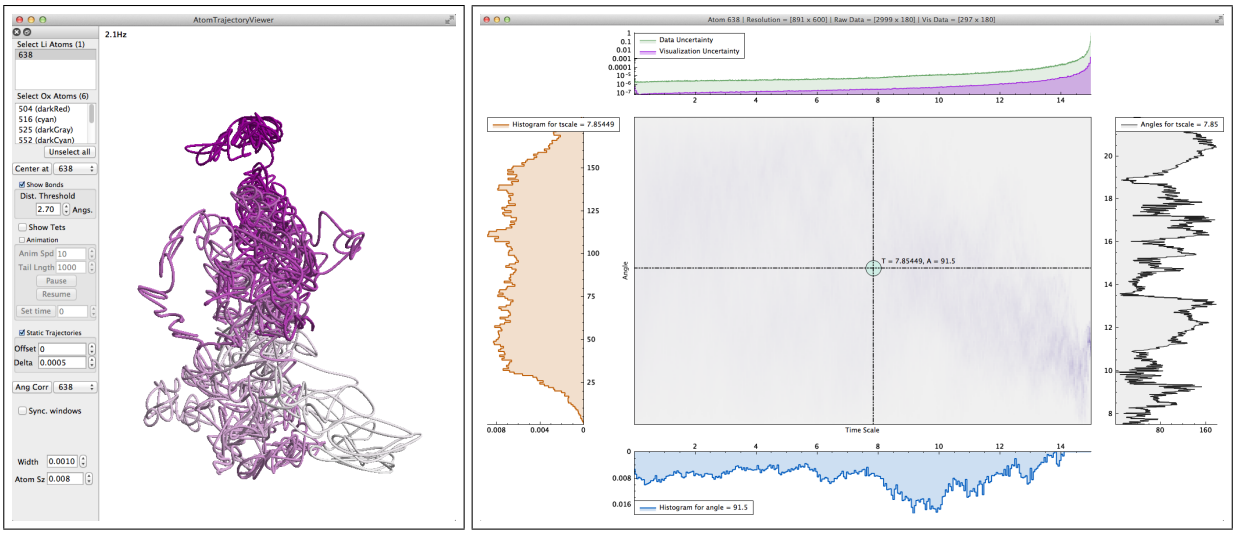
Here, we present an interactive tool for the visualization and analysis of atomic trajectories, also demonstrated in the supplemental video. Illustrated in Figure 11, our tool has two main components: the 3D spatial visualization of trajectories and the statistical visualizations of relative-angle distributions. The user can choose to link both components into an integrated environment of one or more 2D windows for statistical visualization of individual trajectories and a single 3D viewer visualizing all selected trajectories. The input to the main application is a set of atomic trajectories as sequences of space-time positions.

6.1 3D Spatial Visualization of Trajectories

The default application window offers the following features:

Atom selection and centering. To avoid visual clutter, the user can choose the atoms for which trajectories should be displayed. Furthermore, since the domain is periodic in all three dimensions, it is often useful to display the data in the reference frame of a chosen atom. To facilitate this, the user can pick any atom around which the visualization is centered, or view the trajectories in the original reference frame of the simulation.

Chemical bond visualization. Trajectory analysis has been augmented with the visualization of chemical bonds. In the molecular system under consideration, lithium ion bonds with four oxygen atoms, which form a tetrahedron that encloses lithium (cf. Figure 7). Visualizing bonds and tetrahedral shapes helps clarify the interplay between diffusion and solvation.



(a) 3D visualization of trajectories.

(b) 2D visualization of relative-angle distribution statistics.

Figure 11: Screenshots of our integrated analysis and visualization tool for interactive exploration of atomic trajectories.

Animated and static trajectories. Besides showing the entire trajectories with respect to a chosen starting time ("offset") and sampling (Δ), the tool can also animate them to understand the temporal dimension in the system. The user can control the animation speed, manually set time-step for display, and also view a short past history (controlled by the "tail length" parameter in Figure 11a).

6.2 Visualization of Relative-Angle Distributions

The user can interactively trigger the computation and visualization of relative-angle distributions for any atom. For a holistic understanding of the statistics, all information is displayed within a single application window, as illustrated in Figure 11b.

Multiple distributions as a 2D image. The (potentially) reduced data is visualized as a 2D image at the center of the window. The horizontal and vertical axes represent time-scales and angles, respectively. Nevertheless, the available resolution is often wider than it is high, in which case, the resolution available for angles is typically lower than that for the time-scales. We provide the functionality to toggle the orientation so that angles map to the horizontal axis, whereas time-scales vary along the vertical axis. For simplicity, however, the following discussion assumes time-scales on the horizontal axis and angles on the vertical axis.

1D histograms. The user can select a point on the 2D image using a crosshair, which spans the entire image with a circle around the selected point. The 1D histograms corresponding to the selected values of angle and time-scale are shown below and to the left of the image. Since the horizontal axes of both the 2D image and the bottom plot represent time-scales, they are synchronized with respect to zooming and panning. Furthermore, the bottom plot is vertically inverted, i.e., the horizontal axis is at the top instead of the bottom, and the values on the vertical axis increase downward. Similarly, the vertical axes of the 2D image and the left plot are synchronized, and the latter is horizontally inverted. This layout provides a more coherent visualization, leading to a more intuitive interaction, and saves space by making redundant axes dispensable.

Uncertainty plots. A plot containing data and visualization uncertainties is displayed at the top of the 2D image, with a synchronized time-scale axis. The values for visualization uncertainty are sampled using the visualization data, i.e., n time-scales, whereas those for data uncertainty are sampled using the original data, N time-scales. Nevertheless, displaying both errors in the same plot gives a sense of their absolute as well as relative scales.

Angle vs. time. The angle vs. time for the selected time-scale is shown to the right of the center plot. This plot shows the function $\theta(t;\Delta)$ for the time-scale Δ selected, with absolute time t on the long axis. Note that this plot uses raw data, and does not depend on the reduced data for visualization.

User Interactions. The following user interactions are supported:

- The user can *zoom and pan* within the 2D image, with synchronized updates to axes and secondary plots.
- Upon *window resizing*, the new size available for displaying the 2D image is computed. New visualization data is computed by reducing the raw data, and visualizations are updated.
- When the user *moves the selection*, all four graphs around the 2D image are updated with respect to the new selection.

These interactions are linked through multiple visualization windows, such that selection in one is reflected in all others as well.

7 Evaluation

We have deployed our tool to explore FPMD simulations of lithium-ion cells (see Section 3). The main advantage of our approach is that it enables an interactive exploration of large-scale simulation data, which until now has required tedious manual effort. In particular, our collaborators have been able to easily analyze the relative-angle distributions for various atoms in a synchronized manner.

The first goal of the exploration was to confirm some expected features in the relative-angle distribution, which previously could not be tested due to practical time constraints. The first of these features is the attraction and repulsion of surrounding molecules inhibiting Brownian or ballistic motion. Through much of the simulation, the lithium ion is bonded to four oxygen atoms, and previous analysis has shown a strong cyclic behavior of the pair-wise distances between the lithium ion and its bonded oxygens with a period of ≈ 0.1 ps. Zooming into the angle distributions at around $\Delta = 0.05$ ps (Figure 12) shows a strong peak at about 150°, suggesting that lithium has a distinct tendency to reverse at the expected time-scale of traversing the tetrahedron formed by the surrounding oxygen atoms. This confirms the expectation that on short time-scales, the motion is dominated by the constraints due to bonds.

A second assumption was that histograms at longer time-scales are less reliable than those at shorter ones, as fewer samples are available. Although this general trend is observed (cf. Figure 9), the behavior is somewhat unexpected. The number of available

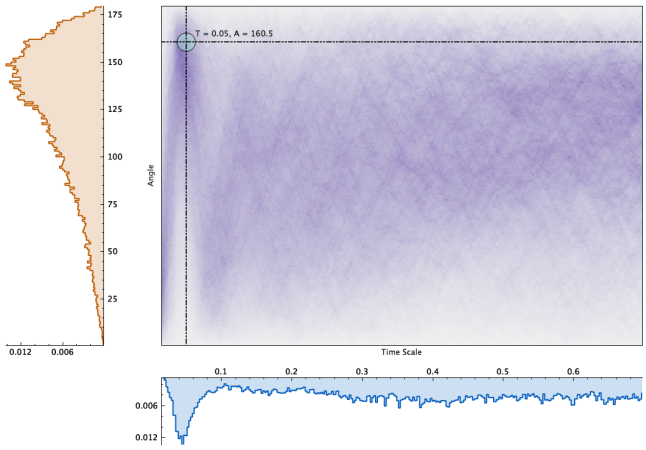


Figure 12: The 2D histogram at $\Delta = 0.05$ ps, and $\theta \approx 150°$, along with the two respective histograms along the corresponding axes, indicate coherent behavior correlated with confined motion.

samples decreases linearly as the time-scale increases, but the error shows highly non-linear behavior, with most increases concentrated in the longest time-scales. This behavior may indicate that this error is not necessarily due to the decreasing number of samples, but more likely due to an increasing complexity of the histograms. The latter argument would also explain why the visualization uncertainty changes even more drastically (though at a smaller magnitude), as the aggregation of more complex and faster-changing histograms will cause more errors. In collaboration with the domain experts, we are currently discussing potential implications.

Next, we exploit interactivity to compare the histograms of different species of atoms. Figure 13a compares the lithium trajectory to that of the phosphorous atom in the PF_6^- ion, which shows markedly different behavior compared to lithium. The peaks in this distribution are not as prominent, and the distribution is wider, indicating a more Brownian-like motion. Comparing oxygen atoms that always bond with lithium with those that never bond also demonstrates interesting behavior. As expected, the relative-angle distribution for bonded oxygen atoms is similar to that of lithium for anything but very short time-scales as their motion is coupled. Nevertheless, the similarities are not as strong as one might expect. In particular, the similarities do not increase for longer time-scales, indicating that the independent local motions still have a significant influence on the relative angle distribution. The unbonded oxygen atoms, on the other hand, show a more uniform histogram, much like phosphorous, reflecting a more diffusive behavior.

7.1 User Feedback

The initial feedback from our collaborators has been positive. The tool has an intuitive interface with a short learning curve and has been used independently by the scientists. In particular, they have been able to explore trajectories in a number of different simulations with different electrolytes and physical parameters, e.g., external voltages. One important goal of the exploration was to understand whether lithium drags the bonded molecules, or switches oxygens by breaking and re-forming bonds as it moves through the liquid. The interactive analysis enabled by our framework has led our collaborators toward the former hypothesis in which lithium's motion is affected by its bonded molecules, which matches the conjecture by Xu et al. [27], and would help to explain why lithium diffuses slower than PF_6^-, which does not bond.

The interactive visualization of relative-angle distribution has been effective as well. By reducing the time to display from minutes to seconds, our tool significantly lowers the barrier to create different histograms. As a result, the domain experts have become interested in cross-comparisons of relative-angle distributions between different atoms, opening a new direction for exploration. Furthermore, dynamic zooming helps in the observation of small structures, such as restricted motion, and the corresponding rescaling of the color map significantly improves the differentiation of small-scale features (see Figure 6). The additional information on the 1D slices as well as the time-dependent angle provide a fast way to explore details of the distributions that are difficult to extract from the 2D image. Similarly, the direct link between the statistical analysis and the 3D trajectories makes the entire process more intuitive and faster, with corresponding improvements in productivity. Finally, the ability to directly evaluate uncertainty in both data and visualization provides an important level of confidence in the analysis and the resulting plots. Our framework, for the first time, allows a holistic and interactive exploration of atomic trajectories and their corresponding statistics. In the words of our collaborators, this tool will save substantial amounts of time and effort, and as discussed, has already led to new questions that would have been too time consuming to explore previously.

8 Conclusion and Future Work

This paper presents an interactive visualization and analysis tool for the exploration of atomic trajectories, along with their relative-angle distributions. Our integrated solution provides a synchronized visualization to enable exploring parameters, such as time-scales, in both 3D space and parameter space, to effectively highlight phenomena, such as Brownian vs. non-Brownian motion, restricted motion due to neighboring molecules, etc.

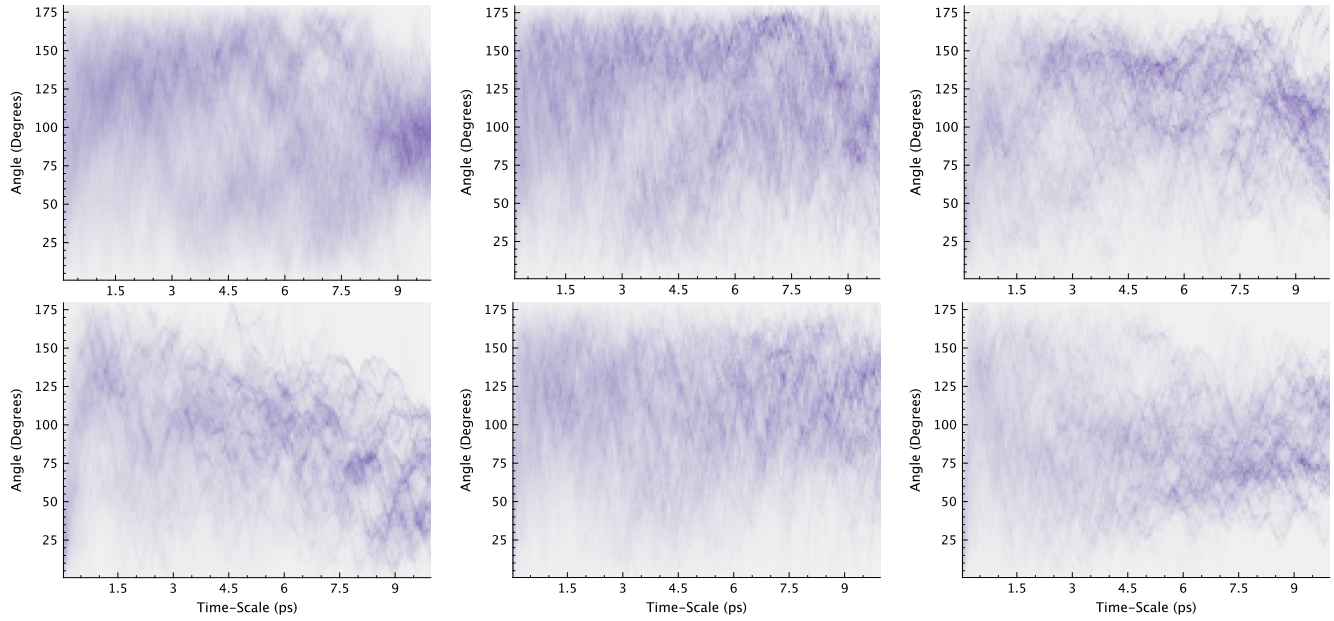
This work addresses important challenges in the visualization of 2D images representing collections of 1D histograms. In particular, the problems of under-plotting and over-plotting, which can cause spurious and misleading visualizations due to blending and interpolation in color space performed by the rendering engine, have been addressed. By systematically recomputing the histograms depending upon the available output resolution, we provide explicit control of the visualization. Furthermore, we visualize the uncertainties in both data and visualization to allow making scientific claims with improved confidence.

We have demonstrated our tool using first-principles molecular dynamics simulation of lithium-ion systems. The tool is available to domain experts and has been effectively utilized to validate hypotheses and explore in more detail phenomena that are not yet well understood. In addition to the application presented in this paper, the tool can be applied, with minor adaptations, to a variety of other applications where the relative-angle distribution describes the behavior of complex trajectories.

Going forward, we are working toward understanding data uncertainty in more detail. In particular, we wish to understand the statistical dependence between adjacent samples along a trajectory and its relevance for the underlying physical phenomena. It is also important to better understand and characterize the various high-frequency variations in the 2D visualizations and separate physically meaningful features from noise.

Acknowledgements

Support for this work was provided through Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy (DOE), Office of Science, Advanced Scientific Computing Research and Basic Energy Sciences. This work is also supported in part by BNSF CISE ACI-0904631, NSG IIS-1045032, NSF EFT ACI-0906379, DOE/NEUP 120341, DOE/Codesign P01180734, DOE/SciDAC DESC0007446, DOE-EXDE-SC0010498, and CCMSC DE-NA0002375. This work was performed under the auspices of the DOE by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. LLNL-CONF-678261.



(a) Li^+ (top) with PF_6^-(bottom) ions.

(b) Two O atoms that always bond to Li.

(c) Two O atoms that never bond to Li.

Figure 13: Comparison of relative angle distributions of different species of atoms in the simulation.

REFERENCES

[1] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1428–1435, Nov 2008.

[2] M. D. Bhatt, M. Cho, and K. Cho. Density functional theory calculations and ab initiomolecular dynamics simulations for diffusion of Li^+ within liquid ethylene carbonate. *Model. Simul. Mater. Sci. Eng.*, 20(6):065004, July 2012.

[3] W. J. T. Bos, B. Kadoch, and K. Schneider. Angular Statistics of Lagrangian Trajectories in Turbulence. *Phys. Rev. Lett.*, 114(21):214502, May 2015.

[4] J. Bresenham. Algorithm for computer control of a digital plotter. *IBM Syst. J.*, 4(1):25–30, 1965.

[5] S. Burov, S. M. A. Tabei, T. Huynh, M. P. Murrell, L. H. Philipson, S. A. Rice, M. L. Gardel, N. F. Scherer, and A. R. Dinner. Distribution of directional change as a signature of complex dynamics. *Proc. Natl. Acad. Sci.*, 110(49):19689–19694, Dec. 2013.

[6] J. Cottam, A. Lumsdaine, and P. Wang. Overplotting: Unified solutions under abstract rendering. In *IEEE Int. Conf. on Big Data*, pages 9–16, Oct 2013.

[7] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, 1997.

[8] S. Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, Oakland, California, 2009.

[9] P. Ganesh, D.-e. Jiang, and P. R. C. Kent. Accurate static and dynamic properties of liquid electrolytes for Li-ion batteries from ab initio molecular dynamics. *J. Phys. Chem. B*, 115(12):3085–3090, 2011.

[10] I. Golding and E. C. Cox. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.*, 96:098102, Mar 2006.

[11] G. Grolemund and H. Wickham. Visualizing complex data with embedded plots. *J. Comput. Graph. Stat.*, 24(1):26–43, 2015.

[12] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1531–1538, Nov. 2009.

[13] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[14] I. Izeddin, V. Récamier, L. Bosanac, I. I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, and X. Darzacq. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife*, 3:e02230, June 2014.

[15] R. Kosara, F. Bendix, and H. Hauser. Timehistograms for large, time-dependent data. In O. Deussen, C. Hansen, D. Keim, and D. Saupe, editors, *IEEE VGTC Symp. on Vis.* The Eurographics Association, 2004.

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

[17] 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.

[18] 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. Phys. Chem. B*, 119(4):1535–1545, Jan. 2015.

[19] M. N. Phadke, L. Pinto, O. Alabi, J. Harter, R. M. Taylor II, X. Wu, H. Petersen, S. A. Bass, and C. G. Healey. Exploring ensemble visualization. In P. C. Wong, D. L. Kao, M. C. Hao, C. Chen, R. Kosara, M. A. Livingston, J. Park, and I. Roberts, editors, *IS&T/SPIE Electronic Imaging*, page 82940B. SPIE, June 2012.

[20] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *IEEE Int. Conf. on Data Mining Workshops*, pages 233–240, Dec 2009.

[21] G. Ramos-Fernandez, J. L. Mateos, O. Miramontes, G. Cocho, H. Larralde, and B. Ayala-Orozco. Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). *Behav. Ecol. Sociobiol.*, 55(3):223–230, Jan. 2004.

[22] J. Savage and G. A. Voth. Persistent Subdiffusive Proton Transport in Perfluorosulfonic Acid Membranes. *J. Phys. Chem. Letters*, 5(17):3037–3042, Sept. 2014.

[23] M. Son, G. Lee, J. Son, S. Choi, Y. Kim, S.-Y. Lee, Y.-R. Yoon, D. S. Yoon, and S. W. Lee. Characterization of anomalous movements of spherical living cells on a silicon dioxide glassy substrate. *Biomicrofluidics*, 9(1):014102, Jan. 2015.

[24] C. Song, Z. Qu, N. Blumm, and A. L. Barabasi. Limits of Predictability in Human Mobility. *Science*, 327(5968):1018–1021, Feb. 2010.

[25] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415, May 1996.

[26] M. Woringer, X. Darzacq, and I. Izeddin. Geometry of the nucleus: a perspective on gene expression regulation. *Curr. Opin. Cell Biol.*, 20:112 – 119, 2014. Molecular imaging.

[27] K. Xu. Nonaqueous Liquid Electrolytes for Lithium-Based Rechargeable Batteries. *Chem. Rev.*, 104(10):4303–4418, Oct. 2004.