Article

Estimates of Electrical Conductivity from Molecular Dynamics Simulations: How to Invest the Computational Effort

Piotr Kubisiak and Andrzej Eilmes*
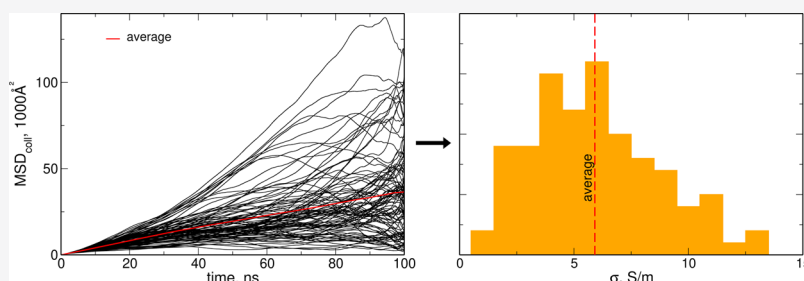
Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Although the electrical conductivity of an electrolyte can be estimated from the molecular dynamics trajectory, it is often a challenging task because of the need to obtain a substantial amount of data to ensure sufficient averaging. Here, we present an analysis on the convergence of results with the number of simulated trajectories. A series of molecular dynamics simulations have been performed for a model electrolyte (NaCl in water) and the Einstein relation has been used to calculate the electrical conductivity. The standard deviation of the conductivity estimates is relatively large compared to the mean value, and it has been shown that the off-diagonal contributions to the collective displacement of ions are responsible for large deviations between systems. It has been found that about 40 independent MD simulations may be required to reduce the errors. A procedure to improve the final estimate of the conductivity has been proposed.

1. INTRODUCTION

Equilibrium molecular dynamics (MD) simulations are today a standard tool for modeling systems in the condensed phase and liquids in particular.[1,2] MD delivers invaluable information on the structure and dynamics of complex molecular systems; therefore, it is frequently used in various disciplines ranging from material science to biochemistry.

A fast growing area of the application of MD is the research on power sources including studies on physicochemical properties of electrolytes. In this context, obtaining estimates of transport-related properties of the electrolyte, diffusion coefficients and electrical conductivity, is of great importance. Appropriate averaging of data is necessary in order to reliably calculate desired values. Generally, it is easier to achieve sufficient averaging in the case of diffusion coefficients because it may be improved by increase in the size of studied systems. Issues related to the derivation of diffusion coefficients have already been discussed in several studies.[3−5]

Computations of electrical conductivity from MD simulations are more demanding. Complications arise because conductivity is a collective property, governed by the response of all ions in the system. Therefore, increasing the system size will not reduce fluctuations efficiently. Similar effects are encountered in estimates of shear viscosity; in such a case, a method based on multiple independent trajectories was proposed.[6]

Because of problems with computations of conductivity from MD trajectories, there are significantly less papers attempting such estimates, compared to studies presenting diffusion coefficients, which are much easier to obtain. Moreover, quite often, such studies do not compute the conductivity exactly but instead use the Nernst−Einstein relationship to obtain the conductivity based on diffusion coefficients derived from MD simulations.[7] Such an approach assumes that the motions of individual ions are not correlated; therefore, it will be problematic in systems where correlations are supposed to be significant, such as concentrated salt solutions or ionic liquids.

In this work, we want to check how well increasing computational effort can improve the quality of conductivity estimates. As the increase in the system size is not an efficient

option, we will ask how many independent repetitions of MD simulations may be sufficient to obtain a reasonable average. Alternatively, we will consider the possibility of increasing the length of the MD trajectory. For this purpose, we will simulate a salt solution in a molecular liquid ($NaCl$ in H_2O) as a model electrolyte. We will compare the approaches based on increasing the number of trajectories or increasing the length of simulations and formulate some suggestions on the investment of computational time. A simple method to correct the estimate will be proposed.

2. COMPUTATIONAL DETAILS

Our model system is 1 M NaCl solution in water. Two sets of structures were prepared. The major part of simulations was performed for systems containing 23 pairs of Na^+ and Cl^- ions in 1165 H_2O molecules. Additionally, we also used a set of structures with a larger number of entities, composed of 93 ion pairs and 4772 water molecules. VMD v. 1.9.2[8] was used to build the initial structures.

Force field parameters of Lennard-Jones potentials for Na^+ and Cl^- ions were taken from refs 9 and 10, respectively. TIP3P parameterization[11] was used for water molecules. It should be noted that in our study, we do not aim for correct reproduction of properties of the solution; therefore, the choice of a particular version of force field is of lesser importance; the parameterization used here was chosen for the simplicity of setting up the simulations.

The NAMD v. 2.12[12] simulation package was used to perform the MD simulations under periodic boundary conditions in the NpT ensemble at $p = 1$ atm and $T = 293$ K with Langevin dynamics and a modified Nose–Hoover–Langevin barostat.[13,14] Electrostatic interactions were treated via the particle mesh Ewald algorithm.[15] A time step of 1 fs was used.

A series of 200 ns long MD simulations was performed for 100 systems with 23 ion pairs. For three systems of this size, we collected much longer trajectories, lasting 2 μs. Hereafter, in the text, we will refer to these sets of data as "short" and "long" trajectories. Three larger systems (with 93 ion pairs) were simulated for 1.1 μs ("big" trajectories). The first 100 ns of each MD trajectory was treated as the equilibration stage, with the remaining part used for the analysis of the conductivity. Therefore, in figure captions, short and long trajectories will be marked as "100 ns" or "1.9 μs", respectively.

Conductivity of the electrolyte was estimated from the MD trajectories using the Einstein relation[16]

$$\sigma = \lim_{t \to \infty} \frac{e^2}{6tVk_BT} \sum_{i,j} z_i z_j \langle [\mathbf{R}_i(t) - \mathbf{R}_i(0)][\mathbf{R}_j(t) - \mathbf{R}_j(0)] \rangle \quad (1)$$

where t stands for time, V is the volume of the simulation box, k_B is Boltzmann's constant, T is the temperature, e is the elementary charge, z_i and z_j are the charges of ions i and j, respectively, $\mathbf{R}_i(t)$ is the position of i-th ion at time t, and the brackets $\langle \rangle$ denote the ensemble average.

The diffusion coefficients $D_{\pm}$ of cations and anions were calculated from the slopes of the time dependences of the mean square displacements (MSDs) of ions

$$D_i = \lim_{t \to \infty} \frac{1}{6t} \langle |\mathbf{R}_i(t) - \mathbf{R}_i(0)|^2 \rangle \quad (2)$$

The conductivity is proportional to the collective ion diffusion coefficient

$$D_{coll} = \lim_{t \to \infty} \frac{1}{6tN} \sum_{i,j} z_i z_j \langle [\mathbf{R}_i(t) - \mathbf{R}_i(0)][\mathbf{R}_j(t) - \mathbf{R}_j(0)] \rangle \quad (3)$$

N is the total number of ions in the system. D_{coll} reduces to the average of anion and cation diffusion coefficients $D_{avg} = (D_- + D_+)/2$ in the case where there are no correlations between movements of different ions, that is, when the off-diagonal ($i \neq j$) terms in eq 3 are negligibly small. In such a case, the Nernst–Einstein equation can provide a good estimate of the conductivity.

The conductivity given by formula 1 may be further broken down into contributions arising from different correlations between movements of ions. For this purpose, we write the total conductivity σ of the system as

$$\sigma = \sigma_{Na} + \sigma_{Cl} + \sigma_{Na-Na} + \sigma_{Cl-Cl} + \sigma_{Na-Cl} \quad (4)$$

The diagonal ($i = j$) terms σ_{Na} and σ_{Cl} are related to self-diffusion of Na^+ and Cl^- ions, respectively. The off-diagonal terms arise from correlations between different ions: cation–cation (σ_{Na-Na}), anion–anion (σ_{Cl-Cl}), and cation–anion (σ_{Na-Cl}). In the literature, the diagonal terms σ_{Na} and σ_{Cl} are named "self" contributions, whereas the off-diagonal terms related to ions of same charge σ_{Na-Na} and σ_{Cl-Cl} are referred to as "distinct" contributions.

It should be noted that the diffusion coefficients can be obtained from MSDs only if the data are sufficiently averaged, that is, over many ions or/and over sufficiently long time. Accordingly, in order to calculate the conductivities or the diffusion coefficients from formula 1 or 2, usually, the data are averaged over all possible choices of time intervals Δt within the total simulation time of the trajectory and the estimates are obtained from the slope of the MSD dependence versus time. The formulas 1–3 use the limit of infinitely long time; however, in practice, the data for long times are noisy because of a smaller number of possible choices of long time intervals, resulting in data averaging much worse than for short times. On the other hand, short-time data may not reflect properly the long-time behavior of the system—a sufficiently long time is needed to observe the diffusive regime. Additional complication arises from the fact that the data for different positions of the Δt within the single trajectory are correlated, which affects the statistics, particularly for short times.[4,5] Therefore, in the choice of the time range used to estimate the conductivity, one seeks an optimal balance of the above-mentioned factors.

3. RESULTS

3.1. Short Trajectories. In Figure 1, we show the collective MSD for 100 individual systems, the diagonal contribution of the Na^+ ions to the MSD, and the off-diagonal contribution arising from Na^+–Cl^- correlations. It is readily noticeable that the curves for different systems lie close to each other only at very short times and diverge with increasing time. The divergences are the smallest for the diagonal cation contributions and simultaneously, in this case, the MSD curves are the most linear. Nonlinearity and differences between systems are larger in the case of the collective MSDs and the off-diagonal anion–cation contribution to the MSD_{coll} may even differ in sign between systems. Approximate linearity of
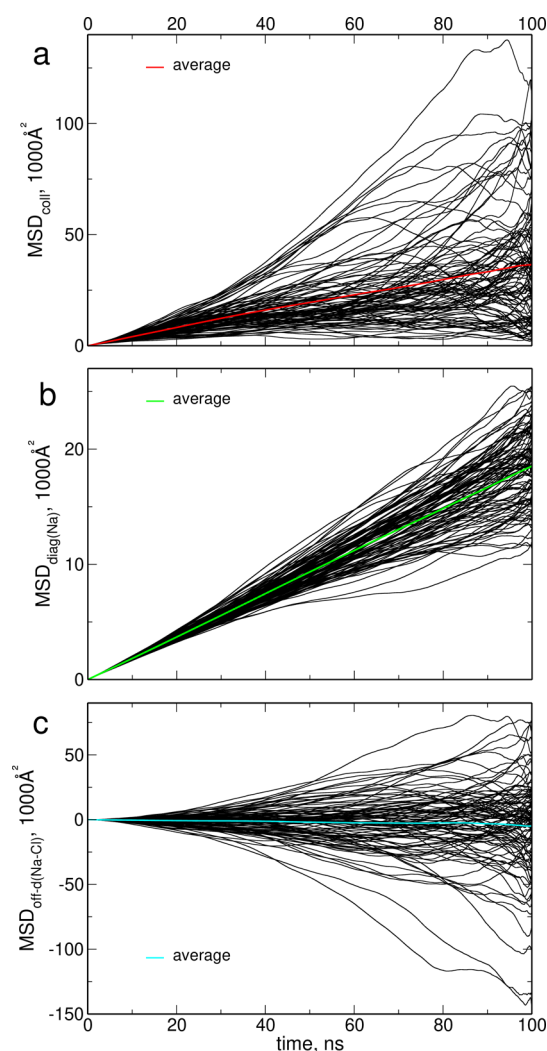
Figure 1. Collective MSD (a), diagonal contribution to MSD_{coll} from Na^+ cations (b), and off-diagonal cation−anion contributions to MSD_{coll} (c) obtained from 100 ns trajectories. Color lines are the averages over 100 systems.

Figure 2. Average collective MSDs and different contributions to MSD_{coll} obtained from 100 short (100 ns) MD simulations. Data for off-diagonal contributions are shown in detail in the bottom panel.

the initial part of the MSD curves (up to 10−20 ns) is a consequence of better data averaging over short time intervals. With the increasing length of simulated trajectories, the averaging over time intervals chosen from the whole trajectory improves the quality of the plot also for larger times and the linear part of the curve becomes longer. However, increasing the length of the trajectory requires an investment of computational time; we will come back to this issue later.

As seen in Figure 1, averaging the data over different systems yields the averaged MSD curves, which are much less noisy and quite linear within the scale of the plot. Only after such an averaging one may conclude that the off-diagonal anion−cation contribution in our system is negative—leads to a decrease in the total conductivity. In Figure 2, we show the different contributions to the collective MSDs averaged over all 100 simulated systems. The averaged diagonal contributions of cations and anions exhibit linear MSD versus time upon 100 ns. These positive contributions are proportional to the diffusion coefficients of Na^+ and Cl^- ions, and therefore, the result is not surprising, confirming common knowledge that the self-diffusion coefficients can be relatively easily estimated from the MD trajectories. All off-diagonal contributions arising
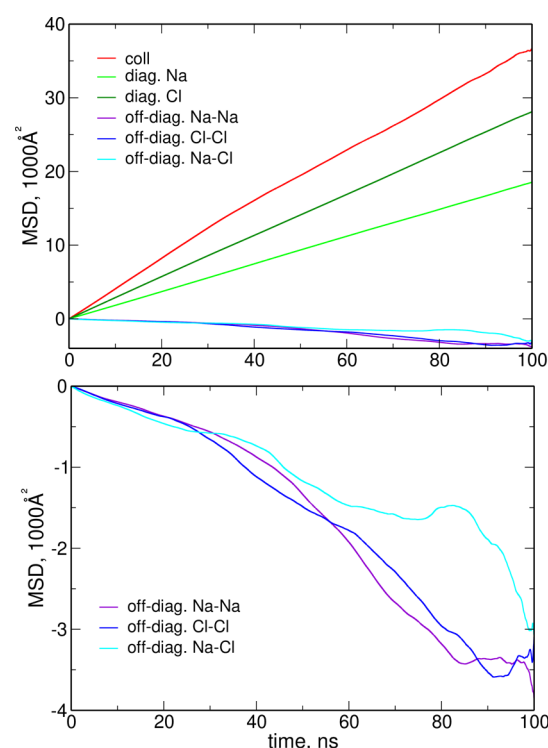
from cation−cation, anion−anion, and cation−anion correlations are negative. It has to be expected that for a salt solution in a molecular liquid, the correlations between ion motions decrease the conductivity of the electrolyte. We should note that the averaging over different systems was necessary to see the negative sign of the off-diagonal contributions because the contributions obtained for individual systems differ in sign, as shown for the Na−Cl contribution in Figure 1c. These contributions describe the effect of correlations between ion motions, and therefore, accidentally obtaining the wrong sign from insufficiently averaged data could lead to unphysical conclusions about ion transport in the system. The off-diagonal contributions to the MSD_{coll} are much noisier that those of diagonal components and the corresponding MSD plots are approximately linear only at short times. The absolute values of the diagonal contributions to the conductivity are about an order of magnitude larger than those of the off-diagonal terms. Note, however, that this statement is true only for the data averaged over different systems or at short times (where averaging over time intervals is sufficiently effective) because as seen in Figure 1b,c, the span of MSD curves for individual systems at long times is comparable for both diagonal and off-diagonal components. The behavior of the collective MSDs is therefore dominated by the diagonal contributions, linearly increasing with time and the MSD_{coll} curve shown in the upper panel of Figure 2 is rather linear. There is some noise at long times, resulting from the fluctuations of the off-diagonal contributions. A closer look at the details reveals also that the line changes slightly its slope between 30 and 40 ns, and again, this feature stems from the behavior of the off-diagonal components.

From Figure 2 and the linearity of the diagonal MSDs up to 100 ns, we may expect that the choice of the time range within

this range will have little impact on the estimates of the diffusion coefficients of ions from the slope of the MSD curve. For calculations of the conductivity, the choice of the time window used to estimate the slope of the MSD_{coll} is more important because the curve is not a perfectly straight line—yet the differences will not be too large (as we will show later). On the other hand, if one wants to estimate the transport properties of individual systems (Figure 1), the time window is rather limited because above about 40−50 ns, the deviations from linearity become quite significant. Nevertheless, obtaining the estimates from individual systems enables us to get some information on the distribution of the values and their deviation (the mean value of all systems will be of course the same as estimated from the system-averaged MSD curve).

In Figure 3, we summarized the information on the conductivity values estimated for individual systems from the
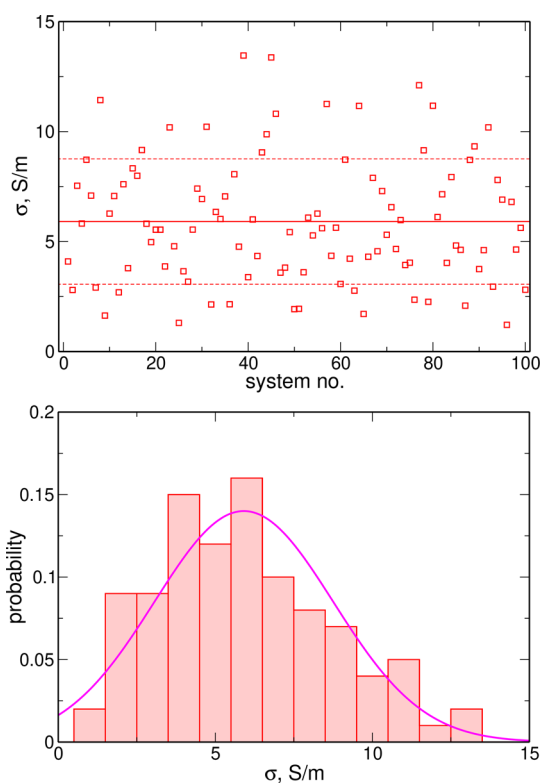

Figure 3. Estimates of the conductivity obtained from individual 100 ns trajectories using the time window $M = 10−20$ ns (upper panel) and histogram of estimated values with a fit of Gaussian distribution (bottom panel). Solid and broken lines in the upper panel mark the mean value and the standard deviation, respectively.

slope of the MSD_{coll} in the time window 10−20 ns. The results span the range 1.2−13.5 S/m with the mean value 5.9 S/m and the standard deviation 2.85 S/m. The bottom part of Figure 3 shows the histogram of the estimates compared with the Gaussian distribution corresponding to the calculated mean value and the standard deviation. The ratio of the standard deviation to the mean value for the conductivity is about 0.5. For the diagonal contributions to the MSD_{coll} (and therefore for the diffusion coefficients of ions), this ratio is much smaller—about 0.1. The large width of the distribution of conductivity values results from the off-diagonal contributions to the collective MSDs, for which the mean value is close to 0, and therefore, the stdev/avg ratio is between 5 and 7.

Given the uncertainty of the calculated values, it is evident that the use of a single 100 ns trajectory to estimate the conductivity carries a significant risk of obtaining the value that significantly differs from the average. Therefore, for such a trajectory length, averaging over different systems is necessary. We analyzed closer how the average and the standard deviation vary with increasing number of samples N. For this purpose, we used different time windows to estimate the conductivity from the slope of the MSD_{coll}, ranging from 10−20 to 10−100 ns. There was no particular ordering of the systems; they were added into the set randomly, in the order in which the electrolyte structures had been generated. Results are collected in Figure 4.

The mean value of the conductivity σ_{avg} stabilizes when the set used for averaging contains 30−40 samples. For smaller numbers of samples, adding a new result may change the average quite significantly. As visible in the bottom panel of
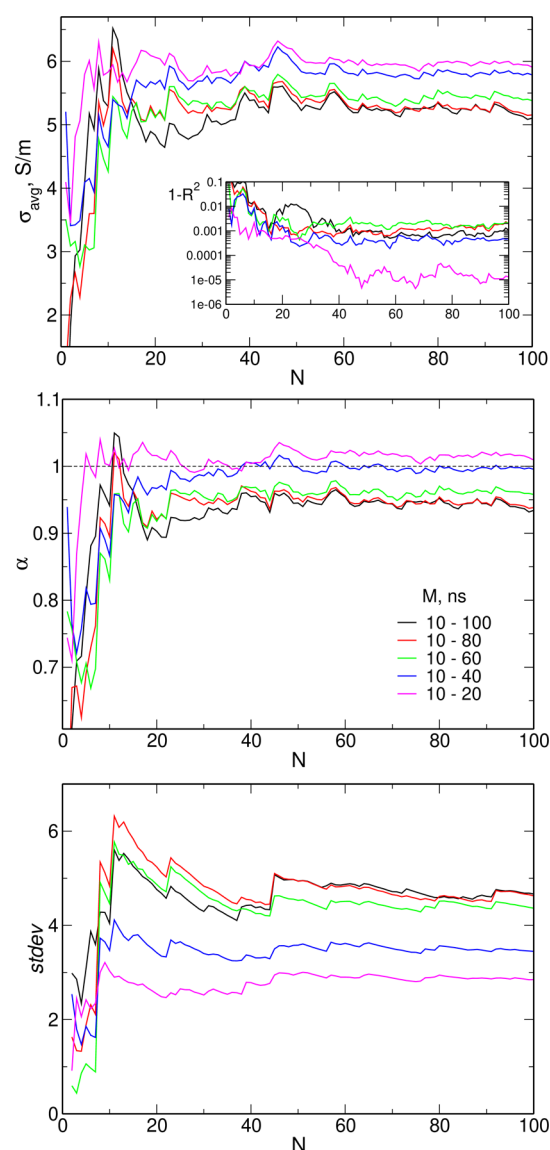

Figure 4. Mean values of the conductivity estimated using time windows M from 100 ns trajectories averaged over N datasets (upper panel); averaged exponent in the $MSD_{coll} \sim t^{\alpha}$ dependence (middle panel); and standard deviation of estimated conductivity (bottom panel).

Figure 4, standard deviation of the conductivity calculated for the dataset increases up to $N = 10$ and then decreases slowly and for $N > 40$ becomes approximately constant, although significant stepwise increase in standard deviation still may be observed when the system far from the average is added to the set. Likewise, the value of the correlation coefficient R^2 does not significantly improve further for more than 40 samples in the dataset. The final mean value of the conductivity obtained for the whole set of 100 systems depends on the time window used to estimate the slope of the MSD_{coll}—wider intervals result in lower conductivities. Values of σ_{avg} and their standard deviations σ_{stdev} obtained for different choices of time window are listed in Table 1. In addition to the data plotted in Figure 4,

Table 1. Mean Values σ_{avg}, Standard Deviations σ_{stdev}, Exponents α, and Corrected Values σ_{corr} of Conductivity Estimated from 100 Short (100 ns) Trajectories Using Different Time Windows M

M, ns	σ_{avg}, S/m	σ_{stdev}, S/m	$\sigma_{stdev}/\sigma_{avg}$	α	σ_{corr}, S/m
10−100	5.104	4.664	0.91	0.934	5.823
10−80	5.156	4.627	0.90	0.938	5.821
10−60	5.380	4.365	0.81	0.958	5.811
10−40	5.786	3.448	0.60	0.995	5.864
10−20	5.909	2.849	0.48	1.009	5.848
40−100	4.954	6.085	1.23	0.908	5.561
40−80	4.835	5.931	1.23	0.881	5.563
40−60	4.858	6.142	1.26	0.868	5.480

we also include in Table 1 the results obtained using the time windows between 40 and 100 ns, yielding lower conductivities (which is a result of the lower slope of MSD_{coll} above 40 ns, cf. Figure 2). As seen in Figure 1, the divergences between MSD curves for different systems and their nonlinearity become more prominent for longer times; accordingly, standard deviations presented in Table 1 increase with the width of the time window and, generally, for windows covering the long-time part of the data. We should also note that because of large standard deviations, at the 95% confidence level one cannot reject the hypothesis that the mean values of the conductivity obtained for different time ranges shown in Table 1 do not differ. Therefore, the data in Table 1 suggest that the conductivity in the system is between 4.8 and 5.9 S/m.

At this point, we would like to comment on the minimal length of the trajectories necessary to obtain reliable results. In order to get some insight into this issue, we repeated the analysis of the 100 ns trajectories but using only the limited part from the beginning of the trajectory to perform the time averaging. The length L of the analyzed part was $L = 20, 40, 60,$ and 80 ns; in all cases, the time window 10−20 ns was used to estimate the conductivity. The results are collected in Figure 5. The upper panel shows the averaged MSD_{coll} and two curves corresponding to the two individual trajectories which are the most deviating from the average. Apparently, for larger L values, the average and the individual curves become more linear. Simultaneously, the extreme individual MSD_{coll} curves become closer to the average, indicating that the width of the distribution of estimated conductivities becomes narrower. This is confirmed by the ratio of the standard deviation of the conductivity to its mean value $\sigma_{stdev}/\sigma_{avg}$ and the R^2 coefficient shown in the bottom panel of Figure 5. The $\sigma_{stdev}/\sigma_{avg}$ ratio decreases with L and for $L \geq 60$, it is lower than 1. From σ_{stdev} and R^2 values, we may conclude that the length $L = 80$ ns of
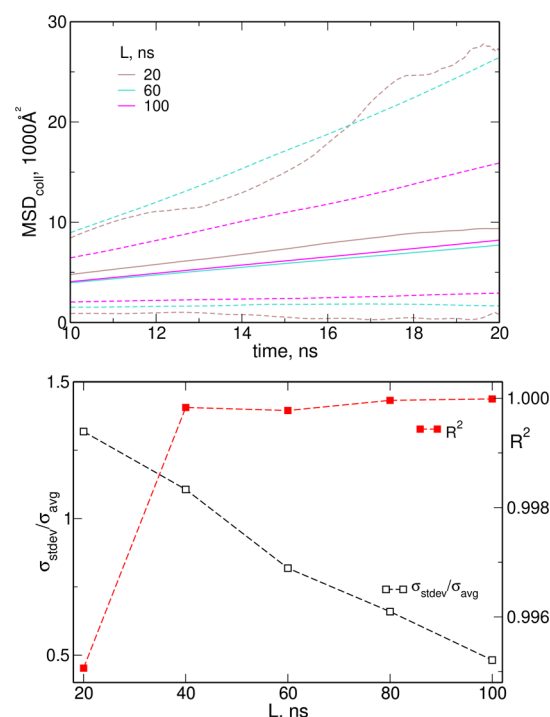

Figure 5. Averaged MSD_{coll} curves (solid lines) and individual trajectories most differing from the average (broken lines) obtained for different lengths L of the trajectory used in the analysis (top) and dependence of the $\sigma_{stdev}/\sigma_{avg}$ ratio and the correlation coefficient on L (bottom).

the trajectory could be sufficient to estimate the conductivity. Here, this conclusion was obtained from a posteriori analysis, but in practice, such an analysis can be performed during MD computations to decide whether the collected trajectories are sufficient for estimates of the conductivity or it is necessary to continue the simulations.

Seeking a way to narrow the 4.8−5.9 S/m interval containing the estimated values of the conductivity, we turned our attention to the dependence on the MSD_{coll} versus time. Generally, one may write it as $MSD \sim t^{\alpha}$. In calculations of the conductivity, we assume that the dependence is linear, that is, $\alpha = 1$ and the conductivity is proportional to the slope of the MSD. This assumption may be verified by fitting the α parameter in the dependence $\log(MSD) \sim \alpha \log t$. The middle panel of Figure 4 shows the values of the exponent α obtained for the increasing number of samples N in the dataset and for different time windows used to estimate the parameters. The values calculated for the whole set of 100 systems are listed in Table 1. Like the mean value and the standard deviation, the fitted value of the exponent α becomes approximately constant for $N > 50$. Only for one of the choices of the time interval ($M = 10−40$ ns), α is close to 1. It is slightly larger than 1 for $M = 10−20$ ns and lower than 1 for broader time ranges. The mean value of the conductivity σ_{avg} positively correlates with the value of the exponent α. This is not surprising because for $\alpha < 1$ and $\alpha > 1$, the MSD curve bends downward or upward, respectively; therefore, the linear fit to the data accordingly yields a slope lower or higher than that in the perfectly linear case $\alpha = 1$.

The deviations of the exponent α from 1 may indicate a special behavior of the system but may as well result from the uncertainty of data. In fact, regardless of the time window, the

diagonal contributions of cations and anions to the MSD_{coll} are described by the dependences with α in the range between 0.99 and 1.01. Therefore, the electrolyte exhibits normal diffusion of Na^+ and Cl^- ions, with $MSD \sim t$. As mentioned earlier, the deviations of the MSD_{coll} from linearity are caused by off-diagonal contributions, which have much larger uncertainties than those of the diagonal components. Let us suppose that the deviations from unity of the exponent α in the dependence of the MSD_{coll} versus time result from the noise in the data. In such a case, the most reliable estimates of the conductivity come from these datasets/time windows, which yield the values of α most close to 1. In Figure 4, this would correspond to the $M = 10-40$ ns time range. We can inspect the data closer, checking the correlation between the value of α and the corresponding conductivity σ_{avg}. The data are displayed in Figure 6.
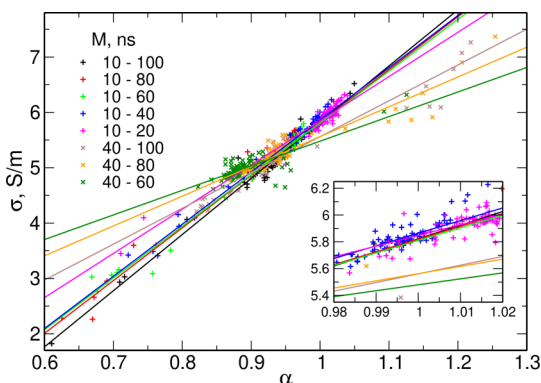

Figure 6. Correlation between the exponent α and the conductivity σ estimated using different time windows M. Lines are linear fits to the data.

Apparently, α and σ_{avg} are correlated and several sets of data (for different choices of M) fall onto one common curve. We used linear fits to the data in Figure 6 to find the conductivity σ_{corr} corresponding to $\alpha = 1$. Corrected values of the conductivity σ_{corr} are listed in Table 1. Both from the numerical values in Table 1 and from the enlarged region of the plot close to $\alpha = 1$ shown in the inset of Figure 6, it is clear that the differences between values corresponding to different time windows are greatly reduced. All five time windows starting at 10 ns yield conductivity within the range between 5.81 and 5.86 S/m. The three other sets with M starting at 40 ns give lower values, from 5.48 to 5.56 S/m, but the difference between these two groups of results is much smaller than that in "uncorrected" data. The data obtained for time windows including long times are less-reliable (because averaging over time intervals is worse in this region), but even including the results for $M = 40-60$, 80, and 100, we find the conductivity in the interval 5.5–5.9 S/m, significantly narrower than that of the previously determined 4.8–5.9 S/m.

The difficulties in computations of a reliable estimate of collective MSDs of ions (and therefore the estimate of the conductivity) together with the fact that the average MSDs of ions (proportional to the self-diffusion coefficients) can be obtained much easier are the origin of a quite common approach in which the conductivity is estimated from the diffusion coefficients of cations and anions. This approximation is reasonable in the systems in which the correlations between motions of ions are expected to be small. A simple check of

whether this assumption is valid is to compare the average of the diffusion coefficients of cations and anions $D_{av} = (D_+ + D_-)/2$ with the collective diffusion coefficient D_{coll}. Following this reasoning, one may calculate the ratio of the uncorrelated to correlated motion of ions $r_{u/c} = D_{av}/D_{coll}$ and use it to rescale the conductivity obtained from the average diffusion coefficient D_{av}. The latter can be usually estimated with satisfactory accuracy even from relatively short trajectories. On the other hand, to estimate the $r_{u/c}$ value, only the time window at short times can be used because for longer times, the MSD_{coll} will be too noisy. Such an approach, used, for example, in refs,[17−19] relies on the assumption that the $r_{u/c}$ value does not change significantly between short and long times. Nevertheless, it may provide an alternative way to estimate the conductivity from the length of the trajectory, which is sufficient only to calculate the diffusion coefficients but shorter than what would be normally necessary to obtain the conductivities directly from the MSD_{coll} curves.

In Figure 7, we show how the $r_{u/c}$ ratio depends on time between 0 and 100 ns for short trajectories. The value averaged
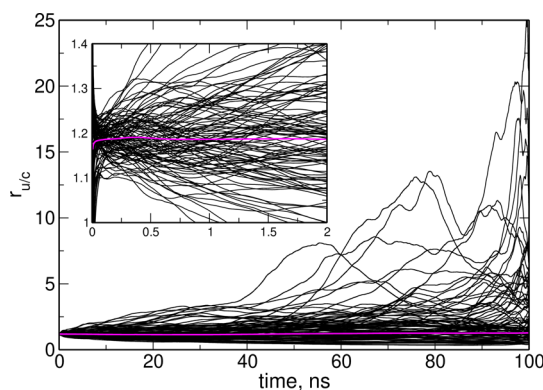

Figure 7. Ratio of uncorrelated to correlated conductivity obtained from 100 ns trajectories. The color line is the average over 100 systems.

over all 100 trajectories is about 1.2, and indeed, it is reasonably constant in time, taking values between 1.15 and 1.25, but the data for individual trajectories differ substantially. Only in a narrow time interval of 0.1–0.2 ns, the $r_{u/c}$ values for all systems are close to the average (between 1.1 and 1.25) but they diverge fast above that time. Already at 1 ns, there are some systems with a $r_{u/c}$ ratio lower than 1, suggesting the opposite effect of correlations (enhancing the conductivity instead of suppressing it). For some trajectories, $r_{u/c}$ reaches values much larger than 1, which would result in significantly underestimated conductivity if used to rescale the conductivity estimated from the D_{av}. The approach discussed here, although generally justified as seen from the average $r_{u/c}$, should be therefore adopted with care, possibly using the average over several systems. It should be avoided when the estimated $r_{u/c}$ ratio significantly exceeds 1, indicating the important role of correlations. In such a case, correlations do not constitute small correction to the conductivity, which therefore should be rather calculated without approximations.

3.2. Long Trajectories. In the preceding part of the paper, we analyzed the results of relatively short 100 ns MD trajectories. To improve the averaging, we used as many as 100 independent simulations. Now, we want to check the performance of the alternative way of investing of the

computational time: using only few long trajectories instead of many short simulations. For this purpose, we analyze the three long (2 μs) trajectories. With this length of the trajectory, averaging of the MSD over different choices of Δt will be about 20 times better than that for individual short trajectories. We tried also the alternative possibility of averaging: the production part of the long trajectory was split into 19 parts, each 100 ns long, and the analysis was performed for these segments in the same way as that applied to short trajectories in the preceding section.

Resulting MSD_{coll} dependences on time are presented in Figure 8. Data time-averaged over one long trajectory (solid
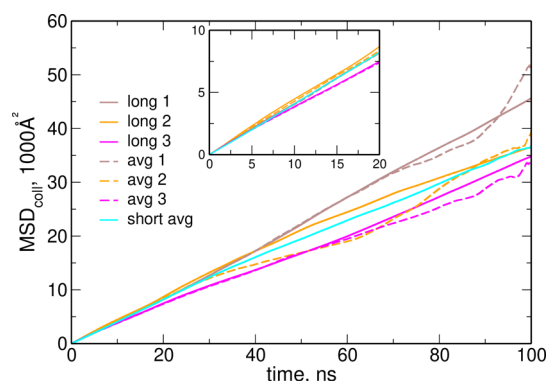

Figure 8. Collective MSD obtained for 1.9 μs trajectories. Data are averaged over the whole trajectory ("long #") or over its 100 ns parts ("avg #"); "short avg" is the average over the 100 individual 100 ns trajectories.

lines) are marked as "long #"; the averages over 100 ns parts of the trajectory (broken lines) are marked as "avg #". For comparison, the average over 100 of short 100 ns trajectories from Figure 1a is also shown. Within the time range shown in Figure 8 (up to 100 ns), the collective MSDs obtained from the whole trajectory are more linear than the MSD_{coll} averaged over parts of the trajectory. Nevertheless, the curves corresponding to the same system lie close together. In the time window up to 20 ns (inset of Figure 8), the differences between the two ways of data averaging are minimal. This is confirmed by the numerical data collected in Table 2,

Table 2. Estimates of the Conductivity Obtained Using the Time Window $M = 10-20$ ns from Long (1.9 μs) Trajectories Treated as a Whole (σ_l) or Split Into 100 ns Parts (σ_{avg} with the Corresponding Standard Deviation σ_{stdev})

trajectory #	σ_l, S/m	σ_{avg}, S/m	σ_{stdev}, S/m	$\sigma_{stdev}/\sigma_{avg}$
1	5.880	5.985	2.198	0.37
2	5.936	5.748	2.671	0.46
3	5.180	5.123	1.935	0.38

presenting the conductivities estimated in the time range $M = 10-20$ ns from the whole long trajectory (σ_l) or as the average over its 100 ns intervals (σ_{avg}). In the latter case, the corresponding standard deviations σ_{stdev} are also given. As seen from the data, the two ways of data averaging yield similar estimates of conductivity, but there are differences (up to 0.8–0.9 S/m) between different systems. Standard deviations are similar to those obtained earlier for short trajectories, in agreement with the observation that for about 20 systems, the

standard deviation becomes close to the value calculated from a large set of data. Both ways of averaging can provide, therefore, a reasonable estimate for the given system, but as there are differences between systems, a single trajectory cannot provide information on whether the estimate represents the typical situation (e.g., the third trajectory apparently yields values lower than the two other). The differences between trajectories reflect different starting configurations; however, for sufficiently long simulations, they become smaller because of better sampling of the configurational space. In particular, this supposition seems to be justified in the approach based on splitting the trajectory into parts because for long times, the segments separated by large time intervals may be viewed as separate simulations initiated from different structures. Nevertheless, the use of multiple relatively short simulations will be more efficient in the averaging.

We also analyzed the data collected from long trajectories in the same way as was done for short trajectories, treating all 100 ns parts as one set of data. The results obtained for different time windows are shown in Figure 9. The picture is similar to that presented in Figures 4 and 6 for short trajectories. The mean values depend on the time window used to extract the slope of the MSD_{coll} and vary between 5.05 and 5.6 S/m. The σ_{avg} correlates with the exponent α, and the linear fits to the dependence σ_{avg} versus α are shown in the bottom panel of Figure 9. Like for the short trajectories, the corrected values σ_{corr} corresponding to $\alpha = 1$ calculated from the fit parameters do not depend much on the time window and fall in the interval from 5.70 to 5.87 S/m. This result agrees well with the estimates obtained from short MD trajectories. Also, the time dependence of the $r_{u/c}$ ratio calculated from long trajectories exhibits features similar to those shown in Figure 7, with the average $r_{u/c}$ close to 1.2.

Finally, we will comment on the trajectories obtained for "big" systems, which were analyzed in the way applied to long trajectories for small systems, that is, treating individual 100 ns parts as separate trajectories. Statistics of averaged data shows that the standard deviation of the conductivity is practically the same for "long" and "big" datasets, that is, there is no improvement for larger system sizes. Inspection of different contributions to the MSD_{coll} and the conductivity reveals that the standard deviations of the diagonal contributions of cations and anions (related to self-diffusion coefficients) for "big" systems are two times smaller than those for the "long" set. The unaffected standard deviation of the conductivity results from the standard deviations of the off-diagonal contributions, which are of similar sizes for both long and big trajectories. As shown earlier, the errors of conductivity estimates are governed by the off-diagonal contributions. Therefore, although increasing the size of the simulated systems improves the accuracy of estimated self-diffusion coefficients, it is ineffective for collective properties, such as the conductivity.

4. DISCUSSION

We presented some results of conductivity estimates for a model system using MD trajectories of different lengths and for different system sizes. Here, we will try to summarize our findings and to formulate some suggestions.

In order to improve the accuracy of properties estimated from the MD simulations, one may invest the computational effort in increasing the system size, the length of the MD trajectory, or the number of realizations of the system. The first possibility, enlargement of the system, may be beneficial for
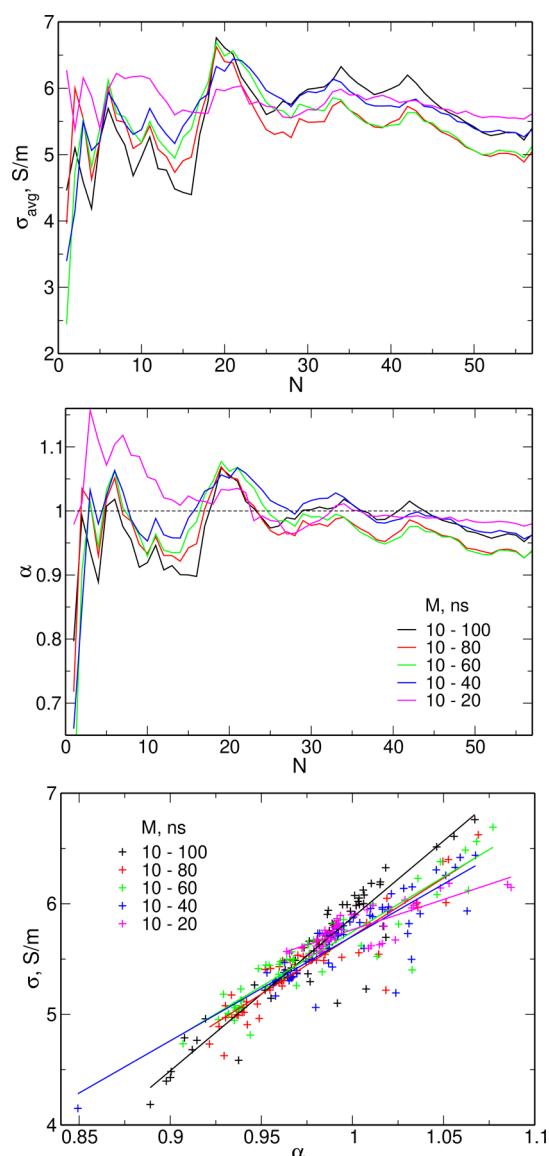
Figure 9. Mean values of the conductivity estimated using time windows M from 100 ns parts of 1.9 μs trajectories averaged over N datasets (upper panel); averaged exponent in the $MSD_{coll} \sim t^{\alpha}$ dependence (middle panel); and correlation between the exponent α and the estimated conductivity (bottom panel).

calculations of diffusion coefficients but will not efficiently improve estimates of the conductivity. Of course, there are other factors which should be considered for the choice of system size, for example, possible finite-size (boundary) effects,[20−23] but the discussion on this issue is beyond the scope of our work.

Once the system size is established, given a limited amount of resources, one faces the choice between the number of investigated samples versus the length of simulations— between running many short simulations or few only but relatively long trajectories. The advantages of increasing the number of realizations are obvious: the improved statistics making it possible to estimate the deviations between samples. With this approach, it is also possible to eliminate the outliers caused, for example, by the nonphysical starting configurations. From the data obtained for our model systems, it seems that at least 10 trajectories should be used in order to get reliable

estimates. However, above a certain number of simulations, adding more trajectories to the dataset does not decrease the standard deviation of the conductivity; therefore, at this point, it seems reasonable not to increase further the number of realizations. The optimal number of independent trajectories seems to be about 30−40 systems, similar to that found in the study on estimating the shear viscosity.[6]

Assuming that the deviations of the MSD_{coll} from the linearity result from insufficient averaging and they are not the intrinsic property of the system (like a subdiffusive regime), one may try to improve the estimate of the conductivity using the procedure outlined in Section 3.1. By establishing the correlation between σ_{avg} and the exponent α, the corrected value of the conductivity may be found from the fit to the data as the value corresponding to $\alpha = 1$. As a proof of concept, we applied this procedure to the conductivity estimated in our recent paper on Me-TFSI electrolytes (Me = Li or Na) in EMIM-TFSI ionic liquid.[24] Results are displayed in Figure 10.
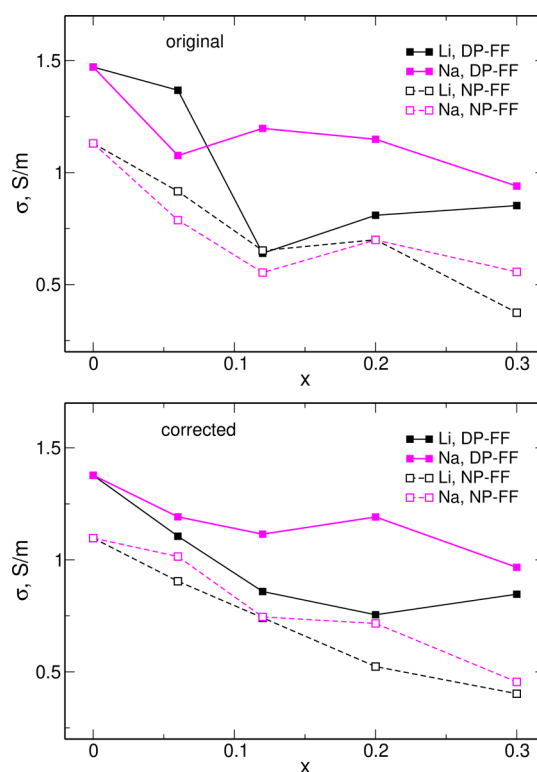

Figure 10. Results of the correction applied to the conductivities of the Me$_x$EMIM$_{(1−x)}$TFSI electrolytes estimated in ref 24.

The upper panel shows the original data, averaged over 10 parts (100 ns each) of a 1.1 μs trajectory. The bottom panel contains the data after the correction. The corrected plot is less-noisy and the trends are more easily noticeable: decrease in the conductivity with increasing Me$^+$ concentration, increase in conductivities for Na$^+$ electrolytes compared to Li$^+$-based systems, or larger conductivity calculated in the polarizable (DP) force field.

The disadvantage of using many short trajectories is the computational time lost for the equilibration of multiple systems. This time is greatly reduced when only few systems are simulated. However, the loss of the computational time used for the equilibration stage is partially compensated by better parallelization. Given the number of available CPU

cores, one may allocate all of them to one MD simulation (in practice probably to few simulations, with a large number of cores allocated to each run) or to divide the CPUs to small groups used to compute simultaneously many trajectories. Because the scaling of the simulation speed with the number of cores is typically worse than linear, the other setup is more efficient in terms of ns of MD trajectories produced in total per day.

If for some reason long trajectories are simulated, they give the opportunity to improve the statistics by averaging over more time intervals within the trajectory. This improves the linearity of the MSD_{coll} (as seen from Figure 8 compared to the top panel of Figure 1). Nevertheless, using only one set of data, even with perfectly linear MSDs, one cannot be sure whether the system is a representative for the population of possible systems, which may be constructed and simulated, or if it is rather an outlier. Splitting the long trajectory into shorter parts and averaging the analyzed data will not answer the abovementioned question but at least will provide some information on the deviations of obtained estimates. It will also make possible the correction of the data using the correlation σ_{avg} versus α. In any case, simulation of only one trajectory per system should be avoided.

5. CONCLUSIONS

We have investigated the issue of calculations of the electrical conductivity from the MD simulations. The results of our study show that the errors of the estimated conductivity result from the off-diagonal contributions to the collective MSD of ions and therefore, as a collective property, cannot be efficiently reduced by increasing the size of simulated systems.

The most effective way to reduce the fluctuations of MSD_{coll} is to average the data over many independent simulations. The minimal number of trajectories is about 10, and the optimum seems to be between 30 and 40 realizations. Adding more than 40 simulations to the dataset is unlikely to further reduce errors because at this number of trajectories, the calculated standard deviation reaches the value roughly corresponding to the population of systems, and therefore, it will not decrease for more samples. If long trajectories are available, it is advisable to split them into shorter parts and to perform the analysis of the averages obtained from the set of segments treated as separate individual trajectories.

We have proposed an ad hoc correction to the estimated conductivity, based on the assumption that the deviations of the system from an ideal behavior are caused by the fluctuations of the off-diagonal contributions to the conductivity and not by the deviations from normal diffusion. The correction is based on fitting correlations between the conductivity and the exponent of the dependence of the MSD_{coll} versus time and finding the estimate corresponding to the linear increase in MSD_{coll}.

We have also shown that the approximate estimates based on the ratio of uncorrelated and correlated displacements may be an effective way to reduce the computational effort using shorter trajectories but with careful choice of a short time window at the beginning of the trajectory in which the $r_{u/c}$ values least differ between systems. For this purpose, comparing several systems is recommended.

AUTHOR INFORMATION

Corresponding Author

Andrzej Eilmes − *Faculty of Chemistry, Jagiellonian University, 30-387 Kraków, Poland;* ● orcid.org/0000-0002-4690-2611; Email: eilmes@chemia.uj.edu.pl

Author

Piotr Kubisiak − *Faculty of Chemistry, Jagiellonian University, 30-387 Kraków, Poland;* ● orcid.org/0000-0002-2680-2461

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.0c07704

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Allen, M. P. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.

(2) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Application*; Academic Press: San Diego, 2002.

(3) Vestergaard, C. L.; Blainey, P. C.; Flyvbjerg, H. Optimal Estimation of Diffusion Coefficients from Single-Particle Trajectories. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2014**, *89*, 022726.

(4) Pranami, G.; Lamm, M. H. Estimating Error in Diffusion Coefficients Derived from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 4586−4592.

(5) Bullerjahn, J. T.; von Bülow, S.; Hummer, G. Optimal Estimates of Self-Diffusion Coefficients from Molecular Dynamics Simulations. *J. Chem. Phys.* **2020**, *153*, 024116.

(6) Zhang, Y.; Otani, A.; Maginn, E. J. Reliable Viscosity Calculation from Equilibrium Molecular Dynamics Simulations: A Time Decomposition Method. *J. Chem. Theory Comput.* **2015**, *11*, 3537−3546.

(7) Bockris, J. O. M.; Reddy, A. K. N. *Modern Electrochemistry 1: Ionics*, 2nd ed.; Plenum Press: New York, 1998.

(8) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(9) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(10) Roux, B. Valence Selectivity of the Gramicidin Channel: a Molecular Dynamics Free Energy Perturbation Study. *Biophys. J.* **1996**, *71*, 3177−3185.

(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(12) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(13) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.

(14) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.

(15) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(16) Müller-Plathe, F. Permeation of Polymers − A Computational Approach. *Acta Polym.* **1994**, *45*, 259−293.

(17) Borodin, O.; Gorecki, W.; Smith, G. D.; Armand, M. Molecular Dynamics Simulation and Pulsed-Field Gradient NMR Studies of Bis(fluorosulfonyl)imide (FSI) and Bis[(trifluoromethyl)sulfonyl]-imide (TFSI)-Based Ionic Liquids. *J. Phys. Chem. B* **2010**, *114*, 6786−6798.

(18) Lesch, V.; Jeremias, S.; Moretti, A.; Passerini, S.; Heuer, A.; Borodin, O. A Combined Theoretical and Experimental Study of the Influence of Different Anion Ratios on Lithium Ion Dynamics in Ionic Liquids. *J. Phys. Chem. B* **2014**, *118*, 7367−7375.

(19) Li, Z.; Borodin, O.; Smith, G. D.; Bedrov, D. Effect of Organic Solvents on Li^+ Ion Solvation and Transport in Ionic Liquid Electrolytes: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2015**, *119*, 3085−3096.

(20) Gabl, S.; Schröder, C.; Steinhauser, O. Computational Studies of Ionic Liquids: Size Does Matter and Time Too. *J. Chem. Phys.* **2012**, *137*, 094501.

(21) Moultos, O. A.; Zhang, Y.; Tsimpanogiannis, I. N.; Economou, I. G.; Maginn, E. J. System-Size Corrections for Self-Diffusion Coefficients Calculated from Molecular Dynamics Simulations: The Case of CO_2, n-alkanes, and Poly(ethylene glycol) Dimethyl Ethers. *J. Chem. Phys.* **2016**, *145*, 074109.

(22) Jamali, S. H.; Wolff, L.; Becker, T. M.; Bardow, A.; Vlugt, T. J. H.; Moultos, O. A. Finite-Size Effects of Binary Mutual Diffusion Coefficients from Molecular Dynamics. *J. Chem. Theory Comput.* **2018**, *14*, 2667−2677.

(23) Shao, Y.; Shigenobu, K.; Watanabe, M.; Zhang, C. Role of Viscosity in Deviations from the Nernst−Einstein Relation. *J. Phys. Chem. B* **2020**, *124*, 4774−4780.

(24) Kubisiak, P.; Wróbel, P.; Eilmes, A. Molecular Dynamics Investigation of Correlations in Ion Transport in MeTFSI/EMIM-TFSI (Me= Li, Na) Electrolytes. *J. Phys. Chem. B* **2020**, *124*, 413−421.