On the calculation of nearest neighbors in activity coefficient models

Citation for published version (APA):

Krooshof, G. J. P., Tuinier, R., & de With, G. (2018). On the calculation of nearest neighbors in activity coefficient models. *Fluid Phase Equilibria*, 465, 10-23. <https://doi.org/10.1016/j.fluid.2018.02.024>

Document license:
TAVERNE

DOI:
[10.1016/j.fluid.2018.02.024](https://doi.org/10.1016/j.fluid.2018.02.024)

Document status and date:
Published: 15/06/2018

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

On the calculation of nearest neighbors in activity coefficient models

Gerard J.P. Krooshof^{a,*}, Remco Tuinier^b, Gijsbertus de With^b

^a DSM Materials Science Center, P.O.Box 1171, 6160 BD Geleen, The Netherlands

^b Laboratory of Physical Chemistry, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:

Received 22 August 2017

Received in revised form

28 February 2018

Accepted 28 February 2018

Keywords:

Lattice theory

UNIQUAC

GEQUAC

COSMOSPACE

UNIFAC

COSMO-RS

COSMOSAC

ABSTRACT

Guggenheim proposed a theoretical expression for the combinatorial entropy of mixing of unequal sized and linear and branched molecules to improve the Flory-Huggins model. Later the combinatorial activity coefficient equation, which was derived from Guggenheim's model, was applied in the UNIQUAC, UNIFAC, and COSMOSAC models. Here we derive from Guggenheim's entropy theory a new function for the number of nearest neighbors of a compound in a multicomponent mixture for which the knowledge of the coordination number and a reference area are not needed. This new relation requires only the mole, volume and surface fraction of the compounds in the mixture. The benefit of the new relation is that both the combinatorial and the residual term in the aforementioned models can be made lattice-independent. We demonstrate that the proposed relation simplifies the Staverman-Guggenheim combinatorial model and can be applied with success to the UNIQUAC and COSMOSPACE model in the description of vapor-liquid phase equilibria and excess enthalpy. We also show that the new expression for the number of nearest neighbors should replace the relative surface area and the number of surface patches in the residual part of the UNIQUAC and the COSMOSPACE model, respectively. As a result a more rigorous version of the UNIQUAC and the COSMOSPACE model is obtained. This could serve as a better basis for predictive models like UNIFAC, COSMO-RS and COSMOSAC.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

In the 1940s Huggins [1] and Flory [2] derived an expression for the combinatorial entropy of mixing of athermal mixtures consisting of unequally sized molecules. Guggenheim [3,4] showed that the FH model overestimates the combinatorial entropy of mixing, because the connectivity of sites in a molecule reduces the number of possible configurations. Consequently, Guggenheim derived a correction term on the Flory-Huggins (FH) model for molecules having no internal contacts, which can be linear or branched. Staverman [5] extended Guggenheim's entropy expression to more complicated molecules, i.e. molecules containing rings. Both theoretical models contain, besides the mole and volume fraction, the surface fraction and require as additional parameter the number of nearest neighbors for each compounds. Abrams and Prausnitz [6] and Magnussen et al. [7] derived from the Staverman-Guggenheim (SG) entropy expressions for the activity coefficient, which have been applied in the combinatorial term of

the UNIQUAC [6], DISQUAC [8], UNIFAC [9,10], COSMOSAC [11], COSMOSPACE [12] and MOQUAC [13] models. In this paper we show that from the Guggenheim entropy of mixing an alternative expression for the combinatorial activity coefficient equation can be obtained, and thereby a new formula for the number of nearest neighbors. With this new equation for the number of nearest neighbors we revise the SG-corrected combinatorial activity coefficient as well as the residual activity coefficient of the UNIQUAC and the COSMOSPACE models. It makes the total activity coefficient, which is defined by the product of the combinatorial and the residual terms, for the UNIQUAC and COSMOSPACE model consistent. The new expression for the combinatorial activity coefficient equation is compared to the original form used in UNIQUAC and COSMOSPACE. In this comparison we will also consider the effect of the molecular size and shape defined by the Pauling bond lengths and the set of van der Waals radii, defined by Rowland and Taylor [14], respectively. Subsequently, the revised UNIQUAC and COSMOSPACE models are evaluated by comparing the description of vapor-liquid phase equilibria and excess enthalpy for alkane-alcohol binary systems.

* Corresponding author.

E-mail address: gerard.krooshof@dsm.com (G.J.P. Krooshof).

2. Theory

2.1. Guggenheim's entropy and the combinatorial activity coefficient

The mixing entropy ΔS for a binary mixture, consisting of molecules A and B, follows from Guggenheim's model [3] as

$$-\frac{\Delta S}{k_B} = \left(1 - \frac{z}{2}\right) \ln \frac{(r_A N_A)! (r_B N_B)!}{(r_A N_A + r_B N_B)!} + \frac{z}{2} \ln \frac{(q_A N_A)! (q_B N_B)!}{(q_A N_A + q_B N_B)!} \quad (1)$$

Here N_j , r_j , q_j are the number, the relative volume and the relative area of molecules j in the mixture, z is the lattice coordination number and k_B Boltzmann's constant. For a multicomponent system, where M is the number of species, and applying Stirling's approximation, the equation of Guggenheim becomes

$$-\frac{\Delta S}{k_B} = \left(1 - \frac{z}{2}\right) \left[\sum_{j=1}^M r_j N_j \ln(\phi_j) \right] + \frac{z}{2} \left[\sum_{j=1}^M q_j N_j \ln(\theta_j) \right], \quad (2)$$

where ϕ_j and θ_j are defined as the volume and area fraction of molecule j in the mixture

$$\phi_j = \frac{x_j r_j}{\sum_{j=1}^M x_j r_j}, \quad (3)$$

$$\theta_j = \frac{x_j q_j}{\sum_{j=1}^M x_j q_j}, \quad (4)$$

and where x_j is the mole fraction of component j . These expressions contain the relative van der Waals volume, r_j , and area, q_j , which require an ad-hoc definition of a reference volume and area. We recommend to use the van der Waals volume and area of the molecule, which can be calculated from molecular data as we will demonstrate in the results section, because the reference volume and area cancel in these equations. Further, it would be better to use the quantity $z_j q_j$ instead of q_j in the expression of the area fraction, but since it is usually assumed that the lattice coordination number is the same for all molecules, it cancels in Eq. (4). Guggenheim introduced for the number of nearest neighbors of linear or branched molecules the formula

$$z q_j = z r_j - 2(r_j - 1). \quad (5)$$

This can be rearranged into:

$$\left(1 - \frac{z}{2}\right) r_j = 1 - \frac{z}{2} q_j. \quad (6)$$

Upon substituting Eq. (6) into Eq. (2) Guggenheim obtained:

$$-\frac{\Delta S}{k_B} = \sum_{j=1}^M N_j \ln(\phi_j) + \frac{z}{2} \left[\sum_{j=1}^M q_j N_j \ln\left(\frac{\theta_j}{\phi_j}\right) \right]. \quad (7)$$

To express the activity coefficient of compound k in a mixture, we subtract first the ideal entropy of mixing to obtain the excess entropy

$$-\frac{\Delta S^E}{k_B} = \sum_{j=1}^M N_j \ln\left(\frac{\phi_j}{x_j}\right) + \frac{z}{2} \left[\sum_{j=1}^M q_j N_j \ln\left(\frac{\theta_j}{\phi_j}\right) \right], \quad (8)$$

and subsequently apply the thermodynamic relation

$$\ln \gamma_k^{\text{comb}} = -\frac{\partial \Delta S^E}{\partial N_k}. \quad (9)$$

This yields [7]:

$$\ln(\gamma_k^{\text{comb}}) = \left[\ln\left(\frac{\phi_k}{x_k}\right) + 1 - \frac{\phi_k}{x_k} \right] - \frac{z q_k}{2} \left[\ln\left(\frac{\phi_k}{\theta_k}\right) + 1 - \frac{\phi_k}{\theta_k} \right]. \quad (10)$$

We recognize as first term the Flory-Huggins contribution and as second term the Staverman-Guggenheim correction [15,16]. Staverman [5] argued that Guggenheim's entropy equation (Eq. (7)) is not valid for more complicated molecules, such as for molecules containing one or more rings. Vera et al. [17] demonstrated, however, that Eq. (10) is still obtained for bulky molecules. In the derivation above the $\left(1 - \frac{z}{2}\right) r_j$ term was replaced by the $\left(1 - \frac{z}{2} q_j\right)$ term in Eq. (2), but we can also replace the $\frac{z}{2} q_j$ term by $r_j \left(\frac{z}{2} - 1\right) + 1$. This gives

$$-\frac{\Delta S^E}{k_B} = \sum_{j=1}^M N_j \ln\left(\frac{\theta_j}{x_j}\right) + \left(1 - \frac{z}{2}\right) \left[\sum_{j=1}^M r_j N_j \ln\left(\frac{\phi_j}{\theta_j}\right) \right]. \quad (11)$$

After applying Eq. (9) we get the alternative expression

$$\ln(\gamma_k^{\text{comb}}) = \left[\ln\left(\frac{\theta_k}{x_k}\right) + 1 - \frac{\theta_k}{x_k} \right] - \left(1 - \frac{z}{2}\right) r_k \left[\ln\left(\frac{\theta_k}{\phi_k}\right) + 1 - \frac{\theta_k}{\phi_k} \right]. \quad (12)$$

This new expression for the SG-correction term no longer contains the number of nearest neighbors, $z q_k$, but the lattice coordination number times the number of sites, i.e. $z r_k$. Since Eq. (10) and Eq. (12) should yield the same activity coefficient for compound k , it follows

$$\begin{aligned} & \left[\ln\left(\frac{\phi_k}{x_k}\right) + 1 - \frac{\phi_k}{x_k} \right] - \frac{z q_k}{2} \left[\ln\left(\frac{\phi_k}{\theta_k}\right) + 1 - \frac{\phi_k}{\theta_k} \right] \\ &= \left[\ln\left(\frac{\theta_k}{x_k}\right) + 1 - \frac{\theta_k}{x_k} \right] - \left(1 - \frac{z}{2}\right) r_k \left[\ln\left(\frac{\theta_k}{\phi_k}\right) + 1 - \frac{\theta_k}{\phi_k} \right]. \end{aligned} \quad (13)$$

With Eq. (5) we can reduce and rearrange the above expression to a property Q_k , which is half the number of nearest neighbors

$$Q_k \equiv \frac{z q_k}{2} = \frac{1 - \frac{\theta_k}{\phi_k} - \frac{\phi_k}{x_k} + \frac{\theta_k}{x_k}}{2 - \frac{\theta_k}{\phi_k} - \frac{\phi_k}{\theta_k}} = \frac{1 - \frac{\phi_k}{x_k}}{1 - \frac{\phi_k}{\theta_k}}. \quad (14)$$

This new relation tells us that Q_k can be calculated from the mole, area and volume fraction, and that $z q_k$ is not a free parameter anymore, as it was considered in the original UNIQUAC model. In Appendix A we show that Eq. (14) yields a constant, despite the fact that it contains functions of mole fraction. It has clear similarity with relation 5, which can be rearranged to

$$Q_k \equiv \frac{z q_k}{2} = \frac{1 - r_k}{1 - \frac{r_k}{q_k}}. \quad (15)$$

In order to apply Eq. (15) one needs either a reference area and a lattice coordination number, or a reference volume and area in order to calculate the relative volume r_k and the relative area q_k , respectively. In the past this factor was obtained by setting $z = 10$ and by calculating q_k using a reference area. We remark that the reference volume $V_{\text{ref}} = 25.17 \text{ \AA}^3$ was never changed after Abrams and Prausnitz [6] introduced it. The reference area, however, has been optimized several times in order to bring the activity models more in agreement with experimental results. Table 1 shows

Table 1
Values of the reference surface areas A_{ref} proposed by various groups.

Author	$A_{\text{ref}}(\text{\AA}^2)$	Model
Abrams and Prausnitz [6]	41.56	UNIQUAC
Wang et al. [18]	79.53	COSMOSAC
Soares [19]	124	COSMOSAC
Bronneberg and Pfennig [13,20]	3.21	MOQUAC

several values for the reference area for different models, in which the SG-combinatorial correction term was implemented. In here the lattice coordination number was always fixed on the value $z = 10$, as was proposed by Abrams and Prausnitz [6].

We remark that with smaller values of the reference area, the number of nearest neighbors of a compound increases. This has little impact on the combinatorial term, but has large effect on the residual activity coefficient. In fact, the binary interaction parameter needs to be adjusted to lower values in order to keep the description of the experimental phase equilibria data in agreement.

2.2. Lattice independent UNIQUAC

Above it has been shown that choosing a reference area and a lattice coordination number is no longer required, and that Eq. (14) defines the combinatorial part of the activity coefficient model. Substituting Eq. (14) into the original combinatorial activity coefficient, Eq. (10), gives

$$\begin{aligned} \ln(\gamma_k^{\text{comb}}) &= \ln\left(\frac{\phi_k}{x_k}\right) + \left(1 - \frac{\phi_k}{x_k}\right) \left[\frac{\ln\left(\frac{\phi_k}{\theta_k}\right)}{1 - \frac{\phi_k}{\theta_k}} \right] \\ &= \ln\left(\frac{\phi_k}{x_k}\right) - Q_k \ln\left(\frac{\phi_k}{\theta_k}\right). \end{aligned} \quad (16)$$

The latter equation is referred to Vera et al. [17] as the restricted form of the UNIQUAC equation. However, we point out that Vera et al. [17] apply Eq. (15), while we use Eq. (14), which requires no definition of the reference area A_{ref} . We mention that in the limit $\theta_k \rightarrow \phi_k$, the term between the square brackets becomes unity, and the Flory-Huggins equation is obtained. This also happens when we set $q_j = r_j$, as this gives by Eq. (15) $r_j = q_j = 1$, from which it follows that $\theta_j = \phi_j$. The derived equation for Q_k is also directly applicable in the residual part of the UNIQUAC model. It is noted that the residual part of the original UNIQUAC model contains an inconsistency. Instead of the frequently published equation, our analysis leads to

$$\ln(\gamma_k^{\text{res,UQ}}) = Q_k \left[1 - \ln\left(\sum_{j=1}^M \theta_j \tau_{jk}\right) - \sum_{j=1}^M \frac{\theta_j \tau_{kj}}{\sum_{i=1}^M \theta_i \tau_{ij}} \right]. \quad (17)$$

This expression contains a factor Q_k instead of q_k , which has been used in the original UNIQUAC model (see Appendix B). In Eq. (17) τ_{ij} denotes the Boltzmann weighting factor given by the contact energy of a lattice cell of compound i and that of compound j

$$\tau_{ij} = \exp\left[-\frac{\Delta U_{ij}}{RT}\right]. \quad (18)$$

The total activity coefficient for the lattice independent UNIQUAC model can therefore be expressed as

$$\begin{aligned} \ln(\gamma_k^{\text{UQ}}) &= \ln\left(\frac{\phi_k}{x_k}\right) - Q_k \ln\left(\frac{\phi_k}{\theta_k}\right) \\ &+ Q_k \left[1 - \ln\left(\sum_{j=1}^M \theta_j \tau_{jk}\right) - \sum_{j=1}^M \frac{\theta_j \tau_{kj}}{\sum_{i=1}^M \theta_i \tau_{ij}} \right]. \end{aligned} \quad (19)$$

The above principle holds also for the UNIFAC model. The difference between principle UNIQUAC and UNIFAC is that in the latter not molecules but molecule fragments are used as interacting entities in the residual part. Although UNIQUAC has been applied with success, we must not forget that the UNIQUAC model is based upon a first order approximation of the local concentration around a molecule in a mixture. This approach is accurate for systems having weak interactions, but for strongly interacting molecules the activity coefficient of the solute is underestimated. In that case the local activity instead of the local concentration of the solute should be used. This has been done in the COSMOSPACE [12] and GEQUAC [21] models.

2.3. Lattice independent COSMOSPACE

For a binary system, where each molecule can have two types of surfaces, the COSMOSPACE activity coefficient is expressed as [12].

$$\ln(\gamma_k^{\text{res,CS}}) = \frac{n_k}{2} \left(\zeta_k^A \ln \left[\frac{\gamma_{A,x}}{\gamma_{A,k}} \right] + \zeta_k^B \ln \left[\frac{\gamma_{B,x}}{\gamma_{B,k}} \right] \right), \quad (20)$$

where n_k is the total number of surface segments of molecule k , and ζ_k^J is the fraction of surface type J on molecule k . The $\underline{x} \equiv (x_1, x_2)$ is used to define the fractions of the two components in the mixture and the pure state. We mention that in the original publication the total number of surface patches is calculated by the relative area as defined in the UNIQUAC model: $n_k = 2q_k$. Since we consider a fully occupied lattice, i.e., there are no empty cells in the lattice, the total number of surface segments on a molecule has to be equal to the number of nearest neighbors. That is to say, we do not subdivide the contact area into smaller patches, as is done in COSMOSAC [11], because the patches of opposite segment sides make simultaneous contact, thereby averaging the interaction energy of the two contacting surface segments. Therefore we consider

$$\frac{n_k}{2} = \frac{zq_k}{2} = Q_k. \quad (21)$$

The variables $\gamma_{J,x}$ and $\gamma_{J,k}$ in Eq. (20) denote the activity coefficient of surface type J in the mixture and the pure state, respectively. They can be calculated by

$$\ln(\gamma_{A,x}) = \ln \left[\frac{1}{\theta_{A,x}} + \frac{1 - \sqrt{(1 + 4\theta_{A,x}\theta_{B,x}\omega)}}{2\omega\theta_{A,x}^2} \right], \quad (22)$$

and

$$\ln(\gamma_{B,x}) = \ln \left[\frac{1}{\theta_{B,x}} + \frac{1 - \sqrt{(1 + 4\theta_{A,x}\theta_{B,x}\omega)}}{2\omega\theta_{B,x}^2} \right], \quad (23)$$

with $\omega = 1/\tau^2 - 1$ reflecting the interaction between the two different surface segments A and B, and where τ is the Boltzmann weighting factor of the COSMOSPACE model which is for weak

interactions related to the interaction parameters of UNIQUAC; $\tau \approx \sqrt{\tau_{21}\tau_{12}}$. In the limit $\omega \rightarrow 0$ the surface activity coefficients become unity, and the binary system becomes an ideal solution. The surface fractions of the two components in the mixture and the pure state are calculated by

$$\theta_{A,x} = \frac{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A}{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A + x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}, \quad (24)$$

and

$$\theta_{B,x} = \frac{x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A + x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}. \quad (25)$$

The total activity coefficient of the lattice independent COSMOSPACE model can be expressed by the sum of Eqs. (16) and (20)

$$\ln(\gamma_k^{CS}) = \ln\left(\frac{\phi_k}{x_k}\right) - Q_k \ln\left(\frac{\phi_k}{\theta_k}\right) + Q_k \left[\zeta_k^A \ln\left[\frac{\gamma_{A,x}}{\gamma_{A,k}}\right] + \zeta_k^B \ln\left[\frac{\gamma_{B,x}}{\gamma_{B,k}}\right] \right]. \quad (26)$$

3. Results and discussion

In this section we evaluate the accuracy of the renewed expression for half the number of neighbors, Eq. (14). For this purpose we chose binary mixtures of n-alkanes and 1-alcohols. Subsequently, we investigate how the UNIQUAC and COSMOSPACE models perform using the new definition for the number of nearest neighbors.

3.1. The number of nearest neighbors of n-alkanes

The calculation of the number of nearest neighbors by Eq. (14) for binary mixtures of alkanes between methane and nonacosane has been carried out for a series of 200 mol fractions between 0.001 and 0.999. Outside this range Eq. (14) suffers from computational inaccuracy due to the vanishing denominator, approaching zero. To calculate the volume and surface fractions in the binary mixture we used the van der Waals volume (V_{vdW}) and surface area (A_{vdW}) given by Bondi [22]. For n-alkanes Bondi gives the following relations

$$V_{vdW} = \frac{6.88 + 10.23N_C}{0.6022}, \quad (27)$$

and

$$A_{vdW} = \frac{1.54 + 1.35N_C}{0.06022}, \quad (28)$$

where we kept the original Bondi parameters and used the factors 0.6022 and 0.06022 to scale from molar size (cm^3/nmol , and cm^3/mol) to molecular area ($\text{\AA}^2$) and volume ($\text{\AA}^3$), respectively. As can be seen from the equidistant lines in Fig. 1, there is a clear linear relationship between zq_k and N_C .

Methane, depicted at the bottom line, has a constant value for zq_k over the entire range from ethane to n-nonacosane. The same holds for the higher alkanes. The explanation for this is given in Appendix A. The open space in Fig. 1 represents the position of the pure liquid, to which Eq. (14) does not apply. However, since the values on the left and the right side of the open space are the same, it seems logical to assign this value as zq_k of the pure compound. This is an important observation, which we will use later. Based on

Fig. 1. The above illustrates the number of nearest neighbors (zq_k) of two n-alkanes between methane and nonacosane calculated by Eq. (14) using the van der Waals volumes and areas of Bondi [22], as has been shown in app. A2. In the bottom row the zq of methane in a mixture of ethane ($N_C = 2$) to nonacosane ($N_C = 29$) is shown, continuing to the top row where the zq of nonacosane in mixture of methane ($N_C = 1$) to octacosane ($N_C = 28$) is given.

the results depicted in Fig. 1, the relation between the number of nearest neighbors of the pure compound and the alkane number is given by

$$zq_k = 4.87 + 4.27N_C. \quad (29)$$

Sayegh and Vera [17,23] proposed the following empirical relation for the number of nearest neighbors based on an analysis of 15 classes of compounds

$$zq_k = 0.4228V_{vdW} + 2(1 - l_k), \quad (30)$$

where l_k is the number of internal contacts in a molecule and the van der Waals volume is expressed in cm^3/mol . The empirical constant 0.4228 has the dimension mol/cm^3 and is equal to the quantity $(z - 2)/V_{\text{ref}}$. We see that the value for the reference volume is dependent on the value of z . E.g. if $z = 10$, then the reference volume is $18.92 \text{ cm}^3/\text{mol}$ or $31.4 \text{ \AA}^3/\text{molecule}$, which gives $A_{\text{ref}} = 48.2 \text{ \AA}^2/\text{molecule}$, which is 20% higher than applied in the original UNIQUAC model (see Table 1). For n-alkanes, where $l_k = 0$, and using the van der Waals volumes defined by the van der Waals radii of Bondi, Eq. (30) becomes

$$zq_k = 4.91 + 4.33N_C. \quad (31)$$

This relation is close to Eq. (29). Although the van der Waals radii as proposed by Bondi are often used in the calculation of molecule dimensions, they are still matter of debate. We refer to the review article of Batsanov [24], in which it is argued that the van der Waals radii of Rowland and Taylor [14] should be considered as more reliable, because a much larger experimental XRD-data set has been used. The van der Waals radii of Rowland and Taylor for hydrogen, carbon and oxygen are $R_H = 1.09 \text{ \AA}$, $R_C = 1.75 \text{ \AA}$, and $R_O = 1.56 \text{ \AA}$, respectively. With these values and the Pauling bond lengths [25] the van der Waals volume and surface area can be calculated. Subsequently, using the same procedure as before, the number of nearest neighbors is obtained

$$zq_k = 4.37 + 3.55N_C \quad (32)$$

For comparison we recall that in the original UNIQUAC model the number of nearest neighbors is calculated by

$$zq_k = z \frac{A_{\text{vdW}}}{A_{\text{ref}}} = \frac{A_{\text{vdW}}}{4.15}, \quad (33)$$

where z is set to 10, $A_{\text{ref}} = 41.5 \text{ \AA}^2$ and A_{vdW} is calculated with the Bondi set of van der Waals radii (i.e. Eq. (28)). Fig. 2 shows a comparison of the aforementioned equations for zq_k of n -alkanes.

We see that the set of van der Waals radii of Rowland and Taylor, i.e. Eq. (32), gives the lowest values for zq_k , while Eq. (33) gives significantly higher values, which might lead to systematic deviations in the UNIQUAC model.

3.2. The number of nearest neighbors of 1-alcohols

Another class of linear molecules are the 1-alcohols. The required van der Waals radii of Bondi [22] and those of Rowland are (in $\text{\AA}$): $R_H = 1.20$, $R_C = 1.70$ and $R_O = 1.52$, and $R_H = 1.09$, $R_C = 1.75$ and $R_O = 1.56$, respectively. Using the same method as described in the section on n -alkanes, this leads to for the Bondi parameter set of van der Waals radii. The set of van der Waals radii of Rowland and Taylor yields

$$zq_k = 6.24 + 3.78N_C \quad (34)$$

$$zq_k = 6.14 + 3.39N_C \quad (35)$$

Results of the various equations for zq_k of the 1-alcohols are depicted in Fig. 3.

Like in the case of the n -alkanes, Eq. (14) gives for the alcohol mixtures lower values for the number of nearest neighbors than those calculated by the original UNIQUAC formula. Comparing Fig. 2 with Fig. 3 it follows that the relation based on the van der Waals radii of Rowland and Taylor are more in line with the expectation that the alcohols have more nearest neighbors than the alkanes with the same carbon number, as long as $N_C < 11$. This

Fig. 2. The number of nearest neighbors (zq_k) of n -alkanes calculated by Eq. (32) (solid line), Eq. (29) (dashed-dotted), Eq. (30) (dotted line) and Eq. (33) (dashed line).

Fig. 3. The number of nearest neighbors (zq_k) of 1-alcohols calculated by Eq. (35) (solid line), Eq. (34) (dashed-dotted line), Eq. (30) (dotted line) and Eq. (33) (dashed line).

suggests that further optimization of the van der Waals radii is still needed. In fact, we expect that the slope of the number of nearest neighbors as function of the carbon number of the alcohols should be the same as that of the alkanes, because in both cases this only depend on the number of the CH_2 groups. Batsanov [24] indicated that the concept of spherical atoms in a molecule probably needs to be adjusted to ellipsoids, because the radius in the bond direction is smaller than the one perpendicular to it, which might also be important in a further refinement.

3.3. The number of nearest neighbors of mixtures of n -alkanes and 1-alcohols

While zq_k of an alkane is independent of the other alkanes in the mixture, it becomes N_C dependent when an 1-alcohol is added.

Fig. 4 shows the case of n -hexane dissolved in an n -alkane and in a 1-alcohol. Although the number of nearest neighbors for n -hexane in a 1-alcohol does not vary with concentration, we observe that it becomes 1-alcohol dependent over a small range of N_C -values. When the carbon number of the alcohol is low, the number of nearest neighbors of n -hexane is slightly below the value of that in an alkane. In the case of Bondi's set of van der Waals radii it even gives negative values when the difference in carbon number becomes smaller, which is unphysical. This artefact is caused by small errors in the calculation of the van der Waals volume and area, which is magnified by the denominator of Eq. (14). This artefact would also occur in the original UNIQUAC method, if the developers had used the right part of Eq. (15), instead of the applied $zq/2$, for which an ad-hoc value of the coordination number ($z = 10$) and the reference area ($a_{\text{ref}} = 41.56 \text{ \AA}^2$) had to be defined. Above the carbon number of hexane the number of nearest neighbors approaches the pure component value. This makes sense, because the alcohol starts to increasingly resemble an alkane. The presence of an apparent singularity around molecules of equal size, however, demonstrates that the choice for a value of the number nearest neighbors is still a critical issue. In order to make the new expression for the UNIQUAC model applicable for all

Fig. 4. Results for Q_k of hexane dissolved in alkanes (green dots/diamonds) and alcohols (blue squares/triangles) with Eq. (14) using Rowland's set (solid curves) or Bondi's set (dashed curves) of van der Waals radii. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

types of solvent mixtures, we take the value obtained from the calculation of the compound in a mixture of molecules of its own class. This is equivalent to the number of nearest neighbors of the compound in its pure state. This choice implies that the coordination of other molecules in the first shell around a central molecule is set equal to that of the pure state.

3.4. The Staverman-Guggenheim correction term

We will now compare the combinatorial activity coefficient using the different expressions for the number of nearest neighbors and apply these to the binary system hexane - hexadecane.

Fig. 5 shows the comparison between the combinatorial activity coefficient of the different models and parameter sets for the binary system of n-hexane (1) and n-hexadecane (2). The activity coefficients of the Flory Huggins model, for which we used the original UNIQUAC parameters $r_1 = 4.5$ and $r_2 = 11.25$, are plotted as dashed curves, which end at the limiting activity coefficient values of $\gamma_1^{\text{comb},\infty} = 0.729$ and $\gamma_2^{\text{comb},\infty} = 0.558$. The different Staverman-Guggenheim models give curves which nearly coincide. The original UNIQUAC model, where $z = 10$, $q_1 = 3.86$ and $q_2 = 9.26$, gives curves that end at $\gamma_1^{\text{comb},\infty} = 0.7404$ and $\gamma_2^{\text{comb},\infty} = 0.5798$. The refined Staverman-Guggenheim model, i.e. Eq. (16) with Q_k defined by Eq. (14), gives using the van der Waals molecular volume and area based on the radii set of Bondi: $\gamma_1^{\text{comb},\infty} = 0.7380$ and $\gamma_2^{\text{comb},\infty} = 0.5752$, while the set of Rowland and Taylor gives: $\gamma_1^{\text{comb},\infty} = 0.7404$ and $\gamma_2^{\text{comb},\infty} = 0.5798$. These values demonstrate that the difference between the original and the refined UNIQUAC models is negligibly small; less than 0.1%. This observation shows that the new equation for the number of nearest neighbors (Eq. (14)) gives nearly the same result as the equation used in the original UNIQUAC equation (Eq. (5)) for which a reference surface area had to be defined. Our approach demonstrates that the lattice model of UNIQUAC can be transformed into a fluid model without the definition of a reference area. This transformation is also observed in the Flory-Huggins model. Initially the number of lattice cells for each molecule was defined, but in the end the activity

Fig. 5. Combinatorial activity coefficient γ^{comb} of the binary n-hexane and n-hexadecane. Flory-Huggins model (small dashed curve, $\gamma_1^{\text{comb},\infty} = 0.729$). The curves of the original UNIQUAC and the refined Staverman-Guggenheim models based on Bondi's and Rowland's set of van der Waals radii nearly coincide.

coefficient equation contains only mole and volume fractions, for which this definition is not needed. Likewise, the Staverman-Guggenheim correction requires the definition of the number of sites of each molecular surface, i.e. zq , but this is later replaced with Eq. (14), and the activity becomes a function of mole, volume and surface fractions only. In other words, in both cases the concept of a lattice has served as a vehicle to reveal the dependency between activity coefficient and concentrations.

3.5. Activity coefficient models

3.5.1. UNIQUAC

In the theoretical section of this work (see Appendix B) we show that the residual part of the original UNIQUAC model contains an error. The correct derivation delivers a residual activity coefficient equation with Q_k and not q_k as molecular dependent variable.

Fig. 6 depicts several experimental isothermal vapor-liquid equilibria data and the UNIQUAC descriptions in three different ways: the original UNIQUAC model (dashed curves), and the lattice independent UNIQUAC model as given in Eq. (19) using Eq. (14). The optimized UNIQUAC interaction parameters Table 2.

The key quantity Q_k is calculated by the volume and the surface area fractions defined by the molecular volumes and areas from Bondi's (solid curves) and Rowland's (dashed-dotted curves) set of van der Waals radii. We observe in Fig. 6 that the three different forms of the UNIQUAC model give equivalent descriptions of the vapor-liquid equilibria. The new equation somewhat overpredicts the pressure in the case of ethanol, and gives lower pressures in the case of pentanol. These deviations are not caused by the new expression, but by the combination of parameters and the first order approximation of the local concentration. This is more clear in the description of the binary n-hexane and ethanol, where all three models yield a heterogeneous azeotrope, while the experimental data show a homogeneous azeotrope. The exaggeration in pressure exists for all UNIQUAC models, but is more present in case of the lattice independent versions. The left plot of Fig. 7 shows that

Fig. 6. Experimental data and UNIQUAC description of binaries between n-hexane(1) and alcohol(2). From left to right, and from top to bottom respectively: methanol (25 °C, from Ref. [26]), ethanol(40 °C, from Ref. [27]), propanol(50 °C, from Ref. [28]) and pentanol(40 °C, from Ref. [29]). Dashed curves (Original UNIQUAC model), solid and dashed-dotted curve (Eq. (19)) with Q_k based on Bondi's and Rowland's set of van der Waals radii, respectively.

Table 2

UNIQUAC binary interaction parameters for n-hexane and a 1-alcohol using number of nearest neighbors of the pure compound as given in Figs. 2 and 3.

1-Alcohol (T °C)	Original		Bondi		Rowland	
	τ_{12}	τ_{21}	τ_{12}	τ_{21}	τ_{12}	τ_{21}
Methanol (25 °C)	0.129	1.017	0.237	1.733	0.165	1.814
Ethanol (45 °C)	0.161	1.289	0.282	1.785	0.261	1.787
Propanol (50 °C)	0.395	1.898	0.432	1.568	0.397	1.578
Pentanol (30 °C)	0.370	1.293	0.528	1.477	0.431	1.601

Table 3

Mean UNIQUAC interaction parameter τ_{298} , and interaction energy, ΔE , values for the binary pair n-hexane - alcohol at 298.15 K. UQ denotes original UNIQUAC model, Bondi and Rowl. the lattice independent UNIQUAC model with Bondi's and Rowland's basis set of van der Waals radii.

2 nd Compound	τ_{298}			$\Delta E_{298}(\text{kJ/mol})$		
	UQ	Bondi	Rowl.	UQ	Bondi	Rowl.
Methanol	0.362	0.641	0.547	2.52	1.10	1.50
Ethanol	0.456	0.710	0.683	1.91	0.85	0.95
1-Propanol	0.584	0.810	0.776	1.33	0.52	0.63
1-Pentanol	0.687	0.881	0.828	0.93	0.31	0.47

the lattice independent UNIQUAC models yield higher activity coefficients than the original UNIQUAC model; for hexane at mole fractions between 0.1 and 0.5 and for ethanol at mole fractions between 0.7 and 0.9. This difference between the two types of UNIQUAC models is caused by the small q_k in the original UNIQUAC model versus the applied correct factor Q_k in the lattice independent models; $q_{\text{Ethanol}} = 1.97$, $q_{\text{Hexane}} = 3.86$, and Rowland's data give $Q_{\text{Ethanol}} = 6.48$ and $Q_{\text{Hexane}} = 12.85$.

The interaction parameters for the original UNIQUAC model are the smallest in value, followed by those of the lattice independent model based on Rowland's and Bondi's basis set of van der Waals radii. This is in line with the fact that the original UNIQUAC model uses the smallest value for the relative area, while in the refined models the van der Waals radii of Bondi give a larger molecular surface area than those of Rowland.

3.5.2. COSMOSPACE

In the article on the COSMOSPACE model [12] it was already made clear that poor results are obtained when the molecular surface of each compound is considered to be made of one type of surface. This so-called homogeneous surface approach fails, because in reality each interacting molecule, even an alkane, consists of at least two different types of surfaces. Therefore the most comprehensive form of the COSMOSPACE model is the so-called

Fig. 7. Activity coefficient of hexane-ethanol. Experimental data (symbols) [27]. Left plot: original UNIQUAC model (dashed curves), new UNIQUAC models with Bondi's (solid curves) and Rowland's (dashed dotted curves) set of van der Waals radii. Right plot: original COSMOSPACE model (solid curves), new COSMOSPACE models with Bondi's (dashed curves) and Rowland's (dashed dotted curves) set of van der Waals radii. To compare the parameters of the different UNIQUAC models, it is convenient to take the mean interaction parameter (i.e. $\tau_{298} = \sqrt{\tau_{12}\tau_{21}}$) at 25 °C. Table 3 shows that with increasing alkane number of the alcohol the binary system becomes more ideal; the mean interaction parameter shifts to unity and the interaction energy, calculated by $\ln(\tau_{298})$, goes to zero.

non-homogeneous double-binary set-up. We note that more surface types of interaction can be taken into account, but this requires more computation [12]. For this work, however, we can show that the non-homogeneous double-binary set-up is sufficient to demonstrate that Eq. (14) provides an accurate description for a series of vapor-liquid equilibria of hexane and an alcohol. For the description of the isothermal experimental vapor-liquid equilibria we have used two parameters: the number of nearest neighbors of the hydroxyl group (n_{OH}) in the alcohol and the alkyl-hydroxyl interaction parameter (τ_{298}) at the reference temperature of 298.15 K. An Arrhenius equation is used to define the interaction parameter τ at other temperatures

$$\ln(\tau) = -\frac{\Delta E_{298}}{RT} = \frac{298.15}{T} \ln(\tau_{298}), \quad (36)$$

where ΔE_{298} is the interaction energy between the alkyl and the hydroxyl surfaces at 298.15 K. Simultaneous optimization of the isothermal vapor-liquid equilibria yields to a set of parameter values listed in Table 4.

Fig. 8 shows that the COSMOSPACE model gives an excellent description of the phase equilibria data. The two lattice independent COSMOSPACE models show hardly any difference between the results obtained with the van der Waals radii of Bondi and Rowland; the curves almost coincide. Statistical information in the form of the average absolute deviation indicates that Bondi's set of van der Waals radii gives slightly better results than Rowland's set (See Table 5).

In comparison to the UNIQUAC approach the COSMOSPACE model clearly gives a better quantitative description of the

experimental vapor-liquid phase equilibria, especially the transition from heterogeneous to homogeneous azeotrope is captured well. This is caused by the better activity coefficient description in the mid concentration range as can be observed by comparing the two plots in Fig. 7. The description of the experimental data is best performed by the lattice independent COSMOSPACE model with Bondi's set of van der Waals radii. With the number of contacts associated with hydrogen bonding, $n_{OH} = 2Q_{OH}$, the change in pair interaction energy is calculated. The obtained values of 13.4, 18.6 and 18.9 kJ/mol indicate that the lattice independent model is closer to the energy of hydrogen bonding of alcohols (20 kJ/mol for methanol/ethanol [33]) than the original COSMOSPACE model.

In Fig. 9 we depict the prediction of the excess enthalpy of the binary systems n-hexane - ethanol at 40 °C [27] and n-nonane - ethanol at 25 °C [34], for which the parameters are taken from the vapor-liquid equilibria optimization. The prediction is qualitatively very good with respect to the location of the maximum.

The average absolute deviation of the predicted excess enthalpy is about 15% (see Table 6). This deviation is understandable from the fact that the induced dipole interaction between the alkyl parts of the alkane and the alkanol has not been taken into account in the calculations. It is expected that the inclusion of a dispersion term, as has been done in the DISQUAC [8,36], will improve the prediction of the excess enthalpy.

Fig. 10 shows the fraction of the number of nearest neighbors, ζ_{OH}^{OH} , that acts as hydroxyl surface according to the optimization of the COSMOSPACE parameter n_{OH} and the total number of nearest neighbors n_k given by Eq. (34) (dashed curve) and Eq. (35) (solid curve). For comparison also the ratio in the van der Waals surface area of the -OH group and the total molecule are plotted. The van der Waals surface areas of the hydroxyl group are 24.3 Å² and 23.7 Å², for the Bondi and Rowland sets of van der Waals radii, respectively. The total van der Waals surface areas of the 1-alcohols are given by the linear relations for the Bondi and Rowland set of van der Waals radii, respectively.

$$A_{vdW} = 37.03 + 22.41N_C, \quad (37)$$

and

$$A_{vdW} = 36.37 + 20.09N_C, \quad (38)$$

Table 4

COSMOSPACE parameters for alkane-1-alcohol systems derived from fitting VLE data (Table 5).

COSMOSPACE Model	Original	New	New
Parameter	Bondi	Bondi	Rowland
$Q_{OH}(-)$	—	1.18	1.25
$n_{OH}(-)$	2.29	2.35	2.50
$\tau_{298}(-)$	0.0979	0.0409	0.0471
ΔE_{298} (kJ/mol)	5.76	7.92	7.58
$n_{OH}\Delta E_{298}$ (kJ/mol)	13.4	18.6	18.9

Fig. 8. Experimental data (symbols) and COSMOSPACE description (curves) of vapor-liquid equilibria between alkane(1) and alcohol(2). From left to right and from top to bottom: hexane-methanol [26], hexane-ethanol [27], heptane-ethanol and nonane-ethanol [30], hexane-propanol [28], and hexane-1-hexanol [31]. Solid curves represent the original COSMOSPACE model, the dashed and dashed-dotted curves are the lattice independent COSMOSPACE model based on Bondi's and Rowland's set of van der Waals radii, respectively.

We observe that the COSMOSPACE model gives lower values for the fraction of hydrogen bonding area, than is predicted by the ratio of van der Waals surface areas. This is logical because only a part of the hydroxyl van der Waals surface area is involved in hydrogen bond breaking. The original model overpredicts the surface area involved in hydrogen bonding, as a result of incorrect use of the q_k parameter, instead of Q_k . This also explains why the incrementation of the hydroxyl group in the UNIFAC method failed for methanol, which would consist of one methylene group and one hydroxyl group. In UNIFAC the surface area of the strongly interacting part of the hydroxyl group was set equal to the van der Waals surface of the hydroxyl group. Extrapolation towards methanol would make it a

strongly hydrophilic compound. Therefore methanol was defined as a new UNIFAC group.

4. Concluding remarks

We have shown that two expressions for the combinatorial activity coefficient can be derived from Guggenheim's model for the entropy of mixing of linear molecules. These expressions yield a unique expression for the number of nearest neighbors of a molecule in a multicomponent mixture. This formula is a function of mole, volume and area fractions only, and makes a choice for the value for the lattice coordination number and the reference surface

Table 5

Average absolute deviation in pressure: $100 \times |1 - P_{calc}/P_{exp}|$ of the various COSMOSPACE (CS) models. NPS = number of mixture data points in regression. Pure data points are used to scale pressure.

Binary VLE	T	NPS	CS(Original)	CS(New)	CS(New)	
Alkane-Alcohol	°C	—	Bondi	Bondi	Rowland	lit.
Hexane-Methanol	25	4	18.7	3.5	5.1	[26]
Hexane-Ethanol	40	38	2.9	1.2	2.4	[27]
Heptane-Ethanol	30	50	6.4	3.3	4.1	[32]
Nonane-Ethanol	25	25	3.8	0.9	1.6	[32]
Hexane-1-Propanol	50	20	3.2	1.4	2.3	[28]
Hexane-1-Butanol	60	24	0.8	1.5	2.4	[30]
Hexane-1-Pentanol	40	13	0.9	1.7	2.2	[29]
Hexane-1-Hexanol	30	21	7.2	4.1	5.4	[31]
Average			4.3	2.1	3.1

Fig. 9. Predicted excess enthalpy of alkane-ethanol binaries with parameters from Table 4. Symbols represent experimental data. From top to bottom: hexane-ethanol [27], heptane-ethanol [35], and nonane-ethanol [34]. The solid curves represent the original COSMOSPACE model, the dashed and dashed dotted curves the lattice independent COSMOSPACE model based on Bondi's and Rowland's set of van der Waals radii, respectively.

superfluous. In other words, in combination with the (revised) residual term as obtained for the COSMOSPACE model, it makes the total activity coefficient model lattice independent. We expect that the new formula also works for more complicated molecules, but this still needs to be verified. We have demonstrated that the new expression for the number of nearest neighbors leads to lattice independent versions of the UNIQUAC and COSMOSPACE models. In order to put the various models in perspective, Table 7 shows a comparison indicating the lattice (in)dependency of the combinatorial and residual activity coefficients. In this table we use the label σ -profiles, by which we indicate the whole procedure starting with QM-calculations, the averaging of surface charge densities of molecules and the collection of surface area into histograms. Hence, the three-dimensional lattice information vanishes in these conversions prior to the calculation of the surface and residual activity coefficients.

For the UNIQUAC model we first had to remove an inconsistency from the residual term of the UNIQUAC model in order to show this possibility. The original and the lattice independent version of UNIQUAC describe the investigated vapor-liquid phase equilibria with the same quality and accuracy as before. A much better description of these phase equilibria is obtained with the lattice independent version of the COSMOSPACE model. For 1-alcohols mixed with n-hexane at moderate temperatures it captures the

transition from a heterogeneous to a homogeneous azeotrope when changing from methanol to ethanol. The number of parameters is substantially lower than those for the original and the lattice independent version of the UNIQUAC model. Especially for strongly interacting systems the lattice independent COSMOSPACE model is recommended above the UNIQUAC model.

The new formula gives a constant value for the number of nearest neighbors of a compound dissolved in a mixture composed of molecules of the same class. This value stays constant over the entire concentration range, except at limiting concentrations, where computational accuracy plays a role. Because the value of the number of nearest neighbors of a compound is independent from the other molecules of the same class in the mixture, one can assign this value as the number of nearest neighbors of the pure compound. This has shown to be a useful choice, because sometimes unphysical values are obtained. This happens when one of the compounds has nearly equal length as the compound of interest. Although this deviation is not significant for the combinatorial term, because the SG-term suppresses this by the ϕ_k/θ_k term, it is not acceptable for the residual term. For the time being the proposed procedure for defining the number of nearest neighbors works well as shown by the results in this work. It is hypothesized that the strong deviation is a result of Guggenheim's entropy model, where a single lattice coordination number was used. It

Table 6

Absolute average deviation in excess enthalpy: $100 \times |1 - H_{calc}^{ex}/H_{exp}^{ex}|$ of the various COSMOSPACE (CS) models.

Binary H^{ex}	T	NPS	CS(Original)	CS(New)	CS(New)	
Alkane - Alcohol	°C	—	Bondi	Bondi	Rowland	lit.
Hexane - Ethanol	40	21	15	9	9	[27]
Heptane - Ethanol	30	26	14	14	15	[35]
Nonane - Ethanol	25	26	14	21	22	[34]
Average			14	15	15

Fig. 10. Fraction of hydroxyl surface to total surface area as function of alkane number. From bottom to top: Lattice-free COSMOSPACE model based on Bondi's (dashed curve) and Rowland's (dashed-dotted curve) set of van der Waals radii, geometric model using on Bondi's (long dashed curve) and Rowland's (dashed-double dotted curve) set of van der Waals radii, and original COSMOSPACE model (solid curve).

Table 7

Comparison of the various models.

Model	Combinatorial term and parameters	Residual term and parameters	Lattice model? ^a
UNIQUAC	FH-SG ^b , $Q = zq/2$	1 st order solution of QCT ^c , $Q = q$	C: Yes
COSMOSPACE	$z = 10, A_{ref}, V_{ref}$ ^g FH-SG, $Q = zq/2$	Self-consistent solution of QCT	R: Yes C: Yes
	UNIQUAC parameters	= CS-solution ^d , $Q = n = 2q$	R: No
COSMO-RS	Generic model with FH and SG factors	CS-solution ^e , $Q = n = A_{COSMO}/a_{eff}$ and σ -profiles	C: Yes ^f R: No
COSMOSAC	$\lambda_0, \lambda_1, \lambda_2$ ^g FH-SG, $Q = zq/2$ $z = 10, A_{ref}, V_{ref}$ ^h	$a_{eff} = 7.1 \text{ \AA}^2$ CS-solution ^e , $Q = n = A_{COSMO}/a_{eff}$ and σ -profiles	C: Yes R: No
This paper	FH-SG $Q = \frac{(1-\phi/x)}{(1-\phi/\theta)}$ no parameters	$a_{eff} = 7.5 \text{ \AA}^2$ revised UNIQUAC residual: $Q = (1 - \phi/x)/(1 - \phi/\theta)$ revised CS residual $n = (1 - \phi/x)/(1 - \phi/\theta)$	C: No R: No

a C = Combinatorial term, R = Residual term. b FH-SG = Flory-Huggins with Staverman-Guggenheim correction. c QCT = Guggenheim's Quasi-Chemical Theory. d Homogeneous and non-homogeneous double binary solution. e General CS-solution. f Contains FH-SG terms, therefore debatable. g Parameters for the combinatorial term in COSMOTerm C1.2. h Reference area as given in Table 1, $V_{ref} = 25.17 \text{ \AA}^3$.

seems that the Guggenheim model can be improved by introducing a molecule dependent lattice number, which is invariant for concentration, and to define the area fraction, θ_k , by $z_k q_k$ instead of q_k .

Appendix A. Analysis of the number of nearest neighbors

In this appendix we demonstrate the usefulness of Eq. (14) by applying it to particular geometric cases.

Appendix A.1. Chains of repeating units in a lattice

Lattices are in theory made out of space-filling structures. Table A.8 summarizes the volume and area of some space-filling structures as function of the number of units (N) in the chain structure. We also included the case of a chain of touching spheres.

Table A.8

Volume and surface area of a chain of structures.

Repeating unit	Volume	Area
Tetrahedron	$N \frac{6}{\sqrt{2}} L^3$	$(N+1) \frac{\sqrt{2}}{2} L^2$
Cube	$N L^3$	$(2N+1) 2 L^2$
Dodecahedron	$N \frac{15+7\sqrt{5}}{4} L^3$	$(5N+1) \frac{3}{8} \sqrt{25+10\sqrt{5}} L^2$
Touching spheres	$N \frac{\pi}{6} D^3$	$N \pi D^2$

In these expressions L is the length of a side of the regular polyhedron, and D the sphere diameter. The mole, area and volume fractions of mixtures of chains of one type of structure have the general forms

$$x_j = \frac{N_j}{\sum_{j=1}^M N_j}, \quad (\text{A.1})$$

$$\theta_j = \frac{s N_j^2 + N_j}{\sum_{j=1}^M (s N_j^2 + N_j)}, \quad (\text{A.2})$$

and

$$\phi_j = \frac{N_j^2}{\sum_{j=1}^M N_j^2}, \quad (\text{A.3})$$

where the parameter s , which denotes the quadratic contribution of the surface area, is 1, 2, and 5 for respectively, a tetrahedron, a cube, and a dodecahedron. Substitution of A.1, A.2, A.3 into Eq. (14) gives for Q

$$Q(N_k) = 1 + s N_k, \quad (\text{A.4})$$

which is concentration-independent. Eq. (A.4) proofs that for the calculation of the combinatorial activity coefficient of a mixture of space-filling chains, of which the beads are made of tetrahedra, cubes, or dodecahedra only, the Staverman-Guggenheim correction term is needed. For a mixture of chains of touching hard-spheres, the situation is different, as in this case the ratio of volume and surface fraction is always unity; ($\theta_j/\phi_j = 1$). This implies that in Eq. (10) the Staverman-Guggenheim correction term becomes zero, and that only the Flory-Huggins term survives.

Appendix A.2. Linear and branched molecules

Eq. (14) was derived by using the connectivity relation of linear and branched structures. Here we show how Eq. (14) leads to expressions for the number of nearest neighbors as function of the number of repeating units. We start with a general formula for the volume and the area of a molecule containing N_k repeating groups and an end group.

$$V_k(N_k) = V_0(1 + \alpha N_k), \quad (\text{A.5})$$

$$A_k(N_k) = A_0(1 + \beta N_k). \quad (\text{A.6})$$

The constants, A_0 and V_0 define the size of the end group, while α and β define the ratio of the volumes of a repeating group and the chain end group. Substitution into the equations for the volume and area fraction yield for mixtures that consists of molecules of one class

$$\phi_k = \frac{x_k(1 + \alpha N_k)}{\sum_{j=1}^M x_j(1 + \alpha N_j)}, \quad (\text{A.7})$$

$$\theta_k = \frac{x_k(1 + \beta N_k)}{\sum_{j=1}^M x_j(1 + \beta N_j)}. \quad (\text{A.8})$$

As a result we obtain

$$Q_k = \frac{1 - \frac{\phi_k}{x_k}}{1 - \frac{\phi_k}{\theta_k}} = \frac{1 - \frac{(1 + \alpha N_k)}{\sum_{j=1}^M x_j(1 + \alpha N_j)}}{1 - \frac{(1 + \alpha N_k)}{\sum_{j=1}^M x_j(1 + \beta N_j)}} = \frac{\alpha}{\alpha - \beta} (1 + \beta N_k). \quad (\text{A.9})$$

This result explains why Sayegh and Vera [17,23] found a linear relation between the van der Waals volume of molecules and the number of nearest neighbors.

As example we evaluate the binary propane - tetradecane the calculation of the number of nearest neighbors using Eq. (14). According to Bondi's table [22] propane has $V_{vdW} = 37.57 \text{ cm}^3/\text{mol}$ and $A_{vdW} = 5.59 \text{ cm}^2/\text{nmol}$, and tetradecane $V_{vdW} = 150.10 \text{ cm}^3/\text{mol}$ and $A_{vdW} = 20.44 \text{ cm}^2/\text{nmol}$. With these values we can calculate the volume and surface area fractions, and subsequently the number of nearest neighbors for each alkane by applying Eq. (14). The step-by-step results are given in Table A.9.

Table A.9

Step-by-step calculation of the number of nearest neighbors of propane(1) and tetradecane(2) using Eq. (14)

x_1	ϕ_1	θ_1	ϕ_1/x_1	ϕ_1/θ_1	ϕ_2/x_2	ϕ_2/θ_2	Q_1	Q_2	zq_1	zq_2
0.50	0.200192	0.214752	0.4004	0.9322	1.600	1.019	8.844	32.34	17.69	64.68
0.75	0.077008	0.083545	0.3080	0.9218	1.231	1.231	8.844	32.34	17.69	64.68

The results of the other linear alkanes are shown in Fig. 1 of this work. Already with the two values from the above example we can write for the n-alkanes

$$zq_k = 64.68 + \frac{64.68 - 17.69}{14 - 3} (N_C - 14) = 4.87 + 4.27N_C \quad (\text{A.10})$$

This is equation (29) in this work. The above demonstrates that for a group of linear molecules there is a linear relation between the number of repeating groups N_k and the number of nearest neighbors Q_k , which can be derived without the definition of lattice coordination number, a reference area and volume.

Appendix B. Corrected residual activity coefficient of the UNIQUAC model

The original UNIQUAC model contains some inconsistencies. In order to understand this we need to go back to the concept of Wilson [37], who stated that molecules in a fluid are not distributed randomly, but tend to cluster more around molecules for which they have a higher affinity. This leads to the concept of local composition. Within this concept the first shell around a central molecule has a different composition than the average concentration of the bulk. The local mole fractions of compounds 1 and 2 around molecule 1 and those around molecule 2 are defined by the quantities x_{11} , x_{21} , x_{12} and x_{22} , respectively. A general expression for the local mole fractions x_{ij} follows from the situation where molecules are distributed randomly. This is certainly the case when the temperature is very high. We denote the temperature of the random case by T_{ra} . In the random configuration all molecules are interacting with same, "random" interaction energy U_{ra} . When we move molecule j from the random system to a system where it has a local configuration and interacts with central molecule i and energy U_{ji} , the ratio of the 2 mol fractions can be quantified by the energy difference as

$$x_{ji} = x_j \exp\left(-\left[\frac{U_{ji}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right). \quad (\text{B.1})$$

Subsequently, the ratio of local mole fraction of molecules 1 and 2 around central molecule 1 can be defined as

$$\frac{x_{21}}{x_{11}} = \frac{x_2 \exp\left(-\left[\frac{U_{21}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right)}{x_1 \exp\left(-\left[\frac{U_{11}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right)} = \frac{x_2}{x_1} \exp\left(-\frac{U_{21} - U_{11}}{RT}\right). \quad (\text{B.2})$$

While Wilson implicitly assumed that all molecules have the same size, Abrams and Prausnitz [6] later took into account that the molecules are different in size and they used surface fractions instead of mole fractions. Hence the excess energy of mixing two pure compounds was defined as

$$U^E = \frac{z}{2} [N_1 q_1 \Theta_{21} (U_{21} - U_{11}) + N_2 q_2 \Theta_{12} (U_{12} - U_{22})], \quad (\text{B.3})$$

where θ_{ji} is the local surface fraction of compound j around central

molecule i in the mixture. We note that the factor zq_j , i.e. the number of nearest neighbors, indicates that the quantity U_{ij} is the interaction energy between one side of compound i and one side of compound j . Further we observe that the lattice coordination number is identical for all types of molecules; $z_1 = z_2 = z$, which is set to 10 in the original UNIQUAC model. However, this approximation made by the developers of UNIQUAC is not necessary for the

residual part, as we will demonstrate below. In fact the approximation $z = 10$ made for the original UNIQUAC, is one of the root causes for systematic deviation in systems where molecules have a lower or higher coordination number. However, the approximation of $z = 10$ is not necessary, as we will demonstrate below. Let us first recall the quantity that denotes half the number of nearest neighbors of molecule j . This is identical to Eq. (14), but now we also include the possibility that the lattice coordination number is different for each compound

$$Q_j = \frac{z_j q_j}{2}. \quad (\text{B.4})$$

We mention that summation of this quantity equals to the total number of interacting pairs. With Eqs. (B.4) and (B.3) can be written as

$$U^E = Q_1 N_1 \Theta_{21} (U_{21} - U_{11}) + Q_2 N_2 \Theta_{12} (U_{12} - U_{22}). \quad (\text{B.5})$$

Here we implicitly assumed that the lattice coordination number of the pure compound is not affected by the mixture. Subsequently we define the local surface fractions. In line with Wilson's concept it means that we unpair from the random configuration one side of molecule j from one side of molecule i , this involves energy U_{ra} , and pair the two in the local configuration where molecule i is the central molecule. This local configuration requires energy U_{ji} . Applying the same steps as in the Wilson model the ratio of the local and random surface fractions can be written as

$$\frac{\Theta_{ji}}{\Theta_{ii}} = \frac{\theta_j}{\theta_i} \exp \left[-\frac{U_{ji} - U_{ii}}{k_B T} \right]. \quad (\text{B.6})$$

This equation differs from the original UNIQUAC equation, where the factor $z/2$ was used in the exponent. The quantity z can not be present in the exponent, because we place one side of a lattice cell from a molecule in random configuration in contact with one side of a lattice cell of the central molecule in the local configuration. The other cell sides of the molecule, which are also taken from the random configuration, can be placed to any other cell side, and do not take part in this calculation step! For convenience we define the interaction energy difference of unequal pairs and equal pairs on a temperature scale by

$$T_{jk} = \frac{U_{jk} - U_{kk}}{k_B}. \quad (\text{B.7})$$

This gives the following two relations

$$\frac{\Theta_{21}}{\Theta_{11}} = \frac{\theta_2}{\theta_1} \exp \left[-\frac{T_{21}}{T} \right], \quad (\text{B.8})$$

$$\frac{\Theta_{12}}{\Theta_{22}} = \frac{\theta_1}{\theta_2} \exp \left[-\frac{T_{12}}{T} \right]. \quad (\text{B.9})$$

The sum of the local surface fractions around a central molecule are unity, as imposed by the full occupancy boundary condition

$$\Theta_{21} + \Theta_{11} = 1, \quad (\text{B.10})$$

and

$$\Theta_{12} + \Theta_{22} = 1, \quad (\text{B.11})$$

from which follows that

$$\Theta_{21} = \frac{\theta_2 \exp \left[-\frac{T_{21}}{T} \right]}{\theta_1 + \theta_2 \exp \left[-\frac{T_{21}}{T} \right]}, \quad (\text{B.12})$$

and

$$\Theta_{12} = \frac{\theta_1 \exp \left[-\frac{T_{12}}{T} \right]}{\theta_2 + \theta_1 \exp \left[-\frac{T_{12}}{T} \right]}. \quad (\text{B.13})$$

Substitution of Eqs. (B.7, B.12 and B.13) into Eq. (B.5) yields

$$\frac{U^E}{k_B} = Q_1 N_1 T_{21} \frac{\theta_2 \exp \left[-\frac{T_{21}}{T} \right]}{\theta_1 + \theta_2 \exp \left[-\frac{T_{21}}{T} \right]} + Q_2 N_2 T_{12} \frac{\theta_1 \exp \left[-\frac{T_{12}}{T} \right]}{\theta_2 + \theta_1 \exp \left[-\frac{T_{12}}{T} \right]}. \quad (\text{B.14})$$

The Helmholtz energy is obtained from the integral

$$\frac{A^E}{k_B T} = - \int_0^1 \frac{U^E}{k_B} d\frac{1}{T}. \quad (\text{B.15})$$

Substitution of Eq. B into Eq. (B.15) gives

$$\frac{A^E}{k_B T} = -Q_1 N_1 \ln \left[\theta_1 + \theta_2 \exp \left[-\frac{T_{21}}{T} \right] \right] - Q_2 N_2 \ln \left[\theta_2 + \theta_1 \exp \left[-\frac{T_{12}}{T} \right] \right]. \quad (\text{B.16})$$

Generalization of this binary case to a multicomponent system results in

$$\frac{A^E}{k_B T} = - \sum_{j=1}^M Q_j N_j \ln \left[\sum_{i=1}^M \theta_i \tau_{ij} \right], \quad (\text{B.17})$$

where we defined

$$\tau_{ij} = \exp \left[-\frac{U_{ij} - U_{jj}}{k_B T} \right]. \quad (\text{B.18})$$

Using

$$\ln \gamma_k = \frac{\partial A^E}{\partial N_k}, \quad (\text{B.19})$$

yields the correct residual term for the UNIQUAC equation

$$\ln(\gamma_k^{\text{res,UQ}}) = Q_k \left[1 - \ln \left(\sum_{j=1}^M \theta_j \tau_{jk} \right) - \sum_{j=1}^M \frac{\theta_j \tau_{kj}}{\sum_{i=1}^M \theta_i \tau_{ij}} \right]. \quad (\text{B.20})$$

The crucial point in the original UNIQUAC derivation is the incorrect use of the lattice coordination number to define the ratio of the local and random surface fractions. Both Kontogeorgis and Folas [38] and Klamt et al. [12] mentioned that the UNIQUAC equation was inconsistent and that the exponent required $z = 2$. This value is unphysical, because in a 3-dimensional system the lowest value is $z = 4$ (i.e. a lattice made of tetrahedrons). In a reply on the comment made by McDermott and Ashton [39] also Maurer and Prausnitz [40] elaborated on the UNIQUAC inconsistency. Their solution was to add a constant of proportionality c into the denominator of the argument of the exponential in Eq. (B.18). The value for c is mixture dependent, and has an average value $c = 0.27$.

This implies that the factor $z/2$ in the Boltzmann factor of the original UNIQUAC model was reduced from 5 to 1.4. The deviation from unity could be caused by the choice of $z = 10$ as well as the use of a fixed reference area A_{ref} in the UNIQUAC model. But, as we have shown here, the origin for the aforementioned inconsistency and the need to use of a proportionality factor to obtain better results lies in the incorrect definition of the ratio of the local and random surface fractions. The correct derivation leads to Eq. (B.6). This also eliminates the use of a single value of $z = 10$.

References

- [1] M.L. Huggins, Solutions of long chain compounds, *J. Chem. Phys.* 9 (5) (1941) 440.
- [2] P.J. Flory, Thermodynamics of high polymer solutions, *J. Chem. Phys.* 10 (1) (1942) 51.
- [3] E.A. Guggenheim, Statistical thermodynamics of co-operative systems (a generalization of the quasi-chemical method), *Trans. Faraday Soc.* 44 (1948) 1007–1012.
- [4] E.A. Guggenheim, *Mixtures*, Clarendon Press Oxford, 1952.
- [5] A. Staverman, The entropy of high polymer solutions. generalization of formulae, *Recl. Trav. Chim. Pays-Bas* 69 (2) (1950) 163–174.
- [6] D.S. Abrams, J.M. Prausnitz, Statistical thermodynamics of liquid mixtures: a new expression for the excess Gibbs energy of partly or completely miscible systems, *AIChE J.* 21 (1) (1975) 116.
- [7] T. Magnussen, P. Rasmussen, A. Fredenslund, UNIFAC parameter table for prediction of liquid-liquid equilibria, *Ind. Eng. Chem. Process Des. Dev.* 20 (2) (1981) 331–339.
- [8] H. Kehiaian, Thermodynamik flüssiger Mischungen von Kohlenwasserstoffen mit verwandte Substanzen, *Berichte der Bunsengesellschaft für physikalische Chemie* 81 (10) (1977) 908–921.
- [9] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.* 21 (6) (1975) 1086.
- [10] A. Fredenslund, P. Rasmussen, J. Gmehling, Vapor-liquid Equilibria Using UNIFAC: a Group-contribution Method, Elsevier, 1977.
- [11] S.-T. Lin, S. Sandler, A priori phase equilibrium prediction from a segment contribution solvation model, *Ind. Eng. Chem. Res.* 41 (5) (2002) 899–913.
- [12] A. Klamt, G.J.P. Krooshof, R. Taylor, Cosmospace: alternative to conventional activity-coefficient models, *AIChE J.* 48 (10) (2002) 2332–2349.
- [13] R. Bronneberg, A. Pfennig, Moquac, a new expression for the excess gibbs energy based on molecular orientations, *Fluid Phase Equil.* 338 (2013) 63–77.
- [14] R.S. Rowland, R. Taylor, Intermolecular nonbonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii, *J. Phys. Chem.* 100 (18) (1996) 7384–7391.
- [15] I. Kikic, P. Alessi, P. Rasmussen, A. Fredenslund, On the combinatorial part of the UNIFAC and UNIQUAC models, *Can. J. Chem. Eng.* 58 (2) (1980) 253.
- [16] G.M. Kontogeorgis, A. Fredenslund, D.P. Tassios, Simple activity coefficient model for the prediction of solvent activities in polymer solutions, *Ind. Eng. Chem. Res.* 32 (2) (1993) 362–372.
- [17] J.H. Vera, S.G. Sayegh, G.A. Ratcliff, A quasi lattice-local composition model for the excess gibbs free energy of liquid mixtures, *Fluid Phase Equil.* 1 (1977) 113–135.
- [18] S. Wang, S.I. Sandler, C.-C. Chen, Refinement of cosmo-sac and the applications, *Ind. Eng. Chem. Res.* 46 (22) (2007) 7275–7288.
- [19] R. d. P. Soares, The combinatorial term for cosmo-based activity coefficient models, *Ind. Eng. Chem. Res.* 50 (5) (2011) 3060–3063.
- [20] R. Bronneberg, A. Pfennig, Improvement of the UNIQUAC combinatorial-entropy term by adjusting the standard segment, *J. Chem. Eng. Data* 56 (4) (2011) 1305–1310.
- [21] K. Egner, J. Gaube, A. Pfennig, Gequac, an excess gibbs energy model describing associating and nonassociating liquid mixtures by a new model concept for functional groups, *Fluid Phase Equil.* 158 (1999) 381–389.
- [22] A. Bondi, *Physical Properties of Molecular Crystals, Liquids, and Glasses*, John Wiley, 1968.
- [23] S. Sayegh, J. Vera, Lattice-model expressions for the combinatorial entropy of liquid mixtures: a critical discussion, *Chem. Eng. J.* 19 (1980) 1–10. An International Journal of Research and Development.
- [24] S. Batsanov, Van der waals radii of elements, *Inorg. Mater* 37(9).
- [25] L. Pauling, in: *The Nature of the Chemical Bond*, third ed., Cornell Univ, Ithaca, 1960. Bond lengths used in this paper: $L_{\text{HC}}=1.10 \text{ \AA}$, $L_{\text{CC}}=1.54 \text{ \AA}$, $L_{\text{OC}}=1.43 \text{ \AA}$, and $L_{\text{OH}}=0.96 \text{ \AA}$.
- [26] A. Iguchi, Kagaku Sochi, 20, 1978, pp. 66–68.
- [27] S. O'Shea, R. Stokes, Activity coefficients and excess partial molar enthalpies for (ethanol + hexane) from 283 to 318 K, *J. Chem. Therm.* 18 (7) (1986) 691–696.
- [28] M. Diaz-Pena, D. Cheda, *An. Quim.* 66 (1970) 747–755.
- [29] M. Ronc, G.R. Ratcliff, Measurement of vapor-liquid equilibria using a semi-continuous total pressure static equilibrium still, *Can. J. Chem. Eng.* 54 (4) (1976) 326.
- [30] C. Berro, M. Rogalski, A. Peneloux, Excess gibbs energies and excess volumes of 1-butanol-n-hexane and 2-methyl-1-propanol - n-hexane binary systems, *J. Chem. Eng. Data* 27 (3) (1982) 352–355.
- [31] H. Wolff, A. Shadiakhy, The vapor pressure behavior and association of mixtures of 1-hexanol and n-hexane between 293 and 373 K, *Fluid Phase Equil.* 7 (3–4) (1981) 309–325.
- [32] C. Berro, M. Rogalski, A. Peneloux, A new ebulliometric technique. vapour-liquid equilibria in the binary systems ethanol-n-heptane and ethanol-n-nonane, *Fluid Phase Equil.* 8 (1) (1982) 55–73.
- [33] T. Hofman, B. Barbes, C. Casanova, Thermodynamic properties of n-alcohol- α -n-alkane mixtures. a comparative study of some group contribution theories, *J. Chem. Soc., Faraday Trans.* 92 (19) (1996) 3565.
- [34] J.J. Christensen, R.M. Izatt, B.D. Stitt, R.W. Hanks, K.D.J. Williamson, Excess enthalpies of seven n-nonane + alcohol mixtures at 298.15 K, *J. Chem. Thermodyn.* 11 (11) (1979) 1029–1034.
- [35] J. Shatas, M. Abbott, H. van Ness, Excess thermodynamic functions for ternary systems ii chloroform-ethanol-n-heptane at 50 °C, *J. Chem. Eng. Data* 20 (1975) 406–409.
- [36] H. Kehiaian, Thermodynamics of binary liquid organic mixtures, *Pure Appl. Chem.* 57 (1) (1985) 15–30.
- [37] G. Wilson, Vapor-liquid equilibrium. xi. a new expression for the excess free energy of mixing, *J. Am. Chem. Soc.* 86 (2) (1964) 127–130.
- [38] G.M. Kontogeorgis, G.K. Folas, *Thermodynamic Models for Industrial Applications: from Classical and Advanced Mixing Rules to Association Theories*, Wiley, 2010.
- [39] C. McDermott, N. Ashton, Note on the definition of local composition, *Fluid Phase Equil.* 1 (1) (1977) 33–35.
- [40] G. Maurer, J. Prausnitz, On the derivation and extension of the UNIQUAC equation, *Fluid Phase Equil.* 2 (2) (1978) 91–99.