Improving the efficiency of Monte Carlo simulations of ions using expanded grand canonical ensembles

Harold W. Hatch,[1, a)] Steven W. Hall,[2] Jeffrey R. Errington,[3] and Vincent K. Shen[1]

[1])*Chemical Informatics Research Group, Chemical Sciences Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8380, USA*

[2])*Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC 29634, USA*

[3])*Department of Chemical and Biological Engineering, University at Buffalo, Buffalo, NY 14260, USA*

While ionic liquids have promising applications as industrial solvents, predicting their fluid phase properties and coexistence remains a challenge. Grand canonical Monte Carlo simulation is an effective method for such predictions, but equilibration is hampered by the apparent requirement to insert and delete neutral sets of ions simultaneously in order to maintain charge neutrality. For relatively high densities and low temperatures, previously developed methods have been shown to be essential in improving equilibration by gradual insertion and deletion of these neutral sets of ions. We introduce an expanded ensemble approach which may be used in conjunction with these existing methods to further improve efficiency. Individual ions are inserted or deleted in one Monte Carlo trial rather than simultaneous insertion/deletion of neutral sets. We show how charge neutrality is maintained and show rigorous quantitative agreement between the conventional and the proposed expanded ensemble approaches, but with up to an order of magnitude increase in efficiency at high densities. The expanded ensemble approach is also more straightforward to implement than simultaneous insertion/deletion of neutral sets, and its implementation is demonstrated within open source software.

I. INTRODUCTION

Ionic liquids have a wide range of potential industrial applications in synthesis, coatings, lubricants, batteries, fuel cells, personal care and pharmaceuticals.[1] By replacing volatile organic compounds traditionally used as industrial solvents, nonvolatile ionic liquids can help reduce environmental pollution and improve sustainability.[2] Because they are costly to synthesize and their properties are difficult to measure (e.g., low concentrations, hygroscopic, etc), it is desirable to screen novel ionic liquids for desired solvent properties and phase behavior using molecular simulations. However, conventional simulation methods often encounter sampling difficulties at low temperatures and high densities.[3] Thus there is a need to improve the efficiency of simulations of ionic liquids to expand their applicability to industrially relevant solvents.

While the calculation of vapor-liquid equilibrium properties or solvation free energies in the canonical ensemble often requires an explicit interface, which can lead to large simulations to avoid system size effects, grand canonical[4,5] and Gibbs ensemble[6,7] Monte Carlo simulations avoid this computationally costly requirement.[8] In both ensembles, molecules must be inserted and deleted from the simulation box(es). While this is generally not an issue with small molecules, the trial molecule insertion/deletion becomes prohibitively expensive for complex molecules and at higher densities and lower temperatures. Ionic systems in particular suffer from poor

a)Electronic mail: harold.hatch@nist.gov

configurational sampling when multiple ions are inserted simultaneously to maintain charge neutrality in periodic boundary conditions. Configurational-bias[9,10] and growth-expanded ensembles[11–14] are ways of alleviating this sampling issue. Continuous fractional component Monte Carlo[15] is also a very promising technique to improve sampling in these cases. The proposed method in this work utilizes flat histogram Monte Carlo and expanded ensembles to greatly improve sampling by removing the requirement of inserting/deleting multiple ions simultaneously to satisfy charge neutrality, and can be used in conjunction with other methods.

While the insertion/deletion of an ion requires a balancing counter charge in order to maintain charge neutrality in periodic boundary conditions, this counter charge does not have to be localized in space (e.g., a point charge). For example, the Ewald summation,[16] or other efficient Fourier-space computations,[17] is often used to compute the long range electrostatic interactions in molecular simulations. When the Fourier-space summation is applied to a collection of localized charges that do not sum to zero, it will neutralize the system with a delocalized cloud of counter charge. Thus, charge neutrality is always maintained with these Fourier-space summations. Several studies have utilized this aspect of the Fourier-space summation. They cover a range of topics, such as the calculation of chemical potentials and ionization free energies,[18–20] confined fluids,[21] and the one-component plasma.[22] However, to correct for the delocalized counter charge, special terms[23] may be required for the pressure, energy and chemical potential. Unfortunately, the application of these corrections may become problematic for grand canonical or Gibbs ensemble sim-

ulations when the ensemble averages contain a variety of different delocalized counter charges.

The use of flat histogram Monte Carlo with expanded ensembles[24] conveniently overcomes this issue of delocalized counter charge corrections. Flat histogram Monte Carlo methods, such as Wang-Landau[25] and Transition-Matrix,[26,27] use a self-consistently generated biasing function to enhance sampling. When the order parameter is the number of ions in a grand canonical ensemble simulation, the macrostate probability distribution obtained from a flat histogram method can be used to determine the free energy as a function of density.[27,28] Histogram reweighting[29] may then be applied to compute thermodynamic and structural properties over a range of conditions. In addition, expanded ensembles may be applied to provide intermediate states to increase the acceptance probability of insertions/deletions. For example, growth expanded ensemble[3,11,30] simulations introduce states with partially grown molecules because they require smaller cavities to accommodate them. This leads to increased acceptance probabilities in dense systems. Because physical systems do not contain partial molecules, only the macrostates with complete molecules are used to compute ensemble-averaged properties. In an analogous way, the expanded grand canonical ensembles approach proposed in this work utilizes intermediate states with a neutralizing cloud to provide a pathway between neutral states and increase acceptance probabilities in dense liquids.

In order to rigorously test the correctness and benchmark the efficiency of the proposed method, we simulate the restricted primitive model (RPM) due to its simplicity and its wealth of previously published data for comparison. The RPM has been very well characterized since the original study of its vapor phase.[31] The vapor-liquid phase diagram has been investigated extensively with Monte Carlo,[32–38] in addition to solid phase equilibria.[39] The RPM was also used as a test case for hyper-parallel tempering.[40] System size effects and boundary conditions have also been well studied,[3,41] and we expect system size effects reported in Ref. 3 to be similar for the method presented in this manuscript. The focus of this manuscript is upon the saturated liquid and vapor properties and the focus is not to estimate the critical point, which has already been investigated extensively.[32,36,37,42–44] While we focus on a monatomic fluid in this work, the proposed approach can also be applied to more complex and atomistic ionic liquid models[45–47] in the future.

We propose an expanded ensemble approach with individual insertion/deletion of ions. The macrostates with net local charge and delocalized counter charge provide a thermodynamic bridge between the neutral states to enhance sampling. After the simulation is complete, the thermodynamics of the neutral states may then be recovered by simply considering only those neutral states in the subsequent analysis, in analogy to growth expanded ensembles. We show that the expanded grand canon-
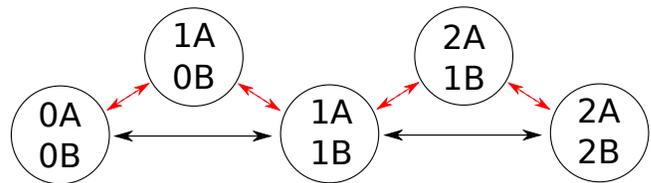
FIG. 1: Schematic of macrostates of a binary mixture of A and B ions shown by circles, and two different paths between these macrostates shown by the two differently colored and styled arrows. The black, solid arrows show the conventional simultaneous insertion and deletion of A-B pairs. The red, dashed arrows show alternating insertion of A and B ions, starting with A, for the expanded ensemble approach.

ical ensembles approach yields results that are in excellent agreement with conventional simulations of the RPM fluid. We also show that the expanded ensemble approach is up to an order of magnitude more efficient than simultaneous insertion of multiple ions. Notably, the implementation of this method does not require special Ewald correction terms. In addition, the expanded ensemble approach presented here can be used in conjunction with previously developed configurational-bias and growth expanded ensemble methods.

The expanded grand canonical ensembles method is described in more detail in Section II. The restricted primitive model electrolyte (RPM) is described in Section III. Details for the flat histogram Monte Carlo simulations are reported in Section IV. In Section V, the proposed method is shown to rigorously reproduce the RPM saturation properties when compared to conventional algorithms and previously published data. Then a discussion of the efficiency and ease of implementation of the proposed method follow in Section VI B, with up to an order of magnitude improvement observed at high density. Finally, we end with a discussion of future work in Section VII.

II. THE EXPANDED ENSEMBLE METHOD FOR IONS.

A schematic of the expanded ensemble approach proposed in this work is shown in Figure 1 for a binary system of equal and oppositely charged ions, which are labeled A and B. For such a binary system, the states with equal numbers of A and B are neutral. Thus, the conventional approach of inserting/deleting simultaneously A-B pairs is shown by the black solid line. For the alternative path proposed in this work (red arrows), the neutral states are bridged by states with one more A than B. We will show that sampling along this alternative path yields identical results for the neutral or even-numbered states.

Thermodynamic or structural quantities may be computed in the same manner as the conventional approach and with the help of histogram reweighting. Also note

that alternative paths may be considered, including growth of the individual ions. A major point for concern is whether or not the states with unequal number of A and B ions, and non-neutral localized system charge, are thermodynamically well-defined such that free energy changes between these states and the neutral states may be computed. We will demonstrate that this is not an issue for long range electrostatic interactions when using the Ewald summation for the restricted primitive model electrolyte (RPM).

III. RESTRICTED PRIMITIVE MODEL ELECTROLYTE

The restricted primitive model electrolyte (RPM)[31] serves as an excellent test case of the proposed expanded ensemble method due to both its simplicity and also its very strong electrostatic interactions at low temperature. The RPM contains a hard sphere and central point charge via the following potential energy, u_{ij} between ions i and j,

$$u_{ij}(r_{ij}) = \begin{cases} \infty & r_{ij} < \sigma_{ij} \\ \frac{z_i z_j e^2}{4\pi\epsilon\epsilon_0 r_{ij}} & r_{ij} \geq \sigma_{ij}, \end{cases} \quad (1)$$

where r_{ij} is the distance between centers of i and j, σ_{ij} is the hard sphere diameter, ze is the charge, e is the charge of an electron, ϵ is the dielectric constant of the implicit solvent and ϵ_0 is the dielectric permeability of a vacuum. In this work, we focus on $z = \pm 1$ but the method is not restricted to this case. Energies are normalized by $Q^* = e^2/(4\pi\epsilon\epsilon_0\sigma)$. A cubic box of length $L = 12\sigma$ is used for all simulations. We chose to compare our resulting equilibrium densities and saturation pressures with those obtained by Rane and Errington,[3] and thus use a real space cutoff distance of $r_c = 4.8913043\sigma$.

Due to the long range nature of this potential, the Ewald summation[16] is utilized with conducting boundary conditions.[3,8] The screening parameter, $\alpha = 6.870984/L$, where L is the cubic periodic box length. The reciprocal-space integer cutoff k_{max}^2 is 38. For systems with net charge, the Ewald summation includes a delocalized cloud of the opposite charge of the system.[8,16] Note that no special Ewald correction[21] is necessary, which helps make the proposed expanded ensemble approach simple to implement in existing codes. Including these special Ewald corrections is also acceptable, which may result in a change in the free energy of the states with odd numbers of ions (for $z = \pm 1$) but not the free energy of states with even numbers (i.e., equal number of A and B ions).

IV. FLAT HISTOGRAM MONTE CARLO SIMULATIONS

Transition matrix Monte Carlo (TMMC) simulations[26,28] were employed in the grand canonical ensemble[4] along with histogram reweighting techniques[29] to calculate the vapor-liquid phase equilibria of RPM. Simulations were conducted with the open source software called the Free Energy and Advanced Sampling Simulation Toolkit (FEASST)[48] using an experimental branch for pair insertions.

The TMMC simulations were parallelized by dividing the macrostate range (e.g., number of ions) into 12 windows. The size of the windows decreased for increasing number of ions, N, by an exponent of approximately 2. The macrostate distribution of the complete range was then recovered by equating the macrostates which overlapped between consecutive windows and renormalizing. The maximum number of ions, N_m, was chosen such that $\ln \Pi(N_l) - \ln \Pi(N_m) > 25$, where $\Pi(N_l)$ is the macrostate distribution and N_l is the location of the saturated liquid peak. Convergence was measured in sweeps. One sweep is completed when each macrostate is visited by another macrostate at least 100 times. In order to improve the comparison with pair insertion/deletion simulations, which only visit even-numbered macrostates, only macrostates with even numbers of ions were considered in the definition of the sweep even for simulations with odd-numbered states. The macrostate distribution was updated every 10^6 trials.

The following Monte Carlo trial moves were employed, as summarized in Table I. The traditional single-particle transitions were attempted. In addition, collective ion moves were also important in the RPM simulations[44,49] and were implemented via rigid cluster translations and rotations. Clusters were defined as all ions within a cutoff distance of 1.05σ from one or more other ions in the same cluster. Note that the sensitivity of the convergence time on this distance cut off parameter for clusters was not investigated in this work, but could be beneficial for further improvements in sampling. However, any improvements in efficiency by optimization of this parameter would be roughly the same for both the old and new methods and thus the relative efficiency would be similar. The clusters were obtained numerically by a flood-fill algorithm. To obey detailed balance, cluster moves which resulted in an ion joining a different cluster were rejected. For all trials moves mentioned above, the parameter associated with the maximum possible translation or rotation was optimized, via 5 % change every 10^6 trials, to yield approximately 25 % acceptance of the trial.

Grand canonical insertion/deletion (also referred to as a transfer) was attempted in three different ways. For the conventional simulations without expanded ensembles, neutral pairs of ions were inserted and deleted from the simulations. Because these conventional simulations were slower to converge than single ion insertion/deletion simulations at low temperature and high density, dual-cut configurational bias (DCCB)[50] pair insertions/deletions were required in order to make feasible comparisons with and without expanded ensembles. DCCB utilized a cell list to monitor ions within a distance σ relative to a central ion of interest to efficiently attempt 8 insertions or deletions for each hard sphere. The Rosenbluth weights

TABLE I: Relative weights for probability of selection of Monte Carlo trials for the three different simulations.

trial	PAIR	EE	EE-AVB
translation	1	1	1
cluster translation	$1/(4N_m)$	$1/(4N_m)$	$1/(4N_m)$
cluster rotation	$1/(4N_m)$	$1/(4N_m)$	$1/(4N_m)$
DCCB pair transfer	$1/4$	0	0
DCCB single transfer	0	$1/4$	$1/4$
AVB single transfer	0	0	1

TABLE II: Summary of simulation conditions. These are below the critical temperature of $T_c^* = 0.0489(3)$ estimated previously.[37]

T^*	$\ln z$	N_m
0.047899461	−13.94	650
0.045421902	−14.5	690
0.042944344	−16	690
0.038815080	−17	750
0.034685816	−19	820

were then used to select a single configuration of the hard sphere only attempts, and this configuration was accepted subject to the more expensive electrostatic interaction. To make precise comparisons, some simulations of the single ion insertions/deletions in the expanded ensemble also used DCCB. Additional simulations of the single ion insertion/deletion simulations also utilized aggregation volume bias (AVB) with 8 DCCB hard sphere insertions.[51–54] The aggregation volume was chosen for center separations in the range σ to 1.5σ. To compare the AVB used in this work with the previously used continuous distance biased method,[3,32,35,49] we note that over 99 % of the bias-selected distances would fall within this aggregation volume region for the temperature range considered in this work.

The three different simulation types described in Table I are as follows. PAIR refers to simulations with insertions and deletions in pairs only. EE refers to the expanded ensemble method proposed in this work with single ion insertion/deletions. EE-AVB is the same as EE but includes an additional AVB trial to further improve sampling. We chose to compare our resulting equilibrium densities and saturation pressures with those obtained by Rane and Errington.[3] Thus, we simulated the five temperatures reported in Table II. These reported temperatures might be rounded to four significant figures elsewhere in the manuscript. The activity, $z = \sigma^3 e^{-\beta\mu}/\Lambda^3$, where $\beta = 1/T^*$, μ is the chemical potential and Λ is the de Broglie wavelength. Each simulation window was run for as many as several hundred sweeps, but for at least 20 sweeps, with the following exception. The PAIR simulations were much slower to converge, and were terminated after two weeks on a dedicated 12-core node. This time limit afforded roughly 5 to 10 billion Monte Carlo trials and at least 5 sweeps but up to 163, depending on the density of the window. The pressure was calculated as described by Equation 20 of Ref. 27.

For statistical analysis, four independent simulations were performed at each condition. Reported error bars represent the standard deviation for these. The standard deviation from four samples is approximately equal to the 95 % confidence interval, because the standard deviation is divided by the square root of four to obtain the standard deviation of the average, which is then multiplied by 1.96 to obtain the 95 % confidence interval.

V. SATURATION PROPERTIES

In this section, we show rigorous quantitative agreement between the proposed expanded ensemble method, the conventional method, and simulation results from previous publications.[3] The macrostate probability distribution for the RPM at $T^* = 0.04294$ is shown in Figure 2a using the three different types of simulations and including the previous work. All four of these agree quantitatively within 95 % confidence.

Because the curves are visually indistinguishable, and the error bars are smaller than the symbols, we take a closer look at the differences by plotting the PAIR, EE and EE-AVB macrostate distributions relative to the previously reported reference simulation, as shown in Figure 2b. The standard deviations were obtained via the propagation of error formula $\sqrt{\sigma_i^2 + \sigma_{REF}^2}$ where i is either PAIR, EE or EE-AVB. In Figure 2b, we see that the error bars are smaller near the vapor and liquid peaks, and larger in the transition region and high density region. This is good because the saturation properties depend most sensitively on the vapor and liquid peaks of the macrostate probability distribution. Overall, the simulations are in agreement with the literature reference within 95 % confidence.

The expanded ensemble simulations also yield the macrostate probability distribution, $\Pi(N)$, for the states with odd numbers of ions, which are shown in Figure 2c. The odd-numbered macrostate probabilities exhibit the largest deviations from their proceeding even-numbered macrostates at the lowest densities. While the difference decreases as N increases, there is a slight trend for the deviation to increase again at very high densities. Also, the differences were observed to be negative over the entire range. Thus, the probability of the odd states is lower than that of the even-numbered states. This trend may be due to the lack of a localized counter charge, which makes the occurrence of lone ions much less favorable than pairs of oppositely charged ions, especially in the vapor phase. Note that this trend may change depending on the model, conditions and the expanded ensemble pathway. In the vapor phase, the differences in the probabilities are large enough to present sampling issues if a flat histogram or biasing method were not used. While the odd numbered macrostates provide a bridge to the even-numbered macrostates, they are not used in the
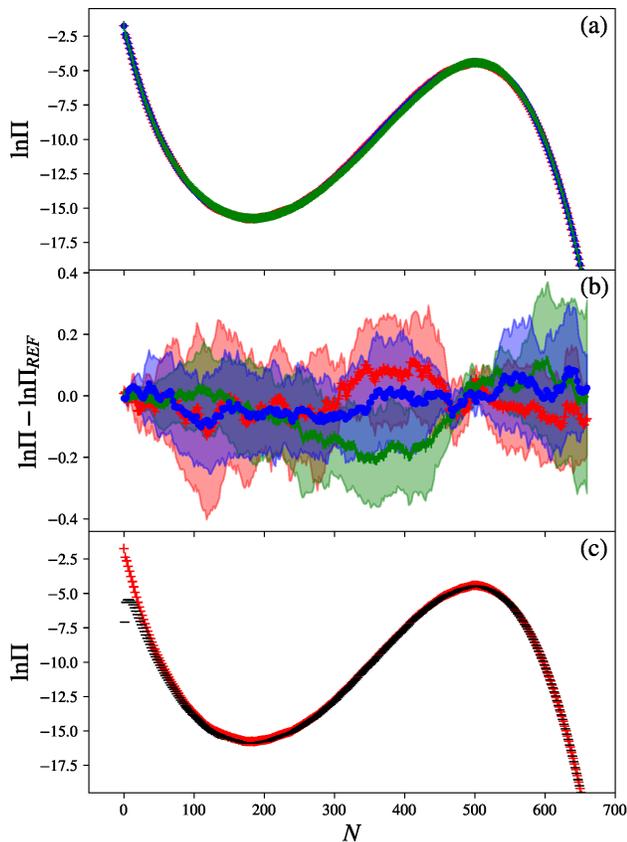
FIG. 2: Comparison of the macrostate probability distribution, $\ln \Pi$, as a function of the number of ions, N, at $T^* = 0.04294$ for the following four different simulation methods: (black x) REF: previously published results using neutral pair transfers,[3] (blue .) PAIR: neutral pair transfers, (red +) EE: single ion transfers with expanded ensemble and (green |) EE-AVB: EE with the additional AVB trial. All data points include error bars that are the standard deviation of four independent simulations. (a) The macrostate distribution for all four methods. (b) The difference between the macrostate distribution and the reference simulation of the previous[3] work, $\ln \Pi_{REF}$. The standard deviation is shown by the shaded regions. (c) The odd number of ion states in the expanded ensemble (EE) simulations are shown by the black horizontal bar symbols.

FIG. 3: Liquid vapor saturation properties of the RPM for four different simulation types with symbols as described in Figure 2. Error bars are the standard deviation of four independent simulations, and are smaller than the size of the symbols.

TABLE III: Liquid vapor saturation properties of the RPM using the various methods. Standard deviations from four independent simulations are shown by subscript with order of magnitude of the last digit.

T^*	simul.	$\ln \rho_v^*$	ρ_l^*	$\ln p^*$	$\ln z$
0.0479	REF	-4.092_{21}	0.178_1	-9.359_9	-27.8818_9
	PAIR	-4.115_{54}	0.1768_7	-9.358_{15}	-27.8824_1
	EE	-4.075_{31}	0.1777_{17}	-9.35_1	-27.8820_4
	EE-AVB	-4.101_{32}	0.1782_7	-9.360_7	-27.8818_{11}
0.0454	REF	-5.140_7	0.2378_5	-9.944_5	-29.3328_9
	PAIR	-5.153_{12}	0.2391_{11}	-9.947_9	-29.3327_7
	EE	-5.168_{27}	0.2387_{11}	-9.941_{22}	-29.3332_{13}
	EE-AVB	-5.145_9	0.2371_{14}	-9.944_4	-29.3332_6
0.0429	REF	-5.972_2	0.2862_3	-10.551_2	-30.9865_4
	PAIR	-5.958_{16}	0.2865_{14}	-10.542_7	-30.9859_5
	EE	-5.981_{31}	0.2858_{17}	-10.557_{23}	-30.9867_{10}
	EE-AVB	-5.974_{10}	0.2874_{10}	-10.553_5	-30.9870_{12}
0.0388	REF	-7.218_{23}	0.3568_5	-11.703_{13}	-34.2929_5
	PAIR	-7.219_7	0.3563_4	-11.704_5	-34.2933_9
	EE	-7.204_{24}	0.3565_6	-11.689_{14}	-34.2924_9
	EE-AVB	-7.226_7	0.3557_9	-11.708_3	-34.2936_{11}
0.0347	REF	-8.673_{21}	0.4241_{16}	-13.164_{11}	-38.5057_{24}
	PAIR	-8.692_{21}	0.4236_{13}	-13.176_{14}	-38.5065_{13}
	EE	-8.728_{27}	0.4231_4	-13.213_{32}	-38.5063_{15}
	EE-AVB	-8.690_4	0.4223_9	-13.174_2	-38.5052_6

calculation of saturation or other thermodynamic properties.

The liquid vapor saturation densities, pressures and activities of the RPM are shown in Figure 3 for a variety of temperatures. Because the error bars are smaller than the symbols, we also report them in Table III. These saturation properties agree quantitatively within 95 % confidence in comparison to the various methods in this work, and also in previous work.[3] Note that previous work[3] used grand canonical temperature-expanded ensemble simulations to determine the low-temperature saturation properties, which avoid the direct calculation of the entire macrostate distribution. This obviates the need to sample the intermediate N range where the system undergoes structural transitions that are difficult to sample.[55] Because the expanded grand canonical ensemble method described here also directly calculates the entire macrostate distribution, it will also encounter difficulties sampling the structural transitions at low temperatures and will also have trouble determining the corresponding saturation properties.

VI. EFFICIENCY

Now that we have shown the expanded ensemble method for ions yields results for the RPM that agree with reference values, we turn our attention to efficiency. We show up to an order of magnitude increase in efficiency at high density using the expanded ensemble method. We also discuss room for improvement for the expanded ensemble simulations at lower temperature and low density. The efficiency measures presented in this section depend sensitively upon the implementation, especially for more specialized techniques such as AVB which require lists of neighbors. More efficient implementations or optimized parameter choices could lead to large changes in these results.

Because the comparison simulations in the previous section were run for different numbers of sweeps on different types of processors, we perform additional benchmark simulations to report on central processing unit (CPU) timings for convergence and trial acceptance probabilities. Benchmark simulations were run for the lowest and highest temperatures and with vapor and liquid densities using windows of size 20 ions. Note that the vapor density simulations were conducted without DCCB because it is unnecessary at low density.

Benchmark simulations in this section were terminated after 10 sweeps, and the time, t, to reach this level of convergence is reported in Table IV. The benchmark simulations were in agreement with the simulations reported in Section V. Note that only macrostates with even numbers of ions were considered in the sweep definition in order to improve the comparison with pair insertion/deletion simulations, which only visit even-numbered macrostates. The acceptance probabilities for ion insertions/deletions, P_{acc}, averaged over each trial type, are also reported. Benchmark simulation averages and standard deviations were obtained with three independent simulations.

Additional details of the hardware are provided for the purposes of reproducibility. Benchmark simulations were run on a 6-core, 3.60 GHz Intel(R) Xeon(R) CPU E5-1650 v4. See the disclaimer at the end of the manuscript regarding company identification. Processor load balancing led to an underestimation of the efficiency gains, because the processor was at full capacity for the shorter simulations, while longer simulations may have experienced a boost in GHz after the short simulations finish.

A. Efficiency comparison of EE and PAIR

To begin, we compare the relative efficiency of the expanded ensemble (EE) simulations and the conventional pair insertion (PAIR) simulations without the complication of including aggregation volume bias. Intuitively, decoupling the transfer of a pair of ions into a dense liquid with two steps instead of one is more efficient because the probability of finding cavities for both ions simulta-

TABLE IV: Convergence time and acceptance probability. Standard deviations from three independent simulations are shown by subscript with order of magnitude of the last digit.

T^*	N range	simul.	t (hours)	P_{acc}
0.0479	0 to 20	PAIR	0.19_1	0.0045_9
		EE	0.16_1	0.032_9
		PAIR-AVB	0.07_1	0.092_1
		EE-AVB	0.05_1	0.027_1
	724 to 744	PAIR	9.8_3	0.00042_4
		EE	1.64_5	0.0055_1
		PAIR-AVB	15.2_7	0.00073_6
		EE-AVB	2.38_7	0.0041_1
0.0347	0 to 20	PAIR	0.8_1	0.0014_2
		EE	0.9_1	0.019_9
		PAIR-AVB	0.08_1	0.027_2
		EE-AVB	0.16_2	0.015_3
	724 to 744	PAIR	71_5	0.000048_6
		EE	6.1_4	0.0013_1
		PAIR-AVB	26_4	0.00031_1
		EE-AVB	14_1	0.0010_1

neously is much lower than the probability of finding a single cavity. This is why the largest efficiency gains can be seen in dense liquids, with a factor of 12 improvement in convergence time at $T^* = 0.0347$ at saturated liquid density when comparing the EE method to the PAIR method.

The overhead of the expanded ensemble approach is negligible. This is because the majority of processor time is spent on computing the interactions between ions, and the cost of computing the interaction of a single ion with the rest of the system is roughly half that of computing the interaction of a pair of ions. Thus, even in the vapor phase, where finding a cavity is not an issue, the method does not decrease the convergence rate due to higher overhead, as might be observed in other specialized methods such as DCCB and AVB.

A potential concern is that single ion insertions in the vapor phase would be less efficient because of the preference for dimers or cluster formation at lower temperature. For example, Figure 2c shows that the probability of observing the macrostates with an odd number of ions is significantly lower than for the even numbers. But the flat histogram method overcomes this barrier via bias on the insertion of the lone ion, which would normally not be favored. A subsequent trial insertion may then find the lone ion to form a new dimer. This allows the convergence times of the PAIR and EE methods to be roughly equivalent for the vapor densities.

Note that the macrostate range of the EE method is twice that of the pair. Based on this observation, one might expect the EE approach to be slower when acceptance probabilities are high. However, the data show that, for low density, the two convergence times are comparable despite this disparity in the macrostate range. An additional concern is that many of the sampled configurations are discarded and not used for ensemble av-

erages. For $z = \pm 1$, the expanded ensemble approach does not utilize the configurations with odd numbers of ions in subsequent calculation of thermodynamic and structure properties. But neighboring configurations in the Markov chain are highly correlated. Thus, removing these correlated configurations with odd numbers of ions (for $z = \pm 1$) does not greatly impact the precision or accuracy of the ensemble averages, especially considering the improvements in sampling.

B. Efficiency comparisons with AVB

The AVB method introduced in Section IV for EE-AVB was also utilized with pair insertions and denoted here as PAIR-AVB. To our knowledge, the following modified version of the AVB algorithm for attractive pairs has not been published previously, but was developed in this work to emulate the distance-biased approach of previous work, which was used to tackle the problem of dimer or cluster formation in the vapor phase.[32] The major difference between these two methods is that AVB has a uniform distance bias within the cutoff distances, while the continuous distance bias scales in various ways.[32,44] The difference in efficiencies between these two methods, and the optimization of their parameters, is beyond the scope of this work. The PAIR-AVB algorithm proceeds as follows. For insertions, the first ion is added anywhere in the system volume, V. The aggregation volume, v^{in}, is then defined as the spherical shell centered on the first ion with an inner and outer radii of σ and 1.5σ, respectively. To compare with distance bias methods at the same temperature, we calculate that over 99% of the bias-selected distances would fall within this aggregation volume region. A second ion is then placed in v^{in} using a uniform random distribution. The trial is accepted with a probability of $\min(1, \chi^{add})$, where $\chi_1^{add} = \frac{V}{N_1+1} \frac{v^{in}}{n_2^{in}+1} \exp[-\beta(\Delta U - \mu_1 - \mu_2)]$, N_1 is the number of ions of the first type, and n_2^{in} is the number of ions of the second type in v^{in}. The ion numbers N_1 and n_2^{in} are for the old configuration and do not include the proposed additions. For deletions, an ion is randomly selected. If this ion contains the opposite type of ion within v^{in}, then both ions are selected for a deletion attempt. Otherwise, the trial is rejected. For the deletion trial, $\chi_{del} = \frac{N_1}{V} \frac{n_2^{in}}{v^{in}} \exp[-\beta(\Delta U + \mu_1 + \mu_2)]$.

We first examine simulation efficiencies at high densities with and without AVB. The EE method converged the fastest at high density among all four types of simulations and both temperatures. The reason the EE-AVB method is slower than EE at high density is likely due to overhead. The overhead of the AVB method increases at high density due to the requirement to keep track of all neighbors, in addition to an increased role of excluded volume in determining acceptance probabilities. However, for PAIR-AVB, there is a nontrivial temperature dependence. At the higher temperature and high density,

PAIR-AVB is slower than PAIR due to the overhead as mentioned above. But at the lower temperature and high density, PAIR-AVB is faster than PAIR. Regardless, EE and EE-AVB are faster than both PAIR and PAIR-AVB at high density.

Now we consider the lower density simulations. At high temperature and low density, both PAIR-AVB and EE-AVB converge at about the same time and they both see about the same increase in efficiency over their non-AVB counterparts. But at the lower temperature, the PAIR-AVB simulation was found to be twice as efficient as EE-AVB. Note that the convergence times for the vapor phase are much smaller than the higher densities, and thus the increased efficiency at high density for the EE simulations outweighs the increase in efficiency at low density for the PAIR-AVB simulations. The PAIR-AVB method implemented in this work is expected to be highly efficient for the vapor phase, because the vapor phase is composed primarily of neutral pairs or dimers. In the PAIR-AVB simulations, these neutral pairs are explicitly biased for insertion and deletion in a single step. In comparison, the EE-AVB simulations, as implemented in this work, do not explicitly target lone ions for deletion or AVB insertion. For example, the EE-AVB single ion insertion/deletion step may attempt the AVB insertion on any ion, instead of targeting lone ions to create pairs. The EE-AVB single ion trial as implemented also does not help with the deletion of lone ions. Thus, the differences in how the AVB method is implemented with EE-AVB and PAIR-AVB contribute significantly to the measures of efficiency. Alternative EE-AVB methods could be developed which are more analogous to PAIR-AVB. Such a method could involve a trial where a lone inserted ion is targeted for AVB insertion of the next ion of the opposite type. Although such an alternative method would be more analogous to PAIR-AVB, the focus of this manuscript is on efficiency at high density. Thus, development of an EE-AVB method which is more similar to PAIR-AVB is beyond the scope of this work.

VII. CONCLUSIONS

We show that the proposed expanded grand canonical ensembles method for ions agrees quantitatively with previous methods and is up to an order of magnitude more efficient at high density in the restricted primitive model electrolyte. Further improvements in sampling are expected as the complexity of the model increases (e.g., all-atom models of room temperature ionic liquids). While some room for improvement of the expanded ensemble method at the low density and low temperature conditions remain, the efficiency gain at high density, which is where most of the CPU time is spent, outweighs the minor differences at low density.

Because the implementation of the flat histogram expanded ensemble method in an existing single-particle grand canonical Monte Carlo simulation code is likely

to be more straightforward than the implementation of pair insertions and deletions, we recommend that future projects utilize this expanded ensemble method. Implementation of pair ion transfers requires careful modification of how interaction energies are computed both between the pair of transferred ions with the rest of the system, and also between themselves. In addition, the Metropolis acceptance criteria must also be modified.[32] However, for the expanded ensemble approach, neither the ion transfer trial nor the Metropolis acceptance criteria need be modified from a traditional single particle trial. And no special Ewald corrections are required for quantities such as energy, pressure and chemical potential. Instead, the acceptance criteria only need to include a bias from a flat histogram approach,[28] with appropriate constraints on the numbers of the particles. The flat histogram approach provides a significant increase in efficiency by itself and is relatively straightforward to implement in an existing code.[7,48]

Future work that remains to be completed includes the simulation of multivalent ions. This would require additional steps in the expanded ensemble, but its implementation is an extension of this work with the introduction of a new path. For example, the bridge between neutral states might involve the addition of a monovalent ion, then an oppositely charged divalent ion, and finally the addition of a second, identical monovalent ion. In addition, the path shown in Figure 1 is not the only possible path even for a monovalent ion. For example, the path between neutral states could branch to include either A or B ions. Our preliminary results show that the free energy of the neutral states is independent of the choice of the pathway, but different pathways may have an effect on the efficiency of the method.

VIII. ACKNOWLEDGEMENTS

[1] N. V.Plechkova and K. R.Seddon, Chemical Society Reviews **37**, 123 (2008).

[2] R. D. Rogers and K. R. Seddon, Science **302**, 792 (2003).

[3] K. S. Rane and J. R. Errington, J. Phys. Chem. B **117**, 8018 (2013).

[4] G. E. Norman and V. S. Filinov, High Temperature **7**, 216 (1969).

[5] S. A. Barr and A. Z. Panagiotopoulos, Phys. Rev. E **86** (2012).

[6] A. Z. Panagiotopoulos, Mol. Phys. **61**, 813 (1987).

[7] A. S. Paluch, V. K. Shen, and J. R. Errington, Ind. Eng. Chem. Res. **47**, 4533 (2008).

[8] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Vol. 1 (Academic, San Diego, 2002).

[9] J. I. Siepmann and D. Frenkel, Mol. Phys. **75**, 59 (1992).

[10] B. Chen and J. I. Siepmann, J. Phys. Chem. B **105**, 11275 (2001).

[11] F. A. Escobedo and J. J. de Pablo, J. Chem. Phys. **105**, 4391 (1996).

[12] K. S. Rane, V. Kumar, and J. R. Errington, J. Chem. Phys. **135**, 234102 (2011).

[13] K. S. Rane, S. Murali, and J. R. Errington, J. Chem. Theory Comput. **9**, 2552 (2013).

[14] H. W. Hatch, W. P. Krekelberg, S. D. Hudson, and V. K. Shen, J. Chem. Phys. **144**, 194902 (2016).

[15] W. Shi and E. J. Maginn, J. Chem. Theory Comput. **3**, 1451 (2007).

[16] S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **373**, 27 (1980).

[17] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[18] G. Hummer, L. R. Pratt, and A. E. García, J. Phys. Chem. **100**, 1206 (1996).

[19] P. Sloth and T. S. Sørensen, Chemical Physics Letters **173**, 51 (1990).

[20] A. Malasics and D. Boda, J. Chem. Phys. **132**, 244103 (2010).

[21] J. S. Hub, B. L. de Groot, H. Grubmller, and G. Groenhof, J. Chem. Theory Comput. **10**, 381 (2014).

[22] J. P. Hansen, Physical Review A **8**, 3096 (1973).

[23] S. Bogusz, T. E. Cheatham, and B. R. Brooks, J. Chem. Phys. **108**, 7070 (1998).

[24] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).

[25] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[26] J. R. Errington, J. Chem. Phys. **118**, 9915 (2003).

[27] V. K. Shen and J. R. Errington, J. Chem. Phys. **122**, 064508 (2005).

[28] V. K. Shen and D. W. Siderius, J. Chem. Phys. **140**, 244106 (2014).

[29] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 4 (1988).

[30] W. P. Krekelberg, V. K. Shen, J. R. Errington, and T. M. Truskett, J. Chem. Phys. **128**, 161101 (2008).

[31] J. P. Valleau and L. K. Cohen, J. Chem. Phys. **72**, 5935 (1980).

[32] G. Orkoulas and A. Z. Panagiotopoulos, J. Chem. Phys. **101**, 1452 (1994).

[33] A. Z. Panagiotopoulos, Fluid Phase Equilibria **76**, 97 (1992).

[34] J. Caillol and J. Weis, J. Chem. Phys. **102**, 7610 (1995).

[35] G. Orkoulas and A. Z. Panagiotopoulos, J. Chem. Phys. **110**, 1581 (1999).

[36] A. Z. Panagiotopoulos and M. E. Fisher, Phys. Rev. Lett. **88** (2002).

[37] A. Z. Panagiotopoulos, J. Chem. Phys. **116**, 3007 (2002).

[38] C. Valeriani, P. J. Camp, J. W. Zwanikken, R. van Roij, and M. Dijkstra, J. Phys. Condens. Matter **22**, 104122 (2010).

[39] B. Smit, K. Esselink, and D. Frenkel, Mol. Phys. **87**, 159 (1996).

[40] Q. Yan and J. J. de Pablo, J. Chem. Phys. **111**, 9509 (1999).

[41] M. Gonzlez-Melchor, F. Bresme, and J. Alejandre, J. Chem. Phys. **122**, 104710 (2005).

[42] J. M. Caillol, D. Levesque, and J. J. Weis, J. Chem. Phys. **107**, 1565 (1997).

[43] E. Luijten, M. E. Fisher, and A. Z. Panagiotopoulos, Phys. Rev. Lett. **88**, 185701 (2002).

[44] A.-P. Hynninen, M. Dijkstra, and A. Z. Panagiotopoulos, J. Chem. Phys. **123**, 084903 (2005).

[45] N. Rai and E. J. Maginn, J. Phys. Chem. Letters **2**, 1439 (2011).

[46] K. S. Rane and J. R. Errington, J. Phys. Chem. B **118**, 8734 (2014).

[47] Q. R. Sheridan, R. G. Mullen, T. B. Lee, E. J. Maginn, and W. F. Schneider, J. Phys. Chem. C **122**, 14213 (2018).

[48] H. W. Hatch, N. A. Mahynski, and V. K. Shen, J. Res. Natl. Inst. Stan **123** (2018).

[49] A.-P. Hynninen and A. Z. Panagiotopoulos, Mol. Phys. **106**, 2039 (2008).

[50] T. J. H. Vlugt, M. G. Martin, B. Smit, J. I. Siepmann, and R. Krishna, Mol. Phys. **94**, 727 (1998).

[51] B. Chen and J. I. Siepmann, J. Phys. Chem. B **104**, 8725 (2000).

[52] B. Chen, J. I. Siepmann, K. J. Oh, and M. L. Klein, J. Chem. Phys. **115**, 10903 (2001).

[53] B. Chen, J. I. Siepmann, K. J. Oh, and M. L. Klein, J. Chem. Phys. **116**, 4317 (2002).

[54] H. W. Hatch, S.-Y. Yang, J. Mittal, and V. K. Shen, Soft Matter **12**, 4170 (2016).

[55] L. G. MacDowell, V. K. Shen, and J. R. Errington, J. Chem. Phys. **125**, 034705 (2006).