Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

A molecular contact theory for simulating polarization: application to dielectric constant prediction

Théophile Gaudin,[*a] and Haibo Ma[*a]

[a]*Key Laboratory of Mesoscopic Chemistry of MOE, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, 210023 China.*

e-mail: gaudin.theophile@gmail.com; haibo@nju.edu.cn

SUPPORTING INFORMATION

1 Technical details about numerical simulation of molecular contacts and dimers

As stated in Section 1 of the main text, the only available computational method to deal with contacts between molecular surfaces up to now is the COSMO-RS method. In this method, the surface is considered as a virtually continuous entity and is only discretized for numerical purposes. As a consequence, contact probabilities between points on the surface of the molecule can be calculated up to virtually infinitesimal dimensions. However, real contacts between molecules should match their coordination number. Within COSMO-RS theory, this is achieved by introducing a universal parameter called effective contact area, a_{eff} (of about 7 Å^2 with some variations depending on the parameterisation[1]), into the statistical mechanical equations. For example, in the case of water, this leads to a coordination number of about 6 based on the surface area of water's COSMO surface (43 Å^2). Though this is not exactly equivalent to the real average coordination number which is slightly above 4 in liquid water[2], this number is of acceptable magnitude notably owing to the fact that all contacts are averaged into a single area. Nevertheless, a direct computation of contact probabilities between two segments from their properties and chemical potentials ultimately only covers their small surface area, which is just an artefact of the discretization of the used molecular surface. For nonpolar or moderately polar surface segments, this is not of huge importance as contacts centered on a pair of points of two molecular surfaces may be reasonably viewed as noncorrelated and thus one may physically interpret them as the probability of a contact of area a_{eff} centered at the contact point between the two (much smaller) segments.

However, for two surface segments of high and opposite polarity (i. e., typically, hydrogen-bonding contacts), the contact probability may be so high that two neighbouring contact segments will frequently be part of the same physical contact (i. e. the correlation between the contact probabilities of two neighboring segments is strong). In this case, some types of contacts are partially excluded from consideration due to the locking effect produced by the strong interaction. Thus, the number of contacts which are not constrained by the hydrogen-bonding scheme will be lower than what would be expected by just considering all possible segment pairs. Since it has been found early that for organic liquids, the polarization effect was low without hydrogen bonding[3], in this study, the choice of neglecting the contributions of the non-hydrogen bonding contacts in the partition function was made. Thus, for any system without hydrogen bonding, $\langle\cos\theta\rangle(s,t)$ ended up being zero by construction.

Moreover, to account for the correlation of contact probabilities related to hydrogen bonding (HB) contacts, a mean field equalization of contact probabilities for each pair of hydrogen donor (HD) and hydrogen acceptor (HA) conceivable for a given mixture was carried out. This required to locate the HA and HD hotspots. In this work, this was done in an automated, but supervised way. The atoms bearing HD and HA groups were identified from the surface charge densities of their above segments and the amount of their surface which was covered in HD or HA segments. In the case of HA, the donor pair geometries were identified using simple VSEPR concepts (e. g. sp^2 oxygen for a C=O group). In a few exceptions (such as acetonitrile) this simple approach misclassified the molecule as being able to donate HB. These exceptions were handled manually. In a future study, the identification of HD and HA hotspots from molecular surfaces will be scrutinized in further detail, but this approach was sufficient for the needs of the present work.

To carry out the mean field equalization, the basic concept explored in this study was to:

- sum every HB contact probability and predict the probability of every HD to interact with every HA (and vice versa);
- replace the individual segment pair contact probabilities by the corresponding HD/HA overall contact probabilities;
- renormalize the full set of probabilities.

In order to model dimerization, we devised an algorithm as follows. First, we computed the angle between all HD and HA pairs situated in the same structure. If it was below a convenient threshold close to parallel, we performed a rough collision test to eliminate dimers that would result in pair interpenetration by computing the distance between the two normal planes associated to the HD/HA pair candidate. If it was below a certain threshold, we assumed that dimerization was possible. For each dimer, the probability of a contact to be a dimerizing one was the sum of all probabilities of the segment to interact with other dimerizing segments of opposite polarity. In the special case of dimerizable conformations of aliphatic acids, guided by the experimental data (even butyric acid has ε_r of 3, close to that of nonpolar alkanes) we considered that the correlation of contacts imposed that these conformations were dimerized to an extent multiplied by the amount of dimerizable acids in the mixture. Indeed, this is likely to occur because it provides an optimal shielding of polar zones from a generally nonpolar environment.

2 Appendix A. Relationship between polarization and electric field

Gauss Law in differential form can be expressed in both following forms, one for both free charges (i. e. charges which do not depend on the presence of any dipole) and bound charges and the other for bound charges only:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} = \frac{\rho_f + \rho_b}{\varepsilon_0} \qquad (S1)$$

$$\nabla \cdot \mathbf{P} = -\rho_b \qquad (S2)$$

where $\mathbf{E}$ is the electric field, ρ is the charge density, ε_0 is the dielectric constant of vacuum ρ_f is the density of free charges, ρ_b is the density of bound charges, i. e. of the charges generated in a polarizable continuum by the presence of a dipole, and $\mathbf{P}$ is the polarization (i. e., the density of dipole moment vectors).

Combining eqs S1 and S2 and rearranging, we get:

$$\nabla \cdot \mathbf{P} = \rho_f - \varepsilon_0 \nabla \cdot \mathbf{E} \qquad (S3)$$

The Gauss law written for free charges only enables to *define* (constitutive equation) the permittivity ε of the medium producing the polarization under study as:

$$\nabla \cdot (\varepsilon \mathbf{E}) = \rho_f \qquad (S4)$$

The product between the permittivity and the gradient of the electric field is called the electric displacement field **D** in electrostatics textbooks but this notion is not important in the present context and it thus skipped.

Combining eqs. S3 and S4 leads to:

$$\nabla \cdot \mathbf{P} = \nabla \cdot \left[(\varepsilon - \varepsilon_0) \mathbf{E} \right] \text{ or } \mathbf{P} = (\varepsilon - \varepsilon_0) \mathbf{E} \qquad (S5)$$

Inserting the definition of relative permittivity $\varepsilon_r = \varepsilon/\varepsilon_0$ into eq. S5 leads to:

$$\mathbf{P} = (\varepsilon_r - 1) \varepsilon_0 \mathbf{E} \qquad (S6)$$

which is eq. 3 in the main text.

3 Appendix B. Low-field relationship between dipole moment and electric field

The power series of a function is:

$$f(x)=\sum_{i=0}^{N}\frac{f^{(n)}(c)}{n!}(x-c)^{i} \qquad (S7)$$

In most cases, c = 0, so the power series of a function is the Mc Laurin series:

$$f(x)=\sum_{i=0}^{N}\frac{f^{(n)}(0)}{n!}x^{i} \qquad (S8)$$

When applied to a function like this:

$$\mathbf{f}(E)=\frac{\sum_{i}\mathbf{m}_i\exp(k_i-l_i\mathbf{m}_i\mathbf{n}E)}{\sum_{i}\exp(k_i-l_i\mathbf{m}_i\mathbf{n}E)} \qquad (S9)$$

The value for E = 0 of this function is:

$$\mathbf{f}(0)=\frac{\sum_{i}\mathbf{m}_i\exp(k_i)}{\sum_{i}\exp(k_i)}=\langle\mathbf{m}\rangle \qquad (S10)$$

Using some algebraic identities, it can be shown that the value of the derivative eq. S10 for E = 0 is:

$$\mathbf{f}'(0)=\mathbf{n}\langle\mathbf{m}^2\rangle-l\mathbf{n}\langle\mathbf{m}\rangle\langle\mathbf{m}\rangle \qquad (S11)$$

Thus, to first power, we have:

$$\mathbf{f}(E)=\frac{\mathbf{f}(0)}{0!}E^0+\frac{\mathbf{f}'(0)}{1!}E^1=\langle\mathbf{m}\rangle+l\left(\langle\mathbf{m}^2\rangle-\langle\mathbf{m}\rangle^2\right)\mathbf{E} \qquad (S12)$$

The average value of dipole moment of any random sample of sufficiently large size under an electric field is computed from the Boltzmann weights of instantaneous configurations:

$$\langle\mathbf{m}\rangle^{E}=\frac{\sum_{i}\exp\left(-\frac{U_i-\mathbf{m}_i\cdot\mathbf{E}}{kT}\right)\mathbf{m}_i}{\sum_{i}\exp\left(-\frac{U_i-\mathbf{m}_i\cdot\mathbf{E}}{kT}\right)} \qquad (S13)$$

where U_i is the interaction energy of the i-th configuration, $\mathbf{m}_i$ is the dipole moment vector of this configuration, **E** is the external electric field, k is the Boltzmann constant, and T is the temperature.

Under low fields and realising that the dipole moment of our random sample at zero fields can be approximated as the null vector as there is no reason for it to be biased in any way in such circumstances, the first order Mc Laurin series of eq. S12 writes, applying eqs. S7-S11:

$$\langle\mathbf{m}\rangle^{E}\approx\frac{\langle m^2\rangle}{3kT}\mathbf{E} \qquad (S14)$$

which is eq. 4 of the main text. Note that in molecular dynamics simulations, the average of the dipole moment is not assumed to be zero because in such situation, an actual box is considered for measurement, which in practice will have a residual numerical dipole moment that should be subtracted as an artifact.

4 Appendix C. Spherical law of cosines and free rotation approximation

Let us assume an orthogonal contact between two molecular surfaces, which are at an azimuthal angle of rotation φ_k with each other (cf. Fig. S1). We know the normal vector of both contact points, and the underlying dipole moment vector of the two molecules in contact. Thus, we can derive the angles between the dipole moment and the normal vector of both molecules (θ_1 and θ_2). The spherical law of cosines allows to calculate the angle between the two dipole moment vectors for this contact and this azimuthal angle:

$$\cos\theta_{12}^{k} = \cos\varphi_k \sin\theta_1 \sin\left(\pi-\theta_2\right) + \cos\theta_1 \cos\left(\pi-\theta_2\right) \qquad (S15)$$

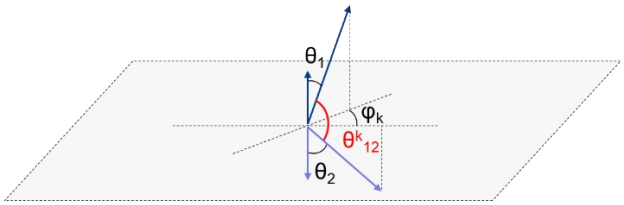

Fig. S1 Geometric depiction of a contact between two molecular surfaces

Assuming that each azimuthal angle has an equal probability (i. e. the free rotation assumption), the average angle for a contact over all azimuthal angles can be calculated as:

$$\left\langle\cos\theta_{12}\right\rangle = \frac{\sum_{k=-\pi}^{\pi}\left[\cos\varphi_k \sin\theta_1 \sin\left(\pi-\theta_2\right) + \cos\theta_1 \cos\left(\pi-\theta_2\right)\right]}{N_{total}} \qquad (S16)$$

where N_{total} is the total number of azimuthal angles which are evaluated (sampled regularly between $-\pi$ and π). This number should be large enough that the result is negligibly different from a continuous evaluation over all azimuthal angles.

Reminding the mathematical property $\cos x = -\cos(\pi - x)$ and realising that the azimuthal angle is defined from $-\pi$ to π, every $\cos\varphi_k$ value is cancelled by an exact opposite and eq. S16 simplifies to:

$$\left\langle\cos\theta_{12}\right\rangle = -\cos\theta_1 \cos\theta_2 \qquad (S17)$$

which is eq. 15 of the main text.

5 Computation of normal vectors

Since the above theory requires to calculate angles between the dipole moment of a structure and the normal vectors of segments, the normal vector has to be calculated. A large set of approaches exist to approximate surface normal on arbitrary surfaces. In this work, a simple approach, based on Singular Value Decomposition (SVD), was implemented. For each segment, a SVD is applied on the position vectors of the ten nearest neighbors. First, the position vectors are arranged into a 10×3 rectangular matrix **X**. This matrix is decomposed using SVD[38] as:

$$\mathbf{X} = \mathbf{U\Sigma V} \qquad (S18)$$

The normal vector is taken as the normalized third column of the 3x3 **V** matrix. The normal vector is then replaced by its negative if it points more towards the nearest atom than away from it. The approach works satisfactorily, as illustrated in Fig. S2 for water. Therefore, this technique is recommended if normal vectors on molecular surfaces are required in any future applications.

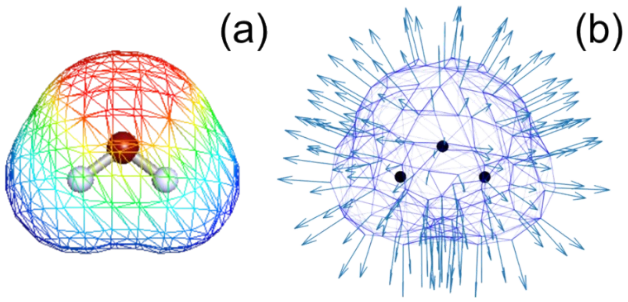

Fig. S2 (a) COSMO surface of water, (b) normal vectors (blue arrows) computed from COSMO surface coordinates using SVD.

References

1. A. Klamt, *COSMO-RS, From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier Science Ltd., Amsterdam, The Netherlands, 2005.
2. S. Swaminathan and D. L. Beveridge, *J. Am. Chem. Soc.*, 1977, **99**, 8392-8398.
3. W. D. Kumler, *J. Am. Chem. Soc.*, 1935, **57**, 600-605.