PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2017, **19**, 18968

Phase and interface determination in computer simulations of liquid mixtures with high partial miscibility†

Marcello Sega *[a] and György Hantal [ab]

Received 3rd May 2017,
Accepted 26th June 2017

DOI: 10.1039/c7cp02918g

rsc.li/pccp

Partially miscible solutions can represent a challenge from the computer simulation standpoint, especially if the mutual solubility of the components is so large that their concentrations do not change much from one phase to another. In this case, identifying which molecules belong to which phase becomes a complicated task. Here, we propose a density-based clustering approach with self-tuning capabilities and apply it to the case of the mixture of an ionic liquid with benzene. The almost linear scaling of the algorithm makes it suitable for the analysis of long Molecular Dynamics or Monte Carlo trajectories.

1 Introduction

The physics and chemistry of fluids and soft materials are strongly influenced by the presence of interfaces separating different phases or micro-phases,[1–3] and many effects that are relevant for determining the macroscopic behaviour of these systems are occurring right at these interfaces, often within the first few molecular layers.[4] Being able to determine and selectively study the molecules at interfaces is therefore very important in order to improve our understanding of multi-phase systems. Computer simulations have been instrumental in changing our picture of fluid interfaces, showing that, for example, far enough from critical points, their local structure is not at all characterized by the slowly varying, sigmoidal-shaped density profile predicted by mean field theories.[5] Rather, the structure of liquid interfaces presents features that closely resemble the molecular pair correlation function. Thermal capillary waves, in fact, smear this sharply structured, intrinsic profile into a smooth function that bridges the two bulk phases across the interface. There are approaches, such as that of Berkowitz,[6] that estimate the average distribution of surface atoms in the reference frame of the simulation box. However, one needs to have access to the set of surface atoms in every frame[7] in order to compute intrinsic quantities.

Determining the interface between two phases in an atomistic computer simulation, however, often requires *a priori* knowledge of which molecules belong to which phase. Far enough from critical points (be that the liquid–vapour critical point for single component systems, or the critical mixing point in bicomponent systems), the problem of distinguishing the two phases is less severe: simulating water/vapour coexistence at a temperature of 300 K in a simulation box with an edge of about 10 nm, one can hardly find a single molecule in the vapour phase. If the vapour phase is composed of few molecules, a possible way to separate the two phases consists of defining clusters of molecules that are within a given cutoff distance from each other. If the cutoff distance corresponds to the location of the first minimum of the liquid-phase radial distribution function, the liquid phase can be reasonably defined as the largest cluster in the system.[8]

This approach fails if the density of the two phases becomes too similar, as this clustering approach will identify a single, large, connected cluster spanning the complete simulation box. Another, more robust approach consists of considering the liquid phase as the largest cluster of molecules, which have at least a given number of neighbours.[7,9] Also this criterion has its drawbacks, as we will show, as it does not consider points with smaller local density, which are still reachable, within the cutoff distance, from the denser regions.

The equivalent situation for bicomponent systems is that of partial miscibility, where the two components A and B can give rise to an A-rich phase Φ_A and a B-rich phase Φ_B, where the respective concentrations are $c_A(\Phi_A) > c_A(\Phi_B)$ and $c_B(\Phi_A) < c_B(\Phi_B)$. The solubility of A in B does not need to be the same as that of B in A, which can lead to extreme situations in which one phase is a single component one (say, $c_A(\Phi_B) = 0$), while in the other phase the two components have similar concentrations, $c_A(\Phi_A) \simeq c_B(\Phi_A)$.

This condition is exemplified by the case of the ionic liquid 1-butyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide (BMIM NTf_2) in combination with the simplest aromatic solvent, benzene. The concentration of BMIM NTf_2 in the benzene-rich

[a] *Faculty of Physics, University of Vienna, Boltzmangasse 5, 1090 Vienna, Austria. E-mail: marcello.sega@univie.ac.at*
[b] *Department of Chemistry, Eszterházy Károly University, Leányka utca 6, H-3300 Eger, Hungary*

† Electronic supplementary information (ESI) available: Phase determination using different algorithms and low vapour pressure. See DOI: 10.1039/c7cp02918g

phase is negligible, while in the ionic liquid-rich phase, benzene is present in an extremely high concentration that reaches a molar ratio $c_{benzene}/c_{total}$ of 0.82.[10] Previously, we investigated the structure of fluid interfaces between ionic liquids and other fluids with little or moderate mutual miscibility.[11] In particular, we were interested in how the intrinsic interface structures of two of the most popular ionic liquid compounds, BMIM NTf_2 and 1-butyl-3-methylimidazolium hexafluorophosphate (BMIM PF_6), change upon varying the polarity of the opposite fluid phase. The interfacial structure of the mixture of BMIM NTf_2 and benzene, however, cannot be studied using the approaches used so far, because of the peculiar mixing properties just described.

This is evident in Fig. 1, where we report a simulation snapshot of the mixture of BMIM NTf_2 and benzene at room temperature. Benzene molecules are represented using a contour joining the six carbon atoms, while atoms of BMIM NTf_2 are represented with a space-filling model using blue spheres (Fig. 1, left panel). The ionic liquid is clearly separating into two phases, one of which does not show the presence of any $BMIM^+$ or NTf_2^- ions. Benzene, on the other hand, is present throughout the simulation box, as one can appreciate by removing the ionic liquid molecules (Fig. 1, right panel), forming a percolating system that cannot easily be separated into two phases using the simple clustering criterion mentioned before.

However, the peculiar properties of this mixture make it a perfect candidate for testing new algorithms that work closer to critical points. This is because the well separated BMIM NTf_2 phase can serve as a clear reference for the other component. One can thus test any algorithm that aims to separate the high benzene concentration phase from the low benzene concentration one by checking that the former is complementary to the ionic liquid rich region.

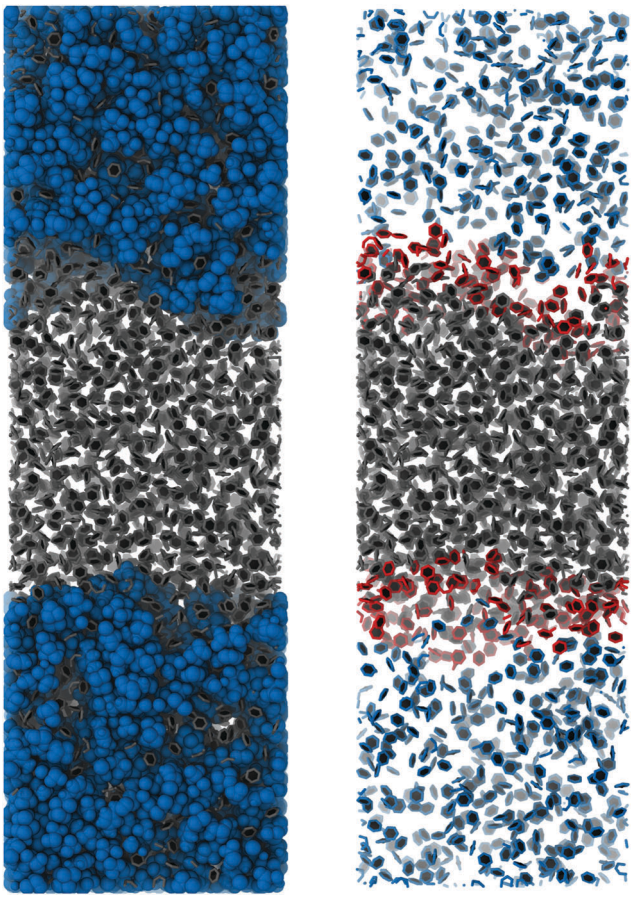

Fig. 1 Simulation snapshot of an equilibrated BMIM NTf_2/benzene mixture. Left panel: All components shown, as blue spheres (BMIM cations and NTf_2 anions) and grey sticks (benzene rings). Right panel: Only the benzene rings are shown (blue: low density phase; red: interfacial rings (as identified by the ITIM algorithm, see text) of the high-density phase; grey: non-interfacial rings of the high-density phase). The two phases have been identified using the present DBSCAN-based algorithm, while the interfacial molecules have been identified using ITIM.

2 A density-based clustering approach

One of the most widespread clustering algorithms for class identification in spatial databases is the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm.[12] According to the basic idea of DBSCAN, the so-called core points of a set are those with, within a cut-off r_c, at least N_c neighbours or, equivalently, a given local threshold density $\rho_c = 3/(4\pi r_c^3)N_c$; directly density-reachable points are then those found within r_c from core points. Two points are said to be density-connected if both can be reached, starting from the same core point, through a chain of directly density-reachable intermediate points. A cluster is finally defined as the set of all points, which are density-connected to one core point.

It is clear that the condition $N_c = 1$ yields the simple clustering algorithm mentioned in the introduction, which is therefore a special case of DBSCAN. A clustering algorithm which is limited to core points, therefore not including density-reachable ones, has been used by Chacón and Tarazona to separate small vapour clusters from the liquid phase.[7,9]

From the point of view of computational complexity, DBSCAN can be performed in a time $\mathcal{O}(N \log N)$,[12] where N is the number of points, because the core points and the directly density-reachable points associated with them can be identified using, for example, a kd-tree based search that takes (on average) $\mathcal{O}(\log N)$ iterations.[13]

In Fig. 2 we show the result of DBSCAN filtering applied to a set of points on the x–y plane, whose coordinates are sampled from two uniform random distributions (upper panel), a more dense one in the interval x from 0 to 1, and a less dense one (1/3 relative density) for x from 1 to 2. Along the y axis, all coordinates are sampled from a uniform random distribution from 0 to 1. Periodic boundary conditions are applied along the y axis. For this example, $r_c = 0.18$ and $N_c = 24$. The DBSCAN filtering (Fig. 2, lower panel) results in the identification of the largest cluster composed of core points (red points) and density-reachable points (black circles), which almost coincide with the set of points sampled at higher density ($x < 1$). Note that although the algorithm does fairly well in recovering the two initial distributions, it is not expected (and in our case, not meant) to do so: at the interface, $x \simeq 1$, the juxtaposition of the

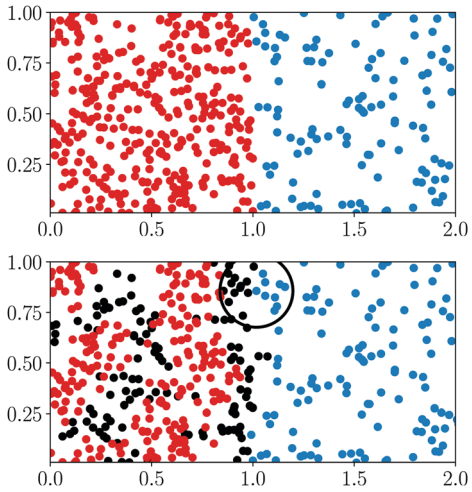

Fig. 2 Example of DBSCAN filtering applied to two uniform distributions. Upper panel: the two initial distributions (red points, $x < 1$: dense phase; blue circles $x > 1$: diluted phase). Lower panel: result of the DBSCAN filtering (red points: core points of the biggest cluster; black circles: density-reachable points of the biggest cluster; blue circles: points not belonging to the biggest cluster). The black circle shows the cutoff radius used to determine the local density.

two distributions creates regions of high local density for $x > 1$, which are reachable from the main cluster. Conversely, regions of low local density for $x < 1$ can occur just because of the random nature of the distribution and, if they occur close to $x = 1$, they can become part of some of the smaller clusters. The importance of considering not only core points (that is, those with a local density higher than the chosen threshold) but also the density-reachable ones can be appreciated by noticing, in Fig. 2, the number of non-core points lying well within the high-density region.

We tested the performances of the algorithm by applying it to similar distributions in the three-dimensional space, by varying the number of points N_{points} from 500 to 5×10^6. The time it takes to perform the filtering, shown in Fig. 3 as blue circles, is virtually indistinguishable from a linear function of N_{points} (dashed line). The prefactors involved in the calculation are also relatively low, as indicated by the timing of about 1 s for 1 million points.

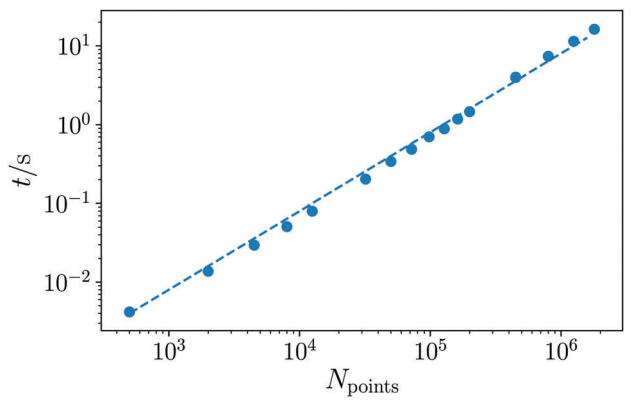

Fig. 3 Computing time needed to apply the DBSCAN filter to systems with a different number of points, N_{points}.

3 Automating the threshold determination

The almost linear scaling of DBSCAN makes it suitable for processing large datasets such as molecular dynamics trajectories. However, differently from the simple neighbour clustering algorithm, two parameters must be provided (r_c and ρ_c), rather than simply the cutoff radius r_c. The choice of this pair of parameters is crucial, as they determine whether the algorithm will be able to distinguish correctly the two phases or not. The identification of proper values of r_c and ρ_c should be the outcome of a simple procedure, if not of an automated one, in order for the algorithm to be of practical use.

A common approach in the simple neighbour clustering algorithm is to use as a cutoff radius the distance at which the first minimum of the radial distribution function (of the dense phase) is located. In this way, two molecules are considered to be in the same cluster if they are located at a distance, which would make them (by definition) belong to each other's shell of first neighbours. In the case of the BMIM NTf_2 benzene mixture, benzene separates in two regions, where its concentration changes from about 7 to 3 molecules per nm^3 (see Fig. 4). The intermolecular radial distribution function for the carbons of benzene rings in the high and low benzene concentration regions are shown in Fig. 5. In the high concentration region, a clear, deep minimum is present at around 0.8 nm, but a less pronounced one appears not far from the close contact distance, at about 0.55 nm. Note that the first peak of the radial distribution function in the high benzene concentration region is lower than the corresponding one in the low benzene concentration region: this could seem counterintuitive at first, but one should not forget that the functions are normalized to their respective bulk concentrations, which have a ratio of 3 : 7. Therefore, the local concentration is always smaller, in absolute terms, in the low benzene concentration region than in the high one.

To understand which approach could be best suited for choosing the cutoff density ρ_c, we calculated the distribution of the number of neighbouring molecules for different

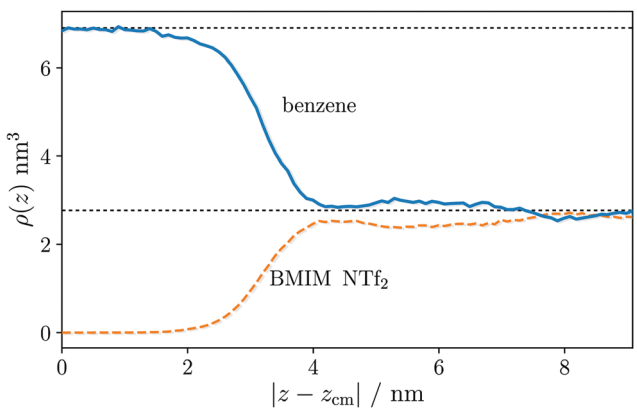

Fig. 4 The concentration profile (in molecules per nm^3) of benzene (solid line) and BMIM and NTf_2 ions (dashed line). Dotted lines are the result of a best fit to the concentration of benzene in the regions far from the interface.

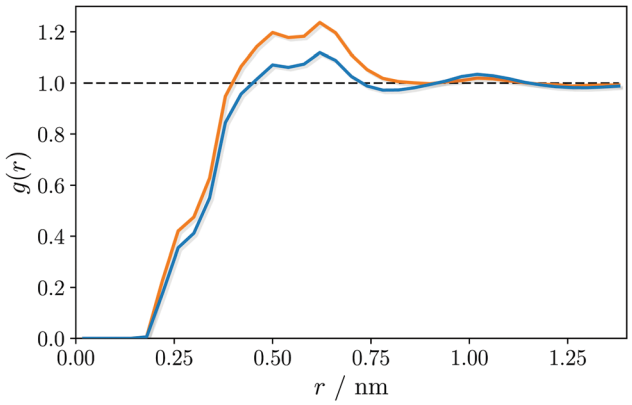

Fig. 5 Intermolecular radial distribution function of the carbon atoms (blue curve: high benzene concentration region; orange curve: low benzene concentration region).

cutoff radii, r_c, calculated over the whole trajectory and over all benzene molecules in the simulation box. The results are presented in Fig. 6. When using a small cutoff radius, $r_c = 0.45$ nm, the distribution of the number of neighbours N_n shows a shoulder at about $N_n = 5$, hinting at the presence of an underlying multimodal distribution, which cannot be properly resolved. By increasing the cutoff to r_c = 0.6 and 0.75 nm, the distribution stretches to a higher number of neighbours, as the available volume increases, and, at the same time, the presence of two peaks is gradually revealed. The two peaks, corresponding to the density distributions in the two concentration regions, are best resolved when a large cutoff is being used. We have chosen to characterize the distribution using a *k*-means clustering algorithm[14] with two modes, which provides two centroids (represented by two circles in Fig. 6 for $r_c = 0.75$ nm). From the average of the positions of the two centroids (denoted by a diamond symbol in Fig. 6) one can obtain a possible estimate of the threshold density ρ_c, as below/above the corresponding number of neighbours most of the distribution is contributed to by atoms in the low/high benzene concentration region, respectively.

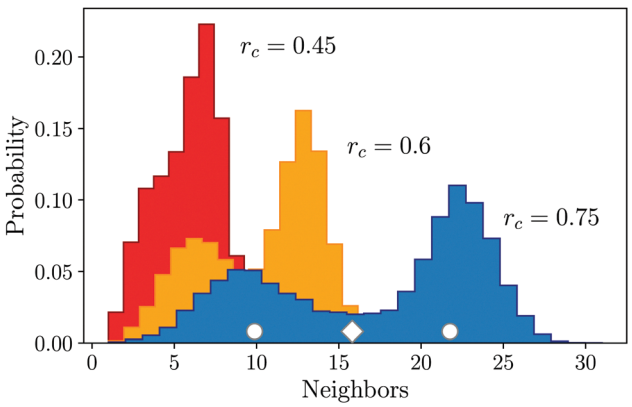

Fig. 6 Histogram of the number of benzene molecule pairs with some atoms closer than a cutoff of 0.45, 0.6 and 0.75 nm, respectively. The marks represent the position of the centroids for the r_c = 0.75 nm case, as determined by bimodal *k*-means clustering (9.9 and 21.8) and their mean (15.8).

Due to the non-miscibility of the ionic liquid in benzene, no ionic liquid should be found in the benzene-rich phase determined by the algorithm. This sets a fairly restrictive criterion to be met by the algorithm. However, by using the midpoint of the two centroids, about 5% of the total number of ionic liquid molecules (51 ions out of about 1024 in the system) can be found within the high benzene concentration region, showing that this choice of ρ_c is not adequate. The reason for this behaviour stems from the fact that by placing the cutoff at the midpoint of the two centroids, some benzene molecules in the right tail of the distribution centred at the lower concentration are identified as belonging to the high concentration region. By using a cutoff concentration corresponding to the upper centroid, instead of the midpoint, the proportion of ions found in the high benzene concentration region drops to about 0.5%. As there are 512 ion pairs in the system, this means that on average about one ion pair (per side) is found across the benzene interface. The choice of the upper centroid concentration as a threshold one does, therefore, improve noticeably the quality of the phase identification. This can be appreciated by inspecting the intrinsic density profile,[9] calculated by constructing the histogram of the distance of molecules from the local position of the interface $\xi(x, y)$, as

$$\rho(z) = \frac{1}{A}\left\langle \sum_i \delta(z - z_i + \xi(x_i, y_i)) \right\rangle \qquad (1)$$

as presented in Fig. 7. The virtual absence of ions inside the benzene rich phase shows that this automatic choice for the threshold density ρ_c is in fact able to distinguish the two phases in a consistent way. The intrinsic density profile shows another interesting feature of the mixture, namely, the presence of a depletion layer of benzene next to the interface, and a corresponding increase in the concentration of the ionic liquid, which cannot be noticed in the non-intrinsic density profile.

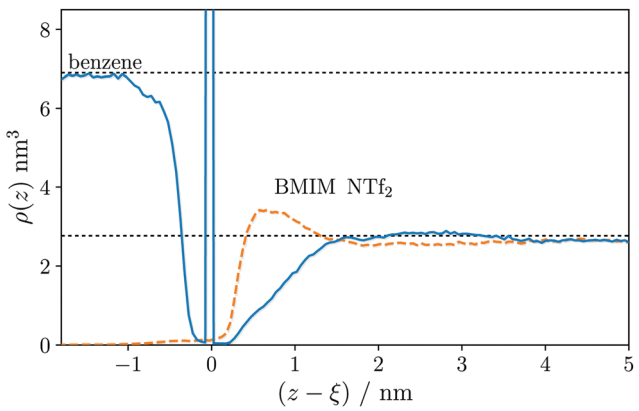

Fig. 7 Intrinsic concentration profile (in molecules per nm^3) of benzene (solid line) and BMIM and NTf_2 ions (dashed line), with respect to the location ζ of the interfacial benzene molecules obtained using as the threshold density ρ_c that of the highest density centroid. Only 0.5% of the ionic liquid molecules (*i.e.* less than one pair on average) are found in the high benzene concentration region.

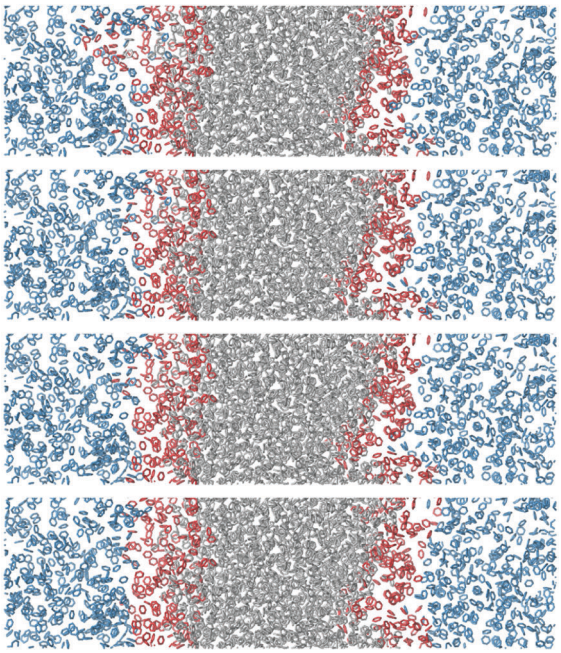

Fig. 8 Benzene molecules in one simulation snapshot assigned to different phases (gray: benzene-rich phase; red: interfacial molecules of the benzene-rich phase; blue: ionic liquid-rich phase) using different values for the cutoff of the DBSCAN algorithm. From top to bottom: cutoff of 0.4, 0.6, 0.8, and 1.2 nm. Notice that the system is translated so that the centre of mass of the largest cluster is located in the middle of the box.

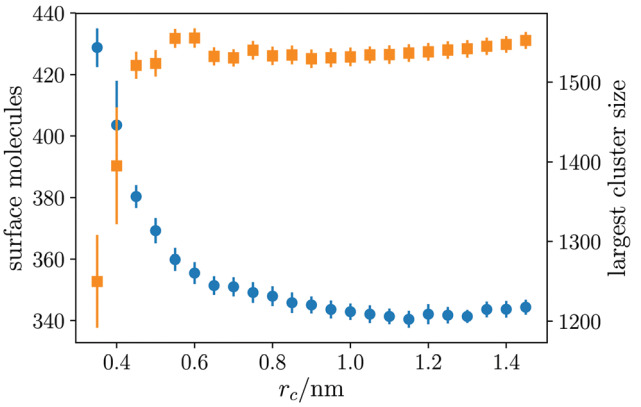

Fig. 9 Number of surface molecules (orange squares) and molecules belonging to the largest cluster (blue circles) as a function of the cutoff radius r_c.

Another important requirement for the algorithm is its stability against small (or large) variation of r_c, the only free parameter left (as ρ_c can be determined automatically). Fig. 8 reports, for the same configuration, but for different values of r_c, the molecules belonging to the largest cluster (grey) and to the other, smaller clusters (blue). In addition, we have run an interfacial analysis using the Identification of Truly Interfacial Molecules algorithm (ITIM),[15] to inspect also the features of the interface between the high and low benzene concentration regions. The atoms belonging to the interface (but still part of the high benzene concentration region) are represented in red. Visual inspection shows that the DBSCAN-based algorithm succeeds in identifying consistently the two regions for all cutoff values, even when the distributions of neighbours are strongly overlapping (see Fig. 6). For cutoff values lower than 0.5 nm, however, the algorithm seems to identify a larger number of benzene molecules, with some protrusion into the ionic liquid-rich phase noticeable in the left interface.

To quantify the difference between the phases obtained with different cutoff values, as well as the differences between the corresponding interfaces, we calculated the number of molecules belonging to the largest cluster (*i.e.*, the benzene-rich phase) and the interface, as a function of r_c, and presented this information in Fig. 9. Two main trends can be seen, namely, a sharp decrease in the number of surface molecules and an increase in the number of molecules in the largest cluster when the cutoff increases from 0.35 to about 0.6 nm. After this point, the number of surface molecules and the size of the biggest cluster change only slightly, and the dependence on the cutoff value suggests the presence of a minimum value for the number of surface molecules. Intuitively, one would consider as an optimal choice for r_c the one that minimizes the surface area[7] or, in this case, the number of surface molecules, which is roughly proportional to it. In principle, one could point out that from r_c = 1 nm up to the largest value r_c = 1.45 nm, the number of surface molecules differs, on average, only by about 5 surface molecules (roughly 1% of the total). These small variations are probably not consequential for many practical applications, and any cutoff radius above 1 nm would be an equivalently good choice. One should bear in mind, however, that the larger the cutoff radius, the more computationally intensive the DBSCAN filtering will be.

The solubility of benzene, defined as the ratio of the benzene molar concentration and total molar concentration in the ionic-liquid rich phase, can be accessed either by counting the number of benzene molecules not in the largest cluster, or through a fit in the bulk region of the density profile. The two values (0.70 ± 0.01 and 0.69 ± 0.01, respectively) compare reasonably well with the experimental solubility of benzene in BMIM NTf_2, 0.82 measured under the same thermodynamic conditions.[10]

It is worth mentioning that the algorithm is not restricted to planar phase separations, but can be applied without any modification in the presence of corrugated interfaces, or when one of the phases is dispersed into the other one, such as droplets of oil dispersed in a water/oil mixture, as long as the cutoff radius r_c is smaller than that of the droplets.

One might wonder how other similar algorithms compare to the present one in determining the phases of a partially miscible system, or, conversely whether the present approach is appropriate also in the fully demixed case. To test similar algorithms we investigated the BMIM NTf_2/benzene system using the ITIM algorithm with simple clustering, and the Generalized ITIM (GITIM) algorithm.[16] Simulation snapshots resulting from these tests are shown in the ESI.† In particular, it clearly emerged that the use of a simple clustering scheme does not allow us to recognize properly the two phases, to the extent that it is impossible for the ITIM algorithm to distinguish upper and

lower sides of the interface. With the simple clustering scheme, all molecules in the simulation box are assigned to the high-density phase, and those tagged as interfacial are not representative at all of the real surface molecules. The GITIM algorithm performs better, as it is able to reveal the small empty pockets, which are abundant in the low benzene concentration phase. The large majority of benzene molecules in the latter phase, however, are identified as surface molecules, and some additional kind of filtering would be required to properly separate the two phases.

The opposite limit of a well-defined phase separation with vanishing vapour pressure is also instructive. We have considered a water/vapour interface formed by 1000 water molecules at 300 K, simulated using the SPC/E water model,[17] and assigned molecules to the two phases using both the simple clustering scheme and the present method. At a temperature of 300 K, in most frames not a single water molecule is clearly seen in the vapour phase, and in these cases the simple clustering approach assigns all water molecules to the liquid phase. The distribution of the number of neighbours, however, is in this case no longer bimodal, which was a prerequisite for using the automatic determination of the threshold density ρ_c. Setting by hand ρ_c to, for example, one water molecule every 35 Å^3 (therefore slightly lower than the bulk number density), the density based clustering algorithm also assigns all molecules to the liquid phase. If, instead, the upper centroid is used to set ρ_c, the following is observed: for a cutoff value of 0.8 nm, some molecules in the outer part of the first molecular layers are identified as vapour ones; by increasing the cutoff value the number of molecules recognized as vapour ones decreases, and for cutoff values larger than 1.5 nm, they completely disappear (simulation snapshots of the water/vapour interface are also shown in the ESI†).

We can therefore conclude that in the extreme case when no or very few molecules are present in the low-concentration phase, the automatic threshold determination should be avoided, in favour of either choosing manually the threshold density, or using the simple clustering method. Either way, the same result is obtained.

4 Methods

Molecular dynamics simulations of the BMIM NTf_2/benzene system were carried out with the GROMACS 4.6.7 program package on the isobaric–isothermal ensemble under ambient conditions (at 298.15 K and 1 bar).[18] During simulations, the cross-sectional area of the rectangular simulation box was kept fixed (5.7 nm × 5.7 nm) and only the direction perpendicular to the plane of the interface was allowed to change (fluctuating around the value of 18.51 nm) according to the prescribed pressure. To simulate the isobaric–isothermal ensemble we applied the Nosé–Hoover thermostat[19,20] in combination with the Parrinello–Rahman barostat, with relaxation constants of 1.0 ps and 2.0 ps, respectively.[21]

To model BMIM NTf_2, the force field published by Logotheti *et al.*[22] was applied, while benzene was described by the general OPLS force field, including all hydrogen atoms.[23] The Lennard-Jones interaction parameters between the ionic liquid and benzene were determined using Lorentz–Berthelot mixing rules. All bond lengths were kept fixed using the LINCS algorithm[24] to make sure we can safely use a time step of 2 fs for the integration of the equations of motion. Non-bonded interactions were truncated beyond 1.2 nm, while the electrostatic interaction was calculated using the PME algorithm[25] in its smooth variant.[26]

Analytical tail corrections for the energy and pressure of the dispersion interactions were also applied. Periodic boundary conditions were applied in all directions. The simulated system contained 512 ion pairs as well as 2500 benzene molecules. Starting from totally unmixed phases, an equilibration was performed for 600 ns. This long equilibration time was needed because of the peculiar mixing behaviour of the system (which requires transport of a large amount of mass through the simulation box), combined with the high viscosity of the mixture, which sets the relaxation time. During the 600 ns equilibration, quantities like the total energy and the concentrations in the two phases (that is, also the chemical potential) relaxed to their equilibrium values. After the equilibration, we sampled configurations from an additional 30 ns long run, saving them to disk for further analysis.

The algorithm for the phase determination was implemented in python/Cython,[27] only slightly modifying the DBSCAN class of the scikit-learn machine learning package[28] to provide also the sizes of clusters and to take into account the presence of periodic boundary conditions. The latter requirement has been fulfilled by using the periodic boundary condition-aware kd-tree neighbour search implemented in recent versions ($\geq$0.18) of scipy. The configurations from the stored trajectories were made accessible to the python script using the MDAnalysis library.[29,30]

5 Conclusions

We have presented a computational approach that is able to tell apart molecules belonging to different phases in solutions with high mutual solubility. The method is based on the DBSCAN clustering algorithm, and introduces an automatic choice for the threshold concentration, which is one of the two parameters needed by DBSCAN. The approach has been shown to be robust against the choice of the remaining input parameter, namely, the distance cutoff used to calculate the local concentration. We have applied the method to a partially miscible solution of an ionic liquid (BMIM NTf_2) with benzene.

This choice has been dictated by two factors, namely (a) the high miscibility of benzene, which yields equilibrium concentrations in a ratio of about 3 : 7 and (b) the virtually zero miscibility of the ionic liquid. On the one hand, the high miscibility of benzene provided a challenging test case, as the equilibrium concentrations are very close to each other, and the interface between the high and low benzene concentration regions is difficult to locate. On the other hand, the low miscibility of the ionic liquid has, as a consequence, the presence of a very

well defined interface between the ionic liquid-rich and poor regions. This provides a remarkable benchmark for the identification of the interface, as the phases as defined using the ionic liquid or benzene are not supposed to interpenetrate in a consistent scheme. In fact, only about 1 ion pair is found, on average, to cross the interface and be found in the high benzene concentration region. This shows that the algorithm is in fact able to identify a meaningful interface also when the miscibility of the components is very high. The robustness against the only remaining free parameter of the method, and the quasi linear scaling as a function of the number of atoms in the system, therefore make it a viable approach for the identification of phases in multicomponent systems with high miscibilities. The code is available free of charge as part of the pytim package at https://github.com/Marcello-Sega/pytim.

Acknowledgements

This work was supported by the Hungarian NKFIH Foundation under Project No. 119732. We thank Joe Donaldson for carefully reading the manuscript.

References

1 I. Benjamin, *Chem. Rev.*, 1996, **96**, 1449–1476.
2 I. Benjamin, *Annu. Rev. Phys. Chem.*, 1997, **48**, 407–451.
3 Q. Du, R. Superfine, E. Freysz and Y. Shen, *Phys. Rev. Lett.*, 1993, **70**, 2313.
4 M. Sega, B. Fabian and P. Jedlovszky, *J. Chem. Phys.*, 2015, **143**, 114709.
5 J. S. Rowlinson and B. Widom, *Molecular Theory of Capillarity*, Oxford University Press, Oxford, 1982.
6 S. Senapati and M. L. Berkowitz, *Phys. Rev. Lett.*, 2001, **87**, 176101.
7 E. Chacón and P. Tarazona, *Phys. Rev. Lett.*, 2003, **91**, 166103.
8 L. B. Pártay, G. Horvai and P. Jedlovszky, *J. Phys. Chem. C*, 2010, **114**, 21681–21693.
9 P. Tarazona and E. Chacón, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2004, **70**, 1–13.
10 A. Arce, M. J. Earle, H. Rodrguez and K. R. Seddon, *J. Phys. Chem. B*, 2007, **111**, 4732–4736.
11 G. Hantal, M. Sega, S. S. Kantorovich, C. Schröder and M. Jorge, *J. Phys. Chem. C*, 2015, **119**, 28448–28461.
12 M. Ester, H. P. Kriegel, J. Sander and X. Xu, Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 1996, pp. 226–231.
13 J. L. Bentley, *Commun. ACM*, 1975, **18**, 509–517.
14 J. MacQueen, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967, pp. 281–297.
15 L. B. Pártay, G. Hantal, P. Jedlovszky, Á. Vincze and G. Horvai, *J. Comput. Chem.*, 2008, **29**, 945–956.
16 M. Sega, S. S. Kantorovich, P. Jedlovszky and M. Jorge, *J. Chem. Phys.*, 2013, **138**, 044110.
17 H. Berendsen, J. Grigera and T. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
18 D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
19 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
20 W. G. Hoover, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1985, **31**, 1695.
21 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
22 G.-E. Logotheti, J. Ramos and I. G. Economou, *J. Phys. Chem. B*, 2009, **113**, 7211–7224.
23 W. L. Jorgensen and J. Tirado-Rives, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6665–6670.
24 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
25 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
26 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
27 S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn and K. Smith, *Comput. Sci. Eng.*, 2011, **13**, 31–39.
28 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
29 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, **32**, 2319–2327.
30 R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney and O. Beckstein, Proceedings of the 15th Python in Science Conference, 2016, pp. 98–105.