A statistical model of hydrogen bond networks in liquid alcohols

Per Sillrén, Johan Bielecki, Johan Mattsson, Lars Börjesson, and Aleksandar Matic

Additional information on J. Chem. Phys.

A statistical model of hydrogen bond networks in liquid alcohols

Per Sillrén,[1,a)] Johan Bielecki,[1] Johan Mattsson,[2] Lars Börjesson,[1] and Aleksandar Matic[1]
[1]*Applied Physics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden*
[2]*School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, United Kingdom*

We here present a statistical model of hydrogen bond induced network structures in liquid alcohols. The model generalises the Andersson-Schulz-Flory chain model to allow also for branched structures. Two bonding probabilities are assigned to each hydroxyl group oxygen, where the first is the probability of a lone pair accepting an H-bond and the second is the probability that given this bond also the second lone pair is bonded. The average hydroxyl group cluster size, cluster size distribution, and the number of branches and leaves in the tree-like network clusters are directly determined from these probabilities. The applicability of the model is tested by comparison to cluster size distributions and bonding probabilities obtained from Monte Carlo simulations of the monoalcohols methanol, propanol, butanol, and propylene glycol monomethyl ether, the di-alcohol propylene glycol, and the tri-alcohol glycerol. We find that the tree model can reproduce the cluster size distributions and the bonding probabilities for both mono- and poly-alcohols, showing the branched nature of the OH-clusters in these liquids. Thus, this statistical model is a useful tool to better understand the structure of network forming hydrogen bonded liquids. The model can be applied to experimental data, allowing the topology of the clusters to be determined from such studies. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3690137]

I. INTRODUCTION

The intermolecular interaction in liquids containing hydroxyl (OH–) groups, such as alcohols, are largely influenced by strongly directional hydrogen (H–) bonds. The strength of H-bonds in alcohols is about 10 $k_B T$ at room temperature, leading to a transient network of H-bonds in the liquid, with bonds breaking and reforming on a timescale of $\tau \sim 10^{-11}$ s.[1] The strong influence of H-bonds on the properties of alcohols becomes evident when they are compared to their alkane analogues; the most obvious difference is perhaps the dramatic increase in freezing and boiling temperatures. In monoalcohols, H-bonds also lead to dynamics not generally observed in liquids, such as relaxations slower than the structural α relaxation, characterised by a Debye-peak observed in dielectric spectroscopy,[2] as well as a relaxation process acting on a time scale intermediate between the Debye and the α relaxation processes, as determined by the NMR-spectroscopy;[3] these relaxation processes are believed to be related to the dynamics of the transient H-bonded clusters.

In alcohols, each OH-group can participate in three H-bonds; the hydrogen can act as an H-bond donor and the two lone pairs as H-bond acceptors. Since the average number of bonds per OH-group, or average coordination number, $\langle n_b \rangle$ is always found to be less than 2 for alcohols,[4–8] it is often assumed that the OH-clusters are chains without significant branching. However, all finite clusters of size n, without loops, contain $n - 1$ bonds, regardless of branching, and thus $2(n - 1)/n < 2$ bonds per OH-group. Even though the coordination number is a quantitative description of the H-bond net-work, it gives no information on the cluster topology or cluster size distribution.

The structure of alcohols has been studied by experimental techniques, such as vibrational spectroscopy,[8–12] neutron-,[5,13] and x-ray diffraction,[6,7,14] as well as computational methods,[4,15] but the lack of an established quantitative model has made it difficult to reconcile the different results and to connect the structures to thermodynamics and dynamics. A family of models that has been used to describe the intermolecular structure of water, as well as other network-forming systems, are the "patchy particle" models,[16] wherein typically hard spherical particles are decorated with attractive patches of one or several kinds. Together with Wertheim's thermodynamic perturbation theory and the related statistical associating fluid theory (SAFT),[17,18] these models have been able to reproduce thermodynamic properties of liquid water,[19] methanol-water mixtures,[20] as well as pure alcohols.[21] These theories use as their starting point assumptions about topology of the network forming liquid, and the cluster size distribution and bonding probabilities can be used as input.[22] A better understanding of the OH-group cluster topology is thus needed, both to interpret experiments and to improve theories.

In this paper, we introduce a simple statistical model aimed at describing hydrogen bonding in alcohols. Each OH-oxygen has two lone pairs and thus two possibilities to bond to neighbouring OH-hydrogens. The probability of forming a first bond, and the conditional probability of forming a second bond given that a first bond already exists, constitute our two model parameters. In terms of these two bonding probabilities, we derive the statistics of the resulting tree-like clusters, including the cluster size distribution, mean cluster size, and the fraction of cluster roots, leaves, and internal hydroxyl groups (see Figure 1). Conversely, the bonding

a)Electronic mail: per.sillren@chalmers.se.

probabilities, and thus the topology of the OH-clusters, can be directly determined when at least two of the latter properties are known, either from simulations or from experimental data. We note that similar models, however, not based on conditional probabilities, have been used in the context of polymer science.[23,24] The formulation in terms of conditional probabilities is more relevant in the case of alcohols which have two identical bonding sites with equal *a priori* probabilities of forming an H-bond.

To investigate the relevance of the tree model, we apply it to configurations from Monte Carlo simulations of methanol, and 1-butanol, as well as a series of mono- and poly-alcohols, each with a three carbon backbone: 1-propanol, 1,2-propanediol (propylene glycol, PG), 1,2,3-propanetriol (glycerol), and 1-methoxy-2-propanol (propylene glycol monomethyl ether, PGME). We find that the tree model generally produces cluster size distributions and average cluster sizes consistent with the simulation data. However, we find clear deviations for the smallest cluster sizes for the monoalcohols. We attribute these deviations to entropy effects not accounted for by the model. It is important to point out that we study networks of OH-groups rather than clusters of molecules. Thus, for monoalcohols that only have one OH-group per molecule there is no difference, but for polyalcohols each molecule has several OH-groups and can thus partake in several clusters.

Hydrogen bonded systems have previously been analysed using a simpler model with only one bonding probability per unit site (OH-group),[8] known as the Anderson-Schulz-Flory (ASF) chain model.[25] The rationale for applying a model that only involves one bonding probability per site to alcohols which have two lone pairs per site, is the assumption that an OH-group with one bonded lone pair has a negligible probability of a second bonded lone pair. However, we show that this approximation is generally not sufficient and bonding also of the second lone pair has to be accounted for to properly characterise the clusters. In the following, we denote the nodes, the OH-groups, in the hydrogen bonded network according to the terminology of Graener et al.,[26] as described in Fig. 1. A root node, by definition, has one or two bonded lone-pairs and a non-bonded hydrogen. A leaf is a node that has no bonded lone-pairs, but a bonded hydrogen (γ). Internal nodes have either one or two bonded lone-pairs, and a bonded hydrogen (δ). A node that has neither its lone pairs nor its hydrogen bonded is a monomer (α). We start by outlining the statistics of the simpler ASF model before we derive the statistics of the branched tree model.

II. THE ANDERSON-SCHULZ-FLORY MODEL

The Anderson-Schulz-Flory model describes chains of particles (nodes) that bond to a neighbouring particle with a probability p_b; no branching or loops are accounted for in the model. The probability is connected to the average number of bonds per particle through
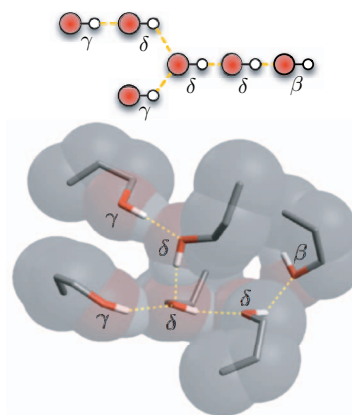
$$p_b = \frac{\langle n_b \rangle}{2}. \tag{1}$$

FIG. 1. A propanol cluster containing 6 hydrogen bonded OH-groups. The topology of the OH-cluster, highlighted in the top part, contains one root (β), two leaves (γ), and three internal OH-groups (δ).

A cluster of size n has $n - 1$ bonds, so the average number of bonds per particle is related to the average cluster size (chain length) through

$$\langle n_b \rangle = 2(1 - 1/\langle n \rangle), \tag{2}$$

so that

$$\langle n \rangle = \frac{1}{1 - p_b}. \tag{3}$$

The cluster size distribution $r(n)$, called the Anderson-Schulz-Flory distribution, is simply given by the probability of forming $n - 1$ consecutive bonds followed by a chain end:

$$r(n) = p_b^{n-1}(1 - p_b), \tag{4}$$

which is normalised so that $\sum_n r(n) = 1$. The mean and variance of the distribution can be calculated by realising that $nr(n) = p_b dr(n)/dp_b + p_b^{n-1}$ which gives

$$\langle n \rangle = \frac{1}{1 - p_b}, \tag{5}$$

as before, and

$$\langle n^2 \rangle - \langle n \rangle^2 = p_b \langle n \rangle^2. \tag{6}$$

We note that for $p_b \lesssim 1$, which has been reported to be the case for alcohols,[1,27–29] this distribution is rather wide, with a standard deviation approximately equal to the mean.

The cluster size distribution $r(n)$ gives the probability that a randomly chosen *cluster* has size n. From $r(n)$ we also get the often more relevant weighted cluster size distribution $R(n)$, which gives the probability that a randomly chosen OH-group belongs to a cluster of size n:

$$R(n) = \frac{nr(n)}{\sum_n nr(n)} = \frac{nr(n)}{\langle n \rangle}. \tag{7}$$

To characterise the clusters, we can determine the fraction of roots (including monomers), $n_{\alpha/\beta}$, leaves (excluding monomers), n_γ, and internal OH-groups, n_δ, where α are monomers, β are chain ends with non-bonded hydrogens, γ are chain ends with non-bonded oxygens, and δ represents internal, doubly bonded OH-groups.[26] Since there is exactly

one α/β per cluster, $n_{\alpha/\beta}$ is just the total number of clusters divided by the total number of OH-groups,

$$n_{\alpha/\beta} = \frac{\sum_n r(n)}{\sum_n n r(n)} = \frac{1}{\langle n \rangle} = 1 - p_b. \quad (8)$$

The same is true for n_γ, except that we have to subtract the monomers,

$$n_\gamma = \frac{1 - r(1)}{\langle n \rangle} = \frac{p_b}{\langle n \rangle} = p_b - p_b^2. \quad (9)$$

All OH-groups that are not chain ends are internal OH-groups, so that

$$n_\delta = 1 - (1 - p_b) - (p_b - p_b^2) = p_b^2. \quad (10)$$

III. A STATISTICAL MODEL OF BRANCHED HYDROGEN BOND NETWORKS

The ASF chain model does not account for branching, which makes it incomplete with respect to describing the hydrogen bonded networks of alcohols, where each OH-group can be H-bonded to three neighbouring OH-groups. To address this, we generalise the ASF chain model by allowing for a third H-bond on each OH-group and derive the properties of such a model.

Instead of the single bonding probability p_b, used above, we define a probability $p_A = p(O_A)$ for accepting a first H-bond from a neighbouring molecule and a probability $p_B = p(O_B|O_A)$ for accepting a second H-bond given that a first bond already exists. The unconditional probability that a second bond exists is given by Bayes' rule,

$$p(O_B) = \frac{p(O_B|O_A)p(O_A)}{p(O_A|O_B)} = p_A p_B. \quad (11)$$

The probability that the hydrogen is bonded, i.e., that the OH-group is donating a hydrogen bond, is given by the addition law of probability,

$$p_H = p(O_A) + p(O_B) - p(O_A \cap O_B) = p_A + p_A p_B, \quad (12)$$

where $p(O_A \cap O_B) = 0$ denotes the probability that the hydrogen is bonded to both an A-site and a B-site. Setting this to zero means that we ignore the possibility of bifurcated H-bonds, i.e., hydrogens bonded to more than one oxygen. For alcohols, such an approximation is motivated since bifurcated H-bonds are found to be very rare.[1,5] The average number of bonds per OH-group $\langle n_b \rangle$ is

$$\langle n_b \rangle = p_H + p_A + p_A p_B$$
$$= 2(p_A + p_A p_B), \quad (13)$$

which clearly, shows that the average coordination number cannot be greater than 2. In analogy with Eq. (3), the average cluster size becomes

$$\langle n \rangle = \frac{1}{1 - p_A - p_A p_B}. \quad (14)$$

The cluster size distribution can be written either in a closed form or in a recursive formulation. The latter starts

from $r(1) = 1 - p_A$, and for $n \geq 2$ we have

$$r(n) = p_A(1 - p_B)r(n - 1) + p_A p_B \sum_{k=1}^{n-2} r(k)r(n - k - 1). \quad (15)$$

The closed form on the other hand is

$$r(n) = \sum_{k=0}^{\infty} a_{n,k} p_A^{n-k-1} (1 - p_B)^{n-2k-1} p_B^k (1 - p_A)^{k+1}, \quad (16)$$

where $a_{n,k} = \frac{1}{n}\binom{n}{k+1}\binom{n-k-1}{k}$ denotes the number of different trees with n nodes and k branches, see Appendix for further details.

The second and third moments, $\langle n^2 \rangle$, and $\langle n^3 \rangle$, of the cluster size distribution, $r(n)$, can be written in terms of the mean, $\langle n \rangle$,

$$\langle n^2 \rangle = -\langle n \rangle + 2(1 - p_A)\langle n \rangle^3, \quad (17)$$

$$\langle n^3 \rangle = \langle n \rangle - 6(1 - p_A)(\langle n \rangle^4 + \langle n \rangle^3)$$
$$+ 12(1 - p_A)^2 \langle n \rangle^5, \quad (18)$$

details are again found in the Appendix. The variance is then

$$\langle n^2 \rangle - \langle n \rangle^2 = -\langle n \rangle - \langle n \rangle^2 + 2(1 - p_A)\langle n \rangle^3. \quad (19)$$

As for the chain model, the weighted cluster size distribution, $R(n)$, is defined by Eq. (7). The mean and variance of $R(n)$, can be written in terms of the moments of $r(n)$. The mean is

$$\langle n \rangle_R = \frac{\langle n^2 \rangle}{\langle n \rangle}, \quad (20)$$

and the variance is

$$\langle n^2 \rangle_R - \langle n \rangle_R^2 = \frac{\langle n^3 \rangle}{\langle n \rangle} - \left(\frac{\langle n^2 \rangle}{\langle n \rangle}\right)^2. \quad (21)$$

To characterise the topology of the clusters, it is useful to know the fraction of OH-groups that have non-bonded hydrogens, i.e., roots of the tree structures or α/β groups, and the fraction of OH-groups that have a non-bonded oxygen, i.e., leaves of the tree structures, or γ groups. The fraction of roots, $n_{\alpha/\beta}$, including monomers, is easy to find since there are exactly 1 root per cluster so that

$$n_{\alpha/\beta} = \frac{\sum r(n)}{\sum n r(n)} = \frac{1}{\langle n \rangle}. \quad (22)$$

The number of leaves n_{leaves} are always one more than the number of branches, so we would like to calculate

$$n_{\text{leaves}} = \frac{1}{\sum_n n r(n)} \sum_n \sum_k (k + 1)y_{n,k}, \quad (23)$$

where the shorthand $y_{n,k}$ for the individual terms in the sum (16) is used. This sum can be evaluated with the help of Eq. (A10),

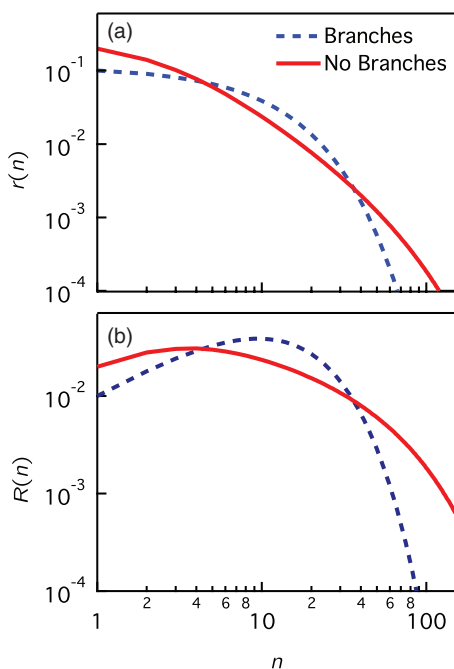$$n_{\text{leaves}} = (1 - p_A). \quad (24)$$

FIG. 2. (a) Two cluster size distributions, $r(n)$, with equal mean, $\langle n \rangle = 10$, but different branching probabilities, $p_B = 0$ (dashed line), and $p_B = 1/8$ (full line). (b) The corresponding weighted cluster size distributions, $R(n)$.

By subtracting the fraction of OH-groups which are monomers, $r(1)/\langle n \rangle = (1 - p_A)/\langle n \rangle$, we get the fraction of γ-groups.

$$n_\gamma = \frac{1}{\langle n \rangle}(1 - p_A)(\langle n \rangle - 1). \qquad (25)$$

To illustrate the role of branching in the distribution functions, two distributions are compared in Figure 2. Both distributions have the same mean cluster size $\langle n \rangle = 10$, but different branching probabilities, $p_B = 0$, and $p_B = 1/8$, respectively. We note that the distributions for clusters with more branching have more small and large, but fewer intermediate sized clusters.

IV. MONTE CARLO SIMULATION DETAILS

To test the relevance of the tree model to H-bond networks in alcohols, we perform Monte Carlo (MC) simulations using the optimised potentials for liquid simulations - all atoms (OPLS-AA) potential model, however, without dihedral restraints. The OPLS-AA model has been designed to reproduce a number of physical properties, such as the density, radial distribution functions, and heat of vaporisation of organic liquids.[30] Molecular dynamics simulations using the closely related optimised potentials for liquid simulations - united atoms (OPLS-UA model have been carried out previously for some alcohols.[1,31] The difference between OPLS-AA and OPLS-UA is that in the AA model, all atoms are treated explicitly, while in the UA model, CH_n groups are treated as one unit to save time.

The OPLS-AA model is based on a pairwise additive potential where the pair interaction between molecules i and j are given by a Coulomb term plus a Lennard-Jones term

$$v_{ij}(\mathbf{r}) = \sum_k^{\text{on } i} \sum_l^{\text{on } j} \frac{q_k q_l}{4\pi\varepsilon_0 r_{kl}} + 4\varepsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{12} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^6\right], \qquad (26)$$

where ε_0 is the vacuum permittivity, q_k is the (partial) charge on atom k, and r_{kl} is the distance between atoms k and l. The total potential energy is given by the sum over all molecules

$$V(\mathbf{r}^N) = \sum_i \sum_{j>i} v_{ij}(\mathbf{r}). \qquad (27)$$

The intramolecular energy comes from harmonic potentials with bond lengths and bond angles as their main input:

$$V_{\text{intra}} = \frac{C}{2} \sum_i \sum_{k \neq l} \frac{(r_{kl} - d_{kl})^2}{w_{kl}^2}, \qquad (28)$$

where

$$w_{kl}^2 = \frac{d_{kl}}{\sqrt{\mu_{kl}}} = d_{kl}\sqrt{\frac{M_k + M_l}{M_k M_l}}, \qquad (29)$$

with d_{kl} being the equilibrium distance between atom k and l, and M_k is the mass of atom k. A good value of $C/2$, determined from matching simulation data with diffraction data,[32] is 65 J $\text{Å}^{-1}(\text{amu})^{-1/2}$.

MC simulations were performed at T = 300 K for methanol, 1-propanol, PG, glycerol, PGME, and 1-butanol using the empirical potential structure refinement (EPSR) software.[32] For each liquid, we simulated 500 molecules under constant *NVT* (number of molecules, volume, temperature) in boxes with periodic boundary conditions. 10 000 MC-cycles were run to equilibrate the systems, whereafter cluster size distributions were calculated every five MC-cycle for another 10 000 cycles; the "clusters" routine provided in the EPSR software was used. Hydrogen bonds were defined by requiring the H–O distance to be between 1.4 and 2.4 Å.[5]

V. RESULTS AND DISCUSSION

An example of a propanol cluster obtained from the MC simulation is shown in Figure 1. This particular cluster has 6 hydrogen bonded hydroxyl groups, two being leaves (γ-groups), three internal OH groups (δ-groups), and one root (β-group). It is clear from inspection of the MC-configurations that branched clusters are present for all investigated liquids. It is thus physically relevant to use a model that allows for branched, tree-like clusters to describe the topology of the networks.

Cluster size distributions, $r(n)$, and weighted cluster size distributions, $R(n)$, determined from the analysis of the MC-configurations are shown in Figure 3. Methanol has the widest distribution, as well as the largest average cluster size, $\langle n \rangle = 5.9$, while PGME has the narrowest distribution as well as the smallest average cluster size $\langle n \rangle = 1.4$. This could be attributed to the fact that methanol has the least steric hindrance, enabling it to easily form hydrogen bonds, while PGME's bulkiness hinders hydrogen bonding. Propanol ($\langle n \rangle = 3.0$), PG ($\langle n \rangle = 2.6$), glycerol ($\langle n \rangle = 3.1$), and butanol ($\langle n \rangle$

= 2.89) which are more bulky than methanol but less than PGME, have average cluster sizes intermediate between those of methanol and PGME. Generally, we find average cluster sizes that are in qualitative agreement with previous simulation and experimental work.[1,27,28,33]

Cluster size distributions for the tree model were calculated with the requirement that they have the same mean and variance as the cluster size distributions from the simulations. This condition, combined with Eqs. (14) and (19), uniquely determines the bonding probabilities, p_A and p_B, and the cluster size distributions $r(n)$ and $R(n)$. Table I lists the average cluster sizes $\langle n \rangle$, and the mean cluster sizes, $\langle n \rangle_R$, that a randomly chosen OH-group belongs to, i.e., the mean of $R(n)$, determined from the MC-configurations together with bonding probabilities, p_A and p_B, obtained from the tree model distributions as well as directly from the simulated configurations.

Figure 3 compares the cluster size distributions, $r(n)$, and the weighted cluster size distributions, $R(n)$, for the tree model, the chain model, and the MC-simulations, respectively. Logarithmic scales are used to highlight and compare the distributions over the large range of cluster sizes, n, present. Figure 4 shows the same weighted cluster size distributions as in Figure 3, but on a linear scale; this highlights

TABLE I. Average cluster sizes, $\langle n \rangle$, average sizes of the cluster a randomly chosen OH-group belongs to, $\langle R(n) \rangle$, and bonding probabilities p_A and p_B obtained by requiring the model distributions to have the same mean and variance as the cluster size distributions from the simulations. The values of p_A and p_B calculated directly from the simulations are shown within parentheses, and were obtained by counting the fraction of oxygens with one ($f_1 = p_A(1 - p_B)$) or two ($f_2 = p_A p_B$) H-bonds.

	Methanol	Propanol	PG	Glycerol	PGME	Butanol
$\langle n \rangle$	5.9	3.0	2.6	3.1	1.4	2.89
$\langle n \rangle_R$	17	5.5	5.7	8.2	2.3	5.12
p_A	0.730	0.642	0.529	0.521	0.240	0.633
	(0.719)	(0.620)	(0.553)	(0.569)	(0.254)	(0.611)
p_B	0.139	0.042	0.169	0.298	0.121	0.033
	(0.135)	(0.058)	(0.147)	(0.239)	(0.067)	(0.093)

the behaviour for the smallest cluster sizes, which contain the largest fraction of the molecules. Overall, we find a very good agreement between the network properties obtained from the MC-simulation and the results of the tree model, see Figure 3 and Table I. On the contrary, the chain model deviates considerably from the simulations, as clearly seen in Figure 3. It is interesting to note that the best agreement between the chain model and the simulations are given by propanol, butanol, and PGME. Considering the low p_B of propanol and butanol (see Table I) this is not surprising. PGME, on the other hand, has a $p_B > 0.1$ (Table I). However, also this result can be understood since the difference in the average cluster size $\langle n \rangle$ (and hence the variance, see Eq. (19)) between the chain and the tree model depends on p_B only through the product $p_A p_B$. A good agreement between the tree and the chain model is thus expected whenever the product $p_A p_B$ is small. From the fit to the simulations, $p_A p_B < 0.04$ is obtained for propanol, butanol, and PGME while $p_A p_B \sim 0.1$ is obtained for the other liquids. A chain model is thus expected to work considerably worse for the latter, just as observed.

The fact that the structure of alcohols are dominated by OH-bonded clusters is evident from the results of the tree model by a large probability for the OH-groups forming the first bond. The exception here is PGME, where p_A is rather low resulting in a large fraction of monomers. The branched topology of the OH-bonded clusters is revealed by a non-zero (3%–30%) probability for forming the second bond, p_B, for all liquids, even though it is significantly lower than p_A. Comparing the results for propanol, PG, and glycerol, that have identical carbon backbones and very similar average cluster sizes, it is found that the branching probability, p_B, increases with increasing number of OH-groups per molecule. This implies that the OH-clusters in propanol are mainly chains, while in PG and glycerol they are more branched. The cluster size distributions for propanol, PG, and glycerol are found to be very similar when we consider clusters of hydroxyl groups. If we instead consider clusters of molecules, the cluster sizes will be larger for PG and glycerol, since they have more than one OH-group per molecule, which could explain why some properties, such as viscosity, vary largely between these liquids even though the OH-clusters are very similar.

By using Eqs. (14) and (A1), we can also use the model to extract bonding probabilities from previous studies where
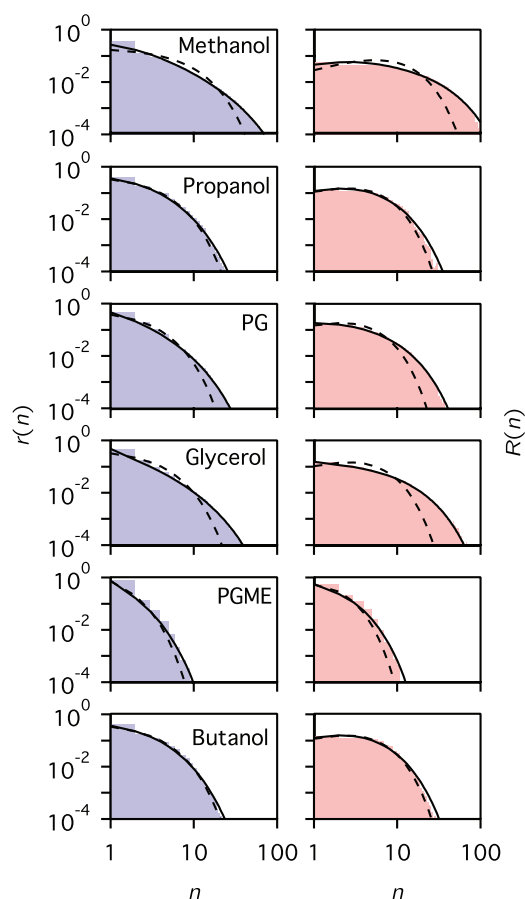

FIG. 3. Left: Cluster size distributions, $r(n)$. Right: Weighted cluster size distributions, $R(n)$. Solid lines show the distributions of the tree model obtained by requiring them to have the same mean and variance as the cluster size distributions from the simulations, which are shown as bars. Dashed lines show the ASF chain distributions with the same mean as the cluster size distributions from the simulations.

the average cluster size $\langle n \rangle$, and the fraction of monomers are given. In this fashion, we obtain for methanol $p_A = 0.56$ and $p_B = 0.15$[14] (x-ray, reverse Monte Carlo), $p_A = 0.78$ and $p_B = 0.019$[27] (MD), $p_A = 0.8$ and $p_B = 0.19$[1] (MD), $p_A = 0.86$ and $p_B = 0.08$ (269 and 290 K)[28] (MD), and $p_A = 0.79$ and $p_B = 0.09$ (339 K)[28] (MD). Even though the results are not directly comparable since a variety of definitions of H-bonds have been used, geometric as well as energetic, it is reassuring that they all confirm the branched nature of the clusters and that they are in qualitative agreement with our results concerning the bonding probabilities. We also note that our value for the average number of H-bonds per glycerol molecule, $3 \times (2 - 2/\langle n \rangle) \approx 4.06$, is intermediate between a previous molecular dynamics study[33] (3.25 H-bonds per molecule) and a more recent neutron diffraction study,[13] where the average number of bonds per molecule was estimated to be 5.68 ± 1.51.

Even though the tree model always reproduces the overall shape of the cluster size distributions there is a discrepancy between the model and the simulated cluster size distributions for the smallest cluster sizes, as shown in Fig. 4. The deviation is largest for methanol, where the number of monomers is significantly larger than the rest of the distribution suggests, and the number of clusters with 2–6 OH-groups are fewer compared to the model distribution. A similar, but less marked observation is made for propanol and butanol. For propanol and butanol, the tree model describes the distribution well for cluster sizes of $n > 3$, while for PG, glycerol, and PGME, the tree model works well over the whole cluster size range.
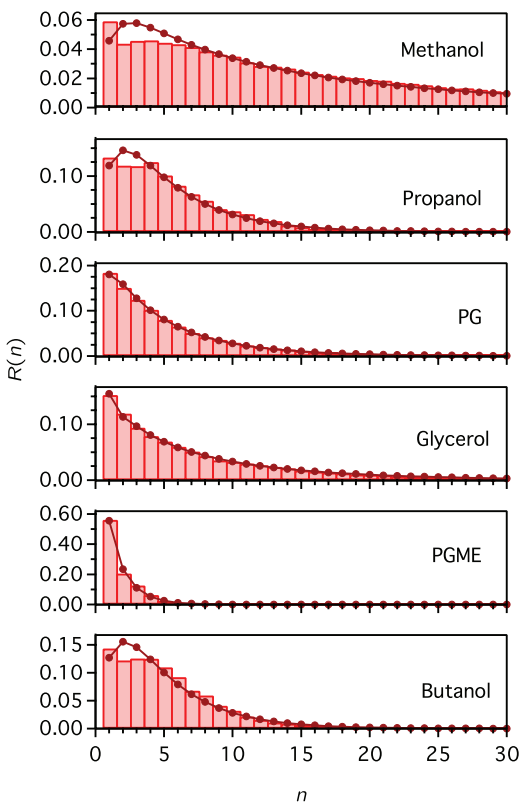

FIG. 4. Weighted cluster size distributions, $R(n)$, on a linear scale. Lines show weighted cluster size distributions obtained from the tree model by requiring them to have the same mean and variance as the cluster size distributions from the simulations.

The origin of this deviation can be understood through entropy arguments. When adding one OH-group to a cluster, the enthalpy, H, of the liquid decreases by an amount corresponding to the energy of the hydrogen bond formed. Also the entropy, S, of the liquid decreases when the bond is formed, leading to a corresponding positive term in the free energy change, $\Delta G = \Delta H - T\Delta S$. While the enthalpy change can be assumed to be largely independent of the size of the cluster, an OH-group joins, the entropy change will depend on the cluster size and be smaller for a larger cluster. Thus, since the tree model does not take these entropic effects into account, and assumes a cluster size independent change in free energy as a monomer joins a cluster and thus a cluster size independent bonding probability, the number of monomers is underestimated and the number of small clusters is overestimated. This argument could also explain why the discrepancy is smaller for the polyalcohols; Even if a polyalcohol molecule has one free OH-group, the other OH-groups of the molecule have a large probability, $1 - R(1)$, of being H-bonded, thus decreasing the translational/rotational entropy of the molecule.

To improve the model, one would need to include bonding probabilities that depend on the size of the cluster a particular OH-group joins, starting on an initial value for small cluster sizes and saturating at a value of n that depends on the particular liquid. A comparison of the simulated cluster size distributions and the tree model distributions for methanol indicates that the bonding probabilities become constant around $n \approx 7$, while for propanol and butanol they become constant already at $n \approx 4$. A model with bonding probabilities depending on the cluster size will certainly better reproduce the computer simulation data, but will also be significantly more complex.

VI. CONCLUSION

A simple statistical model starting from the hydrogen bonding possibilities in alcohols have been described and compared to Monte Carlo simulation data for five different alcohols. The model, based on two bonding probabilities for each OH-group, provides the possibility to form branched, tree-like, H-bonded clusters.

From a comparison to the structural configurations produced by MC-simulations, we find that the model can reproduce the overall cluster size distributions for both mono- and poly-alcohols. The amount of branching in the clusters rangers from 3% to 30%. A slight discrepancy between the tree model and the computer simulation data is found for small clusters in the monoalcohols, which can be attributed to entropic effects, implying cluster size dependent bonding probabilities.

The strength of the tree model lies in the straightforward application to experimental and computer simulation data. Knowing two independent quantities about the cluster size distribution, such as the average cluster size and the number of monomers, often reported in experiments and simulations, the topology of the OH-bonded clusters is given by the model. In addition, a possible application is to use the model in combination with Wertheim's perturbation theory, recently reformulated in terms of bonding probabilities, to calculate

the free energy of the liquid enabling a theoretical study of the structure dependence of thermodynamic properties such as heat capacities and compressibility. Another important application of the model could be as a tool to analyse vibrational OH-stretch spectra from Raman or IR measurements, for which the roots, leaves, and internal OH-group give rise to different bands.[26,34] The relative areas of these bands should thus, with aid of the model, enable determination of bonding probabilities and cluster sizes directly from such experiments.

APPENDIX: DERIVATION OF $r(n)$ AND ITS MOMENTS

In this Appendix, we give a derivation of the cluster size distribution and its moments. Let $r(n)$ denote the cluster size distribution, i.e., the probability that a randomly chosen cluster has size n. The probability of having a monomer, $r(1)$, is just the probability that the chosen OH-group is not bonded at all, i.e.,

$$r(1) = 1 - p_A. \tag{A1}$$

The probability for getting a cluster of size two is obtained by multiplying the probability of the first OH-group having exactly one H-bond, $p_A(1 - p_B)$, by the probability that the second OH-group has no further bonds

$$r(2) = p_A(1 - p_B)(1 - p_A) = p_A(1 - p_B)r(1). \tag{A2}$$

For clusters of size three, there are two options: a linear chain or a branched v-shaped cluster. The linear chain has probability $p_A(1 - p_B)r(2)$ while the branched cluster has probability $p_A p_B r^2(1)$ so that in total

$$r(3) = p_A(1 - p_B)r(2) + p_A p_B [r(1)]^2. \tag{A3}$$

Continuing in this fashion we get, for $n \geq 2$,

$$r(n) = p_A(1 - p_B)r(n - 1) + p_A p_B \sum_{k=1}^{n-2} r(k)r(n - k - 1), \tag{A4}$$

where the first term is the probability of a chain and the second term is the sum over all possible branched clusters of size n.

A closed expression for $r(n)$ can be obtained by noting that

$$a_{n,k} = \frac{1}{n}\binom{n}{k+1}\binom{n-k-1}{k} \tag{A5}$$

is the number of different trees with n nodes and k branches.[35] A given tree with n nodes and k branches must have $n - k - 1$ nodes with a A-bonds, k nodes with B-bonds, $k + 1$ nodes without A-bonds and $n - 2k - 1$ nodes without B bonds: The probability of getting a tree with n nodes and k branches is thus

$$r(n, k) = a_{n,k} p_A^{n-k-1}(1 - p_B)^{n-2k-1} p_B^k (1 - p_A)^{k+1}. \tag{A6}$$

The cluster size distribution is then obtained by summing over the number of branches

$$r(n) = \sum_{k=0}^{\infty} r(n, k). \tag{A7}$$

Differentiating $r(n, k)$ with respect to either p_A or p_B,

$$\frac{\partial r_{n,k}}{\partial p_B} = \frac{k}{p_B} r_{n,k} - \frac{n - 2k - 1}{(1 - p_B)} r_{n,k}, \tag{A8}$$

and then solving for $kr(n, k)$

$$kr(n, k) = \frac{p_B(1 - p_B)}{1 + p_B} \frac{\partial r(n, k)}{\partial p_B} + \frac{p_B(n - 1)}{1 + p_B} r(n, k), \tag{A9}$$

allows us to evaluate

$$\sum_{n,k}(k + 1)r(n, k) = (1 - p_A)\langle n \rangle. \tag{A10}$$

As we could not find an expression for the moment generating function, the moments of $r(n)$ were calculated by brute force starting from Eq. (A7). The sum for the xth moment can be written as

$$\langle n^x \rangle = \frac{1 - p_A}{p_A(1 - p_B)} \sum_{k=0}^{\infty} (s_k S_x), \tag{A11}$$

where

$$s_k = \frac{1}{(k+1)!k!}\left(\frac{p_B(1 - p_A)}{p_A(1 - p_B)^2}\right)^k \tag{A12}$$

is the part of $\langle n^x \rangle$ independent of n and x and

$$S_x = \sum_{n=0}^{\infty} \frac{n^{x-1}n!}{(n - 2k - 1)!}[p_A(1 - p_B)]^n. \tag{A13}$$

To our knowledge, there is no easy way to evaluate these sums directly. Limiting ourselves to the first three moments, a strategy where the three sums involved are rewritten in terms of $\langle n \rangle$ is used. This is achieved by noting the following relations between the different S_x:

$$S_2 = \frac{2k + 1 + \zeta}{\zeta(1 - \zeta)} S_1, \tag{A14}$$

$$S_3 = \frac{4(k + 1)^2 - 4(k + 1) + 6(k + 1)\zeta + (1 - \zeta)^2}{\zeta(1 - \zeta)^2} S_1, \tag{A15}$$

where $\zeta = p_A(1 - p_B)$.

Successive applications of Eqs. (A9) and (A10) makes it possible to rewrite the sums for higher moments in terms of

$$\langle n \rangle = \frac{1 - p_A}{p_A(1 - p_B)} \sum_{k=0}^{\infty} (s_k S_1), \tag{A16}$$

and thus to express $\langle n^2 \rangle$ and $\langle n^3 \rangle$ in terms of $\langle n \rangle$.

[1]J. Padró, J. Mol. Struct. **416**, 243 (1997).
[2]C. Hansen, F. Stickel, T. Berger, R. Richert, and E. Fischer, J. Chem. Phys. **107**, 1086 (1997).
[3]C. Gainaru, R. Meier, S. Schildmann, C. Lederle, W. Hiller, E. Rössler, and R. Böhmer, Phys. Rev. Lett. **105**, 258303 (2010).
[4]E. Tsuchida, Y. Kanada, and M. Tsukada, Chem. Phys. Lett. **311**, 236 (1999).
[5]T. Yamaguchi, K. Hidaka, and A. Soper, Mol. Phys. **96**, 1159 (1999).
[6]K. Vahvaselka, R. Serimaa, and M. Torkkeli, J. Appl. Crystallogr. **28**, 189 (1995).
[7]M. Tomsič, A. Jamnik, G. Fritz-Popovski, O. Glatter, and L. Vlček, J. Phys. Chem. B **111**, 1738 (2007).

[8]F. Palombo, P. Sassi, M. Paolantoni, and A. Morresi, J. Phys. Chem. B **110**, 18017 (2006).

[9]M. Paolantoni, P. Sassi, A. Morresi, and R. Cataliotti, Chem. Phys. **310**, 169 (2005).

[10]M. Paolantoni, P. Sassi, A. Morresi, and R. Cataliotti, J. Raman Spectrosc. **37**, 528 (2006).

[11]V. Crupi, F. Longo, D. Majolino, and V. Venuti, J. Mol. Struct. **790**, 141 (2006).

[12]V. Crupi, F. Longo, D. Majolino, P. Migliardo, and V. Venuti, Philos. Mag. **87**, 705 (2007).

[13]J. Towey, A. Soper, and L. Dougan, Phys. Chem. Chem. Phys. **13**, 9397 (2011).

[14]I. Bakó, P. Jedlovszky, and G. Pálinkás, J. Mol. Liq. **87**, 243 (2000).

[15]R. Ludwig, ChemPhysChem **6**, 1369 (2005).

[16]E. Bianchi, R. Blaak, and C. N. Likos, Phys. Chem. Chem. Phys. **13**, 6397 (2011).

[17]J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, Amsterdam, 2006).

[18]E. Muller and K. Gubbins, Ind. Eng. Chem. Res. **40**, 2193 (2001).

[19]A. Bizjak, T. Urbic, V. Vlachy, and K. A. Dill, J. Chem. Phys. **131**, 194504 (2009).

[20]C. A. Koh, H. Tanaka, J. M. Walsh, K. E. Gubbins, and J. A. Zollweg, Fluid Phase Equilib. **83**, 51 (1993).

[21]K. Xu, Y.-G. Li, and W.-B. Liu, Fluid Phase Equilib. **142**, 55 (1998).

[22]E. Bianchi, P. Tartaglia, E. Zaccarelli, and F. Sciortino, J. Chem. Phys. **128**, 144504 (2008).

[23]P. Flory, J. Am. Chem. Soc. **74**, 2718 (1952).

[24]S. Erlander and D. French, J. Polym. Sci. **20**, 7 (1956).

[25]P. Flory, J. Am. Chem. Soc. **58**, 1877 (1936).

[26]H. Graener, T. Q. Ye, and A. Laubereau, J. Chem. Phys. **90**, 3413 (1988).

[27]W. Jorgensen, J. Phys. Chem. **90**, 1276 (1986).

[28]M. Haughney, M. Ferrario, and I. McDonald, J. Phys. Chem. **91**, 4934 (1987).

[29]I. Svishchev and P. Kusalik, J. Chem. Phys. **100**, 5165 (1994).

[30]W. L. Jorgensen, D. S. Maxwell, and J. TiradoRives, J. Am. Chem. Soc. **118**, 11225 (1996).

[31]L. Saiz and J. Padro, J. Phys. Chem. **101**, 78 (1997).

[32]A. K. Soper, Mol. Phys. **99**, 1503 (2001).

[33]L. Root and F. Stillinger, J. Chem. Phys. **90**, 1200 (1989).

[34]H. Graener, T. Q. Ye, and A. Laubereau, J. Chem. Phys. **91**, 1043 (1989).

[35]P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, 1st ed. (Cambridge University Press, Cambridge, 2009).