

COSMOSPACE: An Alternative to Conventional Activity

Coefficient Models

Andreas Klamt^{1,5}, Gerard J. P. Krooshof², and Ross Taylor^{3,4}

1) COSMOlogic GmbH&CoKG, Leverkusen, Germany

2) DSM Research, Geleen, The Netherlands

3) Department of Chemical Engineering, Clarkson University, Potsdam, New York, USA

4) Department of Chemical Technology, University of Twente,

7500 AE Enschede, The Netherlands

5) Corresponding author: e-mail: andreas.klamt@cosmologic.de

Abstract

In this article, an analytical solution, that we call COSMOSPACE, to the statistical thermodynamics of a model of pair-wise interacting surfaces, is presented. In an implicit form, this solution was initially developed for the *a priori* prediction model COSMO-RS. A comparison of COSMOSPACE with UNIQUAC and with the quasi-chemical theory of Guggenheim reveals the conditions under which the models yield similar results and when they differ very considerably. It is shown that COSMOSPACE is in extremely good agreement with Monte-Carlo simulations for some lattice fluids (where UNIQUAC is particularly poor). The ability of COSMOSPACE to provide good fits to experimental data is shown for 3 binary mixtures including ethanol - cyclohexane where UNIQUAC incorrectly predicts a liquid-liquid phase separation.

1. Introduction

The accurate description of the thermodynamic behavior of liquids and liquid mixtures is of central importance in chemical engineering. Good thermodynamic models are of some economic significance since large-scale process simulation is routinely used in chemical process industry. The *a priori* predictive model COSMO-RS (Klamt, 1995, 1998; Klamt et al., 1998; Klamt and Eckert, 2000) and its implementation in COSMO*therm* (COSMOlogic, 1999) permits the efficient calculation of liquid-liquid and vapor-liquid equilibria based solely on quantum chemical calculations for the involved chemical compounds. This model enables us to predict equilibrium in situations where measurements are not available, even for mixtures of rare, unstable and complicated chemical species.

Although *a priori* predictive models such as COSMO-RS are starting to be used in industry, empirical models fitted to a small number of available experimental data are more routinely applied in process simulation, because such models are more accurate than the current generation of predictive models. For mixtures of non-electrolyte liquids far below the critical region the UNIQUAC equation, introduced by Abrams and Prausnitz (1975) as a generalization of the quasi-chemical equation (QCE) of Guggenheim (1952) (see, for example, Ben Naim, 1987), is widely employed. Other models like those of Flory and Huggins, Wilson, Renon and Prausnitz, van Laar, Scatchard, Hamer and Hildebrand, and Margules (see Walas (1985) for details) can be obtained as special cases of the UNIQUAC equation (Abrams and Prausnitz, 1975). In addition, the group contribution model (GCM) UNIFAC is based on the thermodynamic model of UNIQUAC (see Walas, 1985).

Only recently has the potential of COSMO-RS as a flexible G^E -model become apparent, together with an analytical solution for the binary case. In this paper we present this model,

referred to in what follows as COSMOSPACE (COSMO Surface Pair Activity Coefficient Equation).

2. COSMOSPACE

2.1 General theory of interacting surface pair models

Consider an ensemble of molecules that represents a liquid. Assuming that there is no correlation between molecular interactions and geometrical restraints, the partition sum Z of an ensemble can be factorized into two contributions:

$$Z = Z^C Z^R \quad (1)$$

The first factor on the right hand side of Eq. 1 is called the combinatorial factor. It is the partition sum of the equivalent ensemble of molecules that interact only through steric restraints. The combinatorial factor takes into account all size and shape effects of the molecules. There is no exact expression for Z^C , but by fitting of reasonable models to thermodynamic data of alkane-alkane mixtures reasonable approximations have been derived.

The Staverman-Guggenheim (SG) expression

$$\ln Z_{SG}^C = \sum_i \left[x_i \ln \left(\frac{\Phi_i}{x_i} \right) + \frac{z}{2} x_i q_i \ln \left(\frac{\Theta_i}{\Phi_i} \right) \right] \quad (2)$$

for example, is used in UNIQUAC. The variables x_i , Φ_i , and Θ_i denote the mole-, volume-, and surface-fractions of species i , respectively, with the volume and surface fractions defined in terms of the relative volume, r_i , and relative area, q_i , respectively. Both r_i and q_i are based on van der Waals cavities and normalized in a consistent way. r and q values have been tabulated for many molecules (see Gmehling and Onken, 1977). The UNIFAC group contribution method can also be used to estimate the molecular r and q parameters (see, Walas, 1985; Sandler, 1999). The coordination number, z , commonly is assumed to be 10. For mixtures of compounds that differ by less than a factor 5 in size, the combinatorial factor

usually is of moderate importance and the SG approximation can be considered to be sufficiently accurate for the purposes of this paper.

The second factor, Z^R , in Eq. 1 is called the "residual" contribution and it arises from the interactions between molecules. Despite the name, it is the more important contribution in most liquids. The basic assumption of surface pair interaction models is that residual - i.e., non-steric - interactions can be described by pairs of geometrically independent surface segments. The residual contribution is just the partition sum of an ensemble of pair-wise interacting independent surface segments.

If we consider a system having ν different types of surface, the total number of surface segments n_i on a molecule i is given by

$$n_i = \sum_{\nu} n_i^{\nu} = \frac{q_i}{a_{eff}} \quad (4)$$

where q_i is the total surface area of molecule i and n_i^{ν} denotes the numbers of segments of type ν on molecule i . a_{eff} is the size of the surface segments. There is no simple way to define a_{eff} from first principles and it must be considered to be an adjustable parameter.

If $N_i = N x_i$ is the number of molecules of species i in the system, then the total number, n , of segments in the system is given by

$$n = \sum_i N_i n_i \quad (5)$$

In the same way, the number of segments of type ν is given by

$$n^{\nu} = \sum_i N_i n_i^{\nu} \quad (6)$$

The relative number of segments of type ν is

$$\Theta^v = \frac{n^v}{n} \quad (7)$$

From elementary statistical thermodynamics we know that the Gibbs free energy of the system is

$$G = -kT \ln Z \quad (8)$$

The chemical potential of species i in the mixture is given by

$$\mu_i = \frac{\partial G}{\partial N_i} = -kT \frac{\partial \ln Z}{\partial N_i} \quad (9)$$

In view of our earlier assumption that the residual partition sum Z^R depends only on the segment composition we have

$$\mu_i^R = -kT \frac{\partial \ln Z^R}{\partial N_i} = -kT \sum_{\nu} \frac{\partial \ln Z^R}{\partial n^{\nu}} \frac{\partial n^{\nu}}{\partial N_i} = -kT \sum_{\nu} \frac{\partial \ln Z^R}{\partial n^{\nu}} n_i^{\nu} = \sum_{\nu} n_i^{\nu} \mu^{\nu} \quad (10)$$

where μ^{ν} is the pseudo-chemical potential of a segment of type ν in the ensemble of interacting surface segments. The activity coefficient, γ_i , of species i is defined by

$$kT \ln \gamma_i = \mu_i - \mu_{ii} \quad (11)$$

where μ_{ii} denotes the chemical potential of compound i in a system with $x_i = 1$. In the same way we have

$$kT \ln \gamma^{\nu} = \mu^{\nu} - \mu^{\nu\nu} \quad (12)$$

for the activity coefficient of a segment of kind ν , with $\mu^{\nu\nu}$ being the segment chemical potential at $\Theta^{\nu} = 1$. If we define μ_i^{ν} and γ_i^{ν} to be the chemical potential and activity coefficient of a segment ν in an ensemble of pure compound i , respectively, and apply these definitions to Eq. 10, we obtain

$$\ln \gamma_i^R = \frac{\mu_i^R - \mu_{ii}^R}{kT} = \sum_{\nu} n_i^{\nu} \frac{\mu^{\nu} - \mu_i^{\nu}}{kT} = \sum_{\nu} n_i^{\nu} \frac{(\mu^{\nu} - \mu^{\nu\nu}) - (\mu_i^{\nu} - \mu_i^{\nu\nu})}{kT} = \sum_{\nu} n_i^{\nu} (\ln \gamma^{\nu} - \ln \gamma_i^{\nu}) \quad (13)$$

for the residual part of the activity coefficient of compound i . With the development of Eq. 13 we have reduced the thermodynamics of the system of chemical compounds to the thermodynamics of an ensemble of pair-wise interacting surface segments.

We now consider this ensemble in more detail. Let μ and ν denote different kinds of surface segment and let $u_{\mu\nu}$ be the interaction energy of a pair $\mu\nu$. For a given total pairing, P , of the ensemble the total interaction energy E_{tot} can be written as a sum of pair-wise interactions of segments:

$$E_{tot}(P) = \sum_{\mu\nu} p_{\mu\nu}(P) u_{\mu\nu} \quad (14)$$

In this equation $p_{\mu\nu}(P)$ is the total number of pairs of kind $\mu\nu$. The partition sum Z^R of this ensemble is

$$Z^R = \sum_P \exp \left\{ - \frac{\sum_{\mu\nu} p_{\mu\nu}(P) u_{\mu\nu}}{kT} \right\} \quad (15)$$

where P samples all possible total pairings of the segments.

2.2 The COSMOSPACE Solution

A derivation of an analytical solution for the segment activity coefficients γ^v resulting from the partition sum Z^R of an ensemble of pair-wise interacting segments is given in Appendix A, resulting in the following expression:

$$\frac{1}{\gamma^v} = \sum_{\mu} \tau_{\mu\nu} \Theta^{\mu} \gamma^{\mu} \quad (16)$$

where the interaction parameter $\tau_{\mu\nu}$ is defined by

$$\tau_{\mu\nu} = \exp \left[- \frac{u_{\mu\nu} - \frac{1}{2}(u_{\mu\mu} + u_{\nu\nu})}{RT} \right] \quad (17)$$

For physical consistency we assume that the interaction energies of the segments are given by a symmetric matrix: i.e. $u_{\mu\nu} = u_{\nu\mu}$. Thus: $\tau_{\mu\nu} = \tau_{\nu\mu}$. The derivation in Appendix A is simpler and more convenient than that of Klamt (1995) for the thermodynamics of an ensemble of pair-wise interacting object segments.

Equations (16) are the general COSMOSPACE equations and can be solved iteratively by simple repeated substitution, starting with the assumption of $\gamma^i = 1$ on the right hand side. The result automatically satisfies the Gibbs-Duhem relation because Eq. (16) is a solution to the basic thermodynamic equations. However, for completeness, we demonstrate the thermodynamic (Gibbs-Duhem) consistency of the solution of the general COSMOSPACE model in Appendix B. In Appendix D an explicit formula for the derivatives of the activity coefficients with respect to any state variable (P , T , or the composition variables θ^i) is given. The availability of an analytical derivative is of some practical significance if the COSMOSPACE model is to be used in a process simulator.

2.3 COSMOSPACE Equations for a Binary

It is helpful to consider the case of a mixture that contains only two kinds of segments (A and B in what follows). Here the COSMOSPACE equations become:

$$1 = \Theta^A \gamma^{A^2} + \Theta^B \gamma^B \gamma^A \tau_{AB} \quad (18a)$$

$$1 = \Theta^B \gamma^{B^2} + \Theta^A \gamma^A \gamma^B \tau_{AB} \quad (18b)$$

An analytical solution can be found using standard algebraic methods as shown in Appendix C, with the result:

$$\gamma^A = \sqrt{\frac{1}{\Theta^A} + \frac{1 - \sqrt{1 + 4\Theta^A \Theta^B \omega}}{2\omega \Theta^{A^2}}} \quad (19)$$

where $\omega = \tau_{AB}^{-2} - 1$. The activity coefficient for surface segments of kind B is found by interchanging indices ($A \leftrightarrow B$) in Eq. 19. Note, that γ^A takes the value $1/\tau_{AB}$ for $\Theta^A = 0$.

In Tables 1 and 2 we show the COSMOSPACE equations for what we refer to as a *double binary* mixture; that is a mixture of two different chemical species that involves two kinds of surface segment. In Table 1 we provide the COSMOSPACE equations for what we shall call a homogeneous double binary in which each species has only one type of surface segment.

The numbers of each type of segment is $n_i^v = 2q_i$, if $i = v$, and $n_i^v = 0$, if $i \neq v$. The formulae for the limiting segment activity coefficients are particularly simple in this case (see Table 1) and the residual part of the species activity coefficient is just the corresponding segment activity coefficient. The combinatorial part is given by the combinatorial part of UNIQUAC (Abrams and Prausnitz, 1975; Walas, 1985). A further simplification results when the surface area parameters are equal in which case $\Theta^i = x_i$.

A more complicated case – the non-homogeneous double binary – is summarized in Table 2. In this second case both species involve both different kinds of segment. Multicomponent mixtures that are modelled with just two different types of surface segment represent a minor extension of the model summarised in Table 2. It is, of course, possible to develop even more sophisticated models simply by incorporating even more kinds of surface segment in the model. It should be clear that the COSMOSPACE model is, therefore, inherently more flexible than any conventional activity coefficient model because, even for a binary mixture, the possibility exists to use more than the usual two (or three) interaction parameters per pair of components.

2.3 Comparison of COSMOSPACE with Guggenheim's Quasi-Chemical Theory

The quasi-chemical treatment in Guggenheim's classic treatment of lattice fluids gives the (residual) activity coefficient of compound A in a binary lattice mixture of compounds 1+2 (Equation 4.11.6 of Guggenheim, 1952):

$$\gamma_1^R = \left\{ \frac{\beta + 1 - 2(1 - x_1)}{(1 + \beta)x_1} \right\}^{\frac{z}{2}} \quad (20)$$

where

$$\beta = \sqrt{1 + 4x_1(1 - x_1)\omega} \quad (21)$$

and

$$\omega = \eta^2 - 1 = e^{\frac{2w}{kT}} - 1 \quad (22)$$

In Guggenheim's treatment the molecules are assumed to be of similar size: thus, the q 's may be assumed to be equal in what follows. Now, if we define the following parameters by analogy with the above expressions:

$$\omega' = \frac{1}{\tau^2} - 1 = e^{-\frac{2U_{12}}{kT}} - 1 \quad (23)$$

$$\beta' = \sqrt{1 + 4x_1(1 - x_A)\omega'} \quad (24)$$

Equation (24) can be rearranged to give

$$\omega' = \frac{(\beta' - 1)(\beta' + 1)}{4x_1(1 - x_1)} \quad (25)$$

In terms of the above parameters the homogeneous double binary COSMOSPACE equation in Table 1 becomes:

$$\gamma_A = \sqrt{\frac{1}{x_1} + \frac{1 - \beta'}{2\omega'x_1^2}} = \sqrt{\frac{1}{x_1} + \frac{(1 - \beta')4x_1(1 - x_1)}{2x_1^2(\beta' - 1)(\beta' + 1)}} = \sqrt{\frac{(\beta' + 1) - 2(1 - x_1)}{x_1(\beta' + 1)}} \quad (26)$$

The residual activity coefficient is:

$$\gamma_1^{res} = \gamma_A^{2q_1} = \left(\frac{(\beta' + 1) - 2(1 - x_1)}{x_1(\beta' + 1)} \right)^{q_1} \quad (27)$$

This means that the simplest (homogeneous double binary) COSMOSPACE model is equivalent to Guggenheim's quasi-chemical treatment when $z = 2q_1 = n_1^A$ and when $\beta = \beta'$, which implies: $\omega/z = -U_{12}$.

We conclude that COSMOSPACE model is an exact model of lattice systems when $z = 1$. However, the model appears to give a surprisingly good description of lattice systems and of real off-lattice fluids (reality) for a wide range of parameters, as we shall show in Sections 4 and 5 below.

3. Comparison with UNIQUAC

Although derived from a lattice model, the QCE and the UNIQUAC model are based on the assumption that the interactions of molecules in the liquid state can be described by pair-wise interactions of segments of molecular surface. Thus, there exists a close relationship between COSMOSPACE and UNIQUAC that is more fully explored in this section.

3.1 The UNIQUAC Approximation

The UNIQUAC model incorporates interaction parameters of segments defined in a slightly different way:

$$\tau'_{\mu\nu} = \exp\left[-\frac{u_{\mu\nu} - u_{\mu\mu}}{RT}\right] \quad (28)$$

These interaction parameters generally are not symmetric with respect to μ and ν . The most severe approximation used in UNIQUAC consists of the replacement of the COSMOSPACE Eqns. (16) by

$$\frac{1}{\Psi^{\nu}} = \sum_{\mu} \tau'_{\mu\nu} \Theta^{\mu} \quad (29)$$

In this approximation Ψ^{ν} is a pseudo-activity coefficient (corresponding to the ratio $\Theta_{ii}^{(1)}/\Theta_i$ in Abrams and Prausnitz, 1975), i.e. a kind of availability coefficient that takes the part of the activity coefficient γ^{ν} in the COSMOSPACE model. Note that Eq. 29 can be considered as the analog of Eq. 16, in which the pseudo-activity coefficients Ψ^{μ} are set to unity on the right-hand side. This has the advantage that Ψ^{ν} is obtained explicitly, in contrast to γ^{ν} in the COSMOSPACE equations. This replacement of activities by concentrations is responsible for the inaccuracy of QCE models for strong interactions at high dilution, where the replacement of activity by concentration is most important. In fact, UNIQUAC was called a first approximation by its authors. A second order approximation could be obtained by using Ψ^{μ}

from the first order on the right hand side of Eq. 29, and so on. This procedure would converge to the COSMOSPACE equations, if continued indefinitely.

The pseudo-activity coefficients in Eq. 29 are not thermodynamically consistent. In UNIQUAC the excess Gibbs free energy of the systems is defined as

$$G^{ex} = RT \sum_{\nu} \Theta^{\nu} \ln \Psi^{\nu} \quad (30)$$

and the activity coefficients are derived by differentiation with respect to the numbers of segments:

$$\ln \gamma^{\kappa} = \frac{1}{RT} \frac{\partial NG^{ex}}{\partial N_{\kappa}} = \frac{\partial}{\partial N_{\kappa}} \sum_{\nu} N^{\nu} \ln \Psi^{\nu} = \ln \Psi^{\kappa} - \sum_{\nu} \frac{N^{\nu}}{\Psi^{\nu}} \frac{\partial}{\partial N_{\kappa}} \sum_{\mu} \tau'_{\mu\nu} \frac{N^{\mu}}{N} = \ln \Psi^{\kappa} - \sum_{\nu} \Psi^{\nu} \Theta^{\nu} \tau'_{\kappa\nu} + 1 \quad (31)$$

In UNIQUAC compounds usually are assumed to have only one type of surface segment. In the notation used in this article this means $n_i^{\nu} = 0$, if $i \neq \nu$. However, the thermodynamic model underlying UNIQUAC remains valid even if there is more than one kind of surface per compound. These non-homogeneous UNIQUAC equations are, in fact, used in group-contribution methods like UNIFAC (Fredenslund et al., 1975). Different functional groups in UNIFAC could be considered to represent different types of surface, although from a physical point of view, a single fragment like -OH actually represents extremely different surface types.

There are three further criticisms regarding the derivation of the UNIQUAC approach:

- The factor $\tau'_{\mu\nu}$ in Eq. 29 is intended to take into account the Boltzmann weight for the extra energy arising from $\mu\nu$ pairs, compared with pure $\mu\mu$ and $\nu\nu$ pairing. For the formation of two $\mu\nu$ pairs, we have to break one $\mu\mu$ pair and one $\nu\nu$ pair. The extra energy is $u_{\mu\nu} - (u_{\mu\mu} + u_{\nu\nu})/2$, and not $u_{\mu\nu} - u_{\nu\nu}$. Hence, the definition of the interaction parameter

in Eq. 28 is physically meaningless, and instead the definition given in Eq. 17 is more appropriate. Since the interaction parameters in UNIQUAC are fit to experimental data anyway, the only consequence of this shortcoming is that the asymmetric case of $\tau'_{\mu\nu} \neq \tau'_{\nu\mu}$, which is allowed in UNIQUAC, appears to be physically meaningless. We assume the interaction parameters in UNIQUAC to be identical with those of COSMOSPACE for the comparison below.

- The second inconsistency in the derivation of UNIQUAC concerns the number of independent surface segments. In the beginning of their derivation, Abrams and Prausnitz (1975) consider the number of surface segments on a molecule to be $z q_i$, where $z = 10$ is the coordination number and q_i is the surface area parameter of compound i . This corresponds to an effective contact area of $a_{eff} = 0.1$ in the units of the UNIQUAC surface parameters. However, by replacing $z/2 u_{\mu\nu}$ by a new interaction energy, and by the corresponding reduction of the number of segments originally appearing in Eq. (9) of their paper Abrams and Prausnitz (1975) convert a number factor into a scaling factor of interaction energies. This manipulation is of no importance for the interaction enthalpy considered in Eqs. (9) and (9a) of their work, but it reduces the number of degrees of freedom of the system by a factor $z/2$ and, therefore, shows up in the free energy. Hence, the number of surface segments really taken into account in UNIQUAC is only $2q_i$, corresponding to an effective segment area of $a_{eff} = 0.5$ area units.
- Finally, Abrams and Prausnitz (1975) introduced another inconsistency into the calculation of the Gibbs free energy. In their Eq. (22a) the Gibbs free energy is found from a sum of contributions $q_i \ln \Psi^i$ for each mole of component i . This gives the impression that they are summing one contribution $\ln \Psi^i$ per unit area. But due to the

previous finding that the real coordination number used in the residual part of UNIQUAC is 2, they should have used a contribution of $2q_i \ln \gamma^i$ for each molecule. We conclude that the logarithmic partition coefficients of the segments $\ln \gamma^i$ are half of the value given by Eq. 31.

Summarizing: we find that the UNIQUAC model can be considered to be an interacting surface pair model with a first order approximation regarding 'local composition' effects, if we consider the numbers of segments on compounds to be given by Eq. 4 with $n_i^i = 2q_i$, i.e. and $a_{eff} = 0.5$, and $n_i^\mu = 0$ for $i \neq \mu$, and with

$$\ln \gamma_{UNIQUAC}^\kappa = \frac{1}{2} \left(\ln \Psi^\kappa - \sum_{\nu} \Psi^\nu \Theta^\nu \tau'_{\kappa\nu} + 1 \right) \quad (32)$$

UNIQUAC is physically meaningful only in the symmetric, one-parameter case of $\tau = \tau'_{ij} = \tau'_{ji}$. One may also consider the effective area of an independent surface segment a_{eff} to be an additional fitting parameter because the present choice is rather arbitrary.

For the following analytical and numerical comparisons of COSMOSPACE and UNIQUAC we consider a binary system with only two kinds of segments so that we may use Eq. 19, together with the settings $\tau'_{ij} = \tau'_{ji} = \tau$ and $a_{eff} = 0.5$, that is necessary in order to make the models comparable at all.

3.2 Analytic comparisons

In the symmetric case where $\Theta^A = \Theta^B = 0.5$, we have $\gamma^A = \gamma^B = \gamma$ and Eqs. 18 and 32 reduce to

$$\ln \gamma_{COSMOSPACE} = \ln \gamma_{UNIQUAC} = -\frac{1}{2} \ln(0.5 + 0.5\tau) \quad (33)$$

Thus, we find that COSMOSPACE and UNIQUAC always coincide at $\Theta^A = \Theta^B = 0.5$.

More interesting is the case of infinite dilution activity coefficients, e.g. $\Theta^A = 0$ and $\Theta^B = 1$.

In this limit we find

$$\ln \gamma_{COSMOSPAC}^{1,\infty} = -\ln(\tau) \quad (34a)$$

for the chemical potential of compound 1 from the COSMOSPAC equation, while UNIQUAC yields

$$\ln \gamma_{UNIQUAC}^{1,\infty} = -\frac{1}{2}(\ln(\tau) + (\tau - 1)) \quad (34b)$$

Since $\ln(\tau) \cong \tau - 1$ for $|\tau - 1| \ll 1$, both models yield similar results for weakly interacting systems. However, UNIQUAC underestimates the infinite dilution activity coefficients of strongly interacting systems.

3.3 Consideration of different parameter combinations

Figure 1 shows the behavior of COSMOPAC and UNIQUAC at differing values of the interaction parameter τ . We see here that the behavior of the two models is almost identical in the range $x_1 > 0.1$. Only when we approach the infinite dilution region ($x_1 \rightarrow 0$) do we see distinct differences between the models. For values of $\tau < 1$ COSMOPAC shows a much stronger curvature close to infinite dilution, while for $\tau > 1$ COSMOPAC shows a more moderate behavior than UNIQUAC. Only for values of $\tau > 2.5$ do the two models appear to be quite different over the entire concentration range.

Figure 2 shows that the choice of a_{eff} has some influence on the $\ln \gamma$ curves. The interaction energy per surface area is considered to be constant for all of the curves in Figure 2 ($\ln(\tau) / a_{eff} = \ln(0.5) / 2$). This corresponds to a medium interaction parameter of $\tau = 0.5$ for the default case of $a_{eff} = 0.5$. While the starting slope near $x_1 = 1$ is the same for all curves due to the identical interaction energy per unit area, the curves depend strongly on the choice of a_{eff} in the range $0.2 < x_1 < 0.5$. At infinite dilution a segment of type 1 has no alternative to

having a partner of type 2, no matter how large are the different segments. Thus, the free energy in this limit is determined only by the interaction energy per unit area. Since this quantity is fixed, the curves should end up at the same point. Indeed, we see in Fig. 2 that all of the curves for the COSMOSPACE model end up in the same point at infinite dilution. UNIQUAC, however, shows different infinite dilution activity coefficients as a function of a_{eff} . This has to be an artifact that results from the quasi-chemical approximation.

As mentioned above, the value of $a_{eff} = 0.5$ presently used in UNIQUAC appears to be accidental. The original choice of a coordination number $z = 10$ corresponds to a value of $a_{eff} = 0.1$. As can be seen in Figure 2 this change would have some influence on the shape of the curves. From the COSMO-RS parameterization Klamt et al. (1998) and Klamt and Eckert (2000) found a value of about 0.07 nm^2 for a_{eff} . By correlating the UNIQUAC surface parameters against the corresponding surface areas of the COSMO cavities Klamt and Eckert (2001) found that a UNIQUAC surface unit corresponds to about 0.46 nm^2 of COSMO cavity. Hence, the effective segment size derived from the COSMO-RS fit is $a_{eff} = 0.07 \text{ nm}^2 / 0.46 \text{ nm}^2 = 0.15$ in UNIQUAC units, corresponding to about $z = 6.6$.

The fact that UNIQUAC and COSMOSPACE appear to be in agreement under certain conditions should not be taken to imply that they are always in agreement, as the following section on lattice fluids will show.

4. Modelling of Lattice Fluids

Wu, Cui and Donohue (1998) compared several local composition models to Monte-Carlo simulations for lattice mixtures. The models tested included: UNIQUAC, the AD model for lattice fluids of Aranovich and Donohue (1996), and the Born-Green-Yvon (BGY) model of Lipson (1991). The BGY model is equivalent to UNIQUAC with a coordination number $z=2$ in the Boltzmann exponent. The BGY model is, therefore, in agreement with the modifications of the UNIQUAC model that were discussed in Section 3 above.

Following Wu et al., we calculate the internal energy and the internal energy change of mixing from:

$$U \approx H = \left(\frac{\partial(G/T)}{\partial(1/T)} \right)_{V, N_i} \quad (35)$$

Expressions for the internal energy for the models tested by Wu et al. are given in their paper. For the homogeneous COSMOSPACE model we obtain a particularly simple expression (with $\tau_{ii} = 1$):

$$U = \frac{z}{2} \sum_i \sum_j x_i x_j \gamma_i \gamma_j \epsilon_{ij} \tau_{ij} \quad (36)$$

The mean field internal energy, U_{mf} , is given by:

$$U_{mf} = \frac{z}{2} \sum_i \sum_j x_i x_j \epsilon_{ij} \quad (37)$$

The (dimensionless) internal energy change of mixing ($\Delta U / RT$) is obtained by subtracting the ideal mixture internal energy from the internal energy calculated as given above.

For a cubic lattice mixture of monomers we may set all the q 's to unity, thus forcing the segment fractions to equal the appropriate mole fraction ($\Theta^i = x_i$). The coordination number, the number of nearest neighbors, is 6 in such a lattice mixture. Using specified values of the

interaction energies, $\varepsilon_{11}, \varepsilon_{12} = \varepsilon_{21}, \varepsilon_{22}$, for a number of such systems (in their notation $\varepsilon_{ij}/kT = u_{ij}/RT$ in ours), we calculate the COSMOSPACE interaction parameter from $\tau_{12} = \tau_{21} = \exp\left[-\left(u_{12} - \frac{1}{2}(u_{11} + u_{22})\right)/RT\right]$. The segment activity coefficients follow immediately from the expressions in Table 1 and the internal energy as given above.

Figures 3 and 4 show the internal energy change of mixing for lattice mixtures of monomers with properties as specified in the figure captions (these correspond to the systems in Figs. 2 and 4 of Wu et al., 1998). In addition to the results obtained with COSMOSPACE, we also show the results from our own Monte-Carlo simulations as well as those obtained using UNIQUAC (which here is identically equal to the Wilson model), the AD model and the BGY model. From the results shown here (which appear to be in complete agreement with those of Wu et al., 1998) we see that the original UNIQUAC is the worst of the models (when compared to the Monte Carlo simulations). The BGY method (which is the proper way to use UNIQUAC) is significantly better than UNIQUAC, but it is not a particularly good approximation to the Monte-Carlo simulations. It is impossible to distinguish between COSMOSPACE and the AD model in all cases shown here and both models are in extremely close agreement with the Monte Carlo simulations. Our results for two other systems, those in Figs. 1 and 3 of the paper of Wu et al. show equally good agreement between COSMOSPACE, the AD model and the Monte Carlo simulations.

Figures 5 and 6 shows the deviation in the activity coefficients predicted by COSMOSPACE and the BGY model from those obtained directly from the Monte-Carlo simulations using an addition to our MC code that allows us to evaluate the activity coefficients of the components. We see from these results that COSMOSPACE is in much better agreement with the Monte-

Carlo simulations than is the BGY model. We have not calculated the activity coefficients for the AD model since it is not a model for the excess Gibbs energy.

Figures 7 and 8 show the internal energy change of mixing for polymer/solvent mixtures, again with properties as specified in the figure captions (these are the systems in Figs. 7 and 9 of Wu et al., 1998). The coordination number for the pure solvent is 6, but the coordination number for the polymer is $4 + 2/C$, where C is the polymer chain length. The coordination number for the mixture is the volume fraction weighted sum of these numbers. As before, we see that COSMOSPACE and the AD method are in very close agreement with each other and are very close to the Monte Carlo simulations. The BGY model also gives reasonable predictions. As before. UNIQUAC is by far the worst of the models. Note here that COSMOSPACE (and the other analytical models) can be used to calculate the internal energy at any volume fraction whereas it appears that the Monte Carlo method fails at some value of the volume fraction above 0.8 (see Figs. 7 and 9 in Wu et al., 1998).

It is worth asking the question: are COSMOSPACE and the AD model always in close agreement? The answer is yes, as long as the absolute values of the interaction energies (ϵ_{ij}/kT) are less than unity. For interaction energies greater than unity the models can differ very considerably indeed, although differences begin to appear at interaction energies lower than one. Figure 9 shows the deviations from Monte-Carlo simulations for COSMOSPACE and the AD model as a function of ϵ_{12}/kT for a lattice mixture of monomers in which $\epsilon_{11} = \epsilon_{22} = 0$. The mole fraction is kept constant at 0.02 for these calculations. We see in Fig. 9 that up to $\epsilon_{12}/kT = 0.2$ both methods are within the statistical noise of the Monte-Carlo calculations. At higher interaction energies the deviation from Monte-Carlo of the AD model increases exponentially (a factor 2 for each 0.1 increase in interaction energy) while

COSMOSPACE stays within the noise up to $\varepsilon_{12}/kT = 0.6$, after which the deviations increase, probably due to agglomeration effects that make the appearance in the Monte-Carlo simulations. Phase separation occurs at $\varepsilon_{12}/kT = 0.7$.

If we calculate liquid-liquid equilibrium for a cubic lattice mixture of molecules we find that phase separation occurs starting at $\varepsilon_{12}/kT = 0.42$ (this explains the difficulties with MC starting from $\varepsilon_{12}/kT = 0.5$) at $x_1 = 0.5$ and we find a liquid-liquid phase boundary at $x_1 = 0.02$ with $\varepsilon_{12}/kT = 0.7$. This explains why we observe deviations from the Monte-Carlo method starting from that interaction energy. These results show that COSMOSPACE is more accurate than the AD model. Furthermore, as long as we are in the single-phase region, COSMOSPACE is in close agreement with the Monte-Carlo simulations for these systems.

COSMOSPACE cannot account for differences between 3D cubic lattices and 2D-hexagonal lattices both with $z = 6$ because the model does not include any lattice structure information other than the number of faces belonging to one unit (molecule), i.e. other than z . A pair of A-hexagons creates two neighbouring sites that have two A-neighbours, i.e. where a third A-hexagon can get two favourable interactions at once. The situation of correlated two-site interactions is not covered by COSMOSPACE and, hence, it fails earlier in such lattices than it does for a cubic lattice where only special triples of A-particles form a favourable double A-neighbour site.

5. Application to Real Systems

In this section we show how COSMOSPACE performs in the interpretation of experimental data. Results are shown for three systems:

1. cyclohexane - ethanol
2. chloroform –methanol
3. benzene – acetonitrile

5.1 Cyclohexane - Ethanol

The system ethanol(1) – cyclohexane(2) is an interesting binary to study because it is one of the few systems that shows strong positive deviations from Raoult's law while not exhibiting liquid-liquid phase separation. Binaries that show liquid-liquid phase separation make a comparison difficult because the finite activity coefficients in a particular concentration range are missing.

Activity coefficients for this system at 293.15 K are shown in Figure 10. These activity coefficients were calculated from the P - x - y data of Asperion et al. (1993) and Nagai and Ishii (1935) using the saturation pressure of ethanol (43.80 mmHg) and cyclohexane (79.80 mmHg) from Anderson and Prausnitz (1978). Also shown in Fig. 10 are infinite dilution activity coefficients, determined by interpolation of other data as described later in this section (see Figure 11). The (interpolated) activity coefficients at infinite dilution of ethanol and cyclohexane at 293.15 K are: $\gamma_1^\infty = 79.6 \pm 7.9$ and $\gamma_2^\infty = 9.9 \pm 1.0$.

A least squares fit of the experimental activity coefficients in the ethanol concentration range of $0.15 \leq x_1 \leq 0.85$ gave UNIQUAC parameters of $A_{12} = -82.39$ cal/mol and $A_{21} = 861.41$ cal/mol. Activity coefficients calculated from UNIQUAC are shown in Fig. 10; the deviation of UNIQUAC at concentrations outside this range can clearly be seen. Furthermore,

UNIQUAC predicts liquid-liquid phase separation between an ethanol mole fraction of 0.18 and 0.65 that in reality does not exist. For this reason chemical engineers usually would choose the Wilson model for this kind of situation since the Wilson model is incapable of predicting liquid-liquid phase equilibrium.

In order to describe the activity coefficients by COSMOSPACE the number and types of surface segments have to be defined. The simplest COSMOSPACE model is the homogeneous double binary model (in which each molecule has only one type of surface) in Table 1. Using the tables in Bondi (1964) and the definition of Abrams and Prausnitz (1975) (i.e. $q_i = A_i / 2.5$) we find:

$$n_1^A = 2q_1 = 2 \times 4.93 / 2.5 = 3.944, \quad n_1^B = 0, \quad n_2^A = 0, \quad n_2^B = 2q_2 = 2 \times 8.10 / 2.5 = 6.48$$

The combinatorial activity coefficients at infinite dilution are computed from the UNIQUAC model as required by COSMOSPACE and are:

$$\gamma_{C,1}^\infty = \gamma_{C,2}^\infty = 0.94$$

Thus, the experimental residual activity coefficients at 293.15 K are:

$$\gamma_{R,1}^\infty = 79.6 / 0.94 = 84.4 \quad \gamma_{R,2}^\infty = 9.9 / 0.94 = 10.5$$

The interaction parameter τ_{AB} in the homogeneous double-binary model can be calculated directly from the activity coefficients at infinite dilution (Table 1):

$$\tau_{AB} = \left(\frac{1}{\gamma_{R,1}^\infty} \right)^{\frac{1}{n_1^A}} = \left(\frac{1}{\gamma_{R,2}^\infty} \right)^{\frac{1}{n_2^B}} \quad (38)$$

These expressions lead to two very different estimates of the COSMOSPACE interaction parameter of

$$\tau_{AB} = 0.61 \quad \text{or} \quad \tau_{AB} = 48.7$$

The large difference between these values suggests that the homogeneous COSMOSPACE model will give a poor fit of the experimental data. The next step then is to try the non-homogeneous double-binary COSMOSPACE model summarized in Table 2, with ethanol being divided into two surface types A and B (–OH and an alkyl group), whereas cyclohexane consists of only the alkyl surface type.

The ratio between the A and B surfaces in ethanol could be obtained from the full COSMOTHERM model (as will be done later in Section 5.3), or by using the UNIFAC surface parameters. However, we have chosen to employ an adjustable parameter, f , defined as the fraction of the ethanol surface that is the alkyl surface type. The value for this parameter ranges from 0 to 1. In the non-homogeneous double-binary model the surface segments become:

$$n_1^A = 3.944(1 - f), \quad n_1^B = 3.944f, \quad n_2^A = 0, \quad n_2^B = 6.48$$

Regression of the same activity coefficient data as used above for fitting the UNIQUAC parameters gives $\tau_{AB} = 0.1102$ and $f = 0.4322$. The solid lines in Fig. 10 show the goodness of the fit with the COSMOSPACE model using these parameters, especially outside the finite concentration range used to fit the parameters. The activity coefficients at infinite dilution from COSMOSPACE are $\gamma_1^\infty = 41.91$ and $\gamma_2^\infty = 9.66$. The activity coefficient at infinite dilution of cyclohexane in ethanol is within the experimental error, but that of ethanol in cyclohexane is too small compared to the interpolated data point shown in Fig. 10. Despite this, the two-parameter non-homogeneous COSMOSPACE model provides a better fit to the experimental data than does the two-parameter UNIQUAC model. It is also important to note that, the non-homogeneous COSMOSPACE model does not predict any liquid-liquid phase separation, whereas UNIQUAC does. The upper critical solution temperature predicted using

COSMOSPACE (and the parameters above) is 264.95 K (well below the temperature of the data) - at an ethanol mole fraction of 0.31.

The value of τ_{AB} is equivalent to an interaction energy $U_{12} - \frac{1}{2}(U_{11} + U_{22}) = 1.285$ kcal/mol.

The high value reflects the relative high repulsion between the alkyl surface and the hydroxyl surface. When ethanol is at infinite dilution, the hydroxyl group always encounters alkyl surfaces, and on average the interaction energy is high. On the other hand, when cyclohexane is at infinite dilution, it interacts in part with the alkyl group in ethanol, thereby reducing the (average) interaction energy, as seen in its lower infinite dilution activity coefficient. The estimated fraction for the alkyl surface type, $f = 0.4322$, is roughly equivalent to the surface area of a CH_3 group. This could imply that the CH_2 in ethanol is different from CH_2 in cyclohexane.

We now turn to an examination of the temperature dependence of the activity coefficients for this system. Figure 11 depicts the infinite dilution activity coefficients of ethanol (γ_1^∞) and cyclohexane (γ_2^∞) as a function of the system temperature. Interpolation of the data in Fig. 4 and using the values of the combinatorial contribution to the activity coefficients already computed ($\gamma_{C,1}^\infty = \gamma_{C,2}^\infty = 0.94$) gives the ratio of logarithmic residual activity coefficients:

$$\ln(\gamma_{R,1}^\infty)/\ln(\gamma_{R,2}^\infty) = \ln(\gamma_1^\infty)\ln(\gamma_{C,2}^\infty)/\ln(\gamma_2^\infty)\ln(\gamma_{C,1}^\infty) = 1.88 \text{ at } 293.15 \text{ K}$$

$$\ln(\gamma_{R,1}^\infty)/\ln(\gamma_{R,2}^\infty) = 1.26 \text{ at } 350 \text{ K}$$

Now, the ratio of logarithmic residual infinite dilution activity coefficients in the homogeneous COSMOSPACE model is given by (see Table 1 again):

$$\frac{\ln(\gamma_{R,1}^\infty)}{\ln(\gamma_{R,2}^\infty)} = \frac{n_1^A \ln(1/\tau_{AB})}{n_2^B \ln(1/\tau_{AB})} = \frac{n_1^A}{n_2^B} = \frac{2q_1}{2q_2} = \frac{A_1}{A_2} \quad (39)$$

which is a constant. Using the data in Bondi (1964), the ratio of the Van der Waals surface areas of ethanol (A_1) and cyclohexane (A_2) is $3.944 / 8.1 = 0.49$. This lies outside the range of the experimental values estimated above. This is further evidence that the homogeneous double-binary COSMOSPACE model is not likely to provide a good fit to the experimental data. For the ratio of activity coefficients to be temperature dependent requires at least one of the molecules to have more than one type of surface. The simplest model capable of such a representation is the non-homogeneous double-binary COSMOSPACE model as discussed above.

In the UNIQUAC model the ratio of logarithmic infinite dilution activity coefficients is temperature dependent

$$\frac{\ln(\gamma_2^\infty)}{\ln(\gamma_1^\infty)} = \frac{q_1 \ln(\tau_{21}) - (1 - \tau_{12})}{q_2 \ln(\tau_{12}) - (1 - \tau_{21})} \quad (40)$$

where $\tau_{12} \neq \tau_{21}$. However, the asymmetric case has no physical meaning, as explained in subsection 2.3. Eq. 40 reduces to Eq. 39 in the limit of small interactions (i.e. $|\tau_{ij}-1| \ll 1$).

Interpolation of the data in Fig. 11 gives values for the activity coefficients at infinite dilution of ethanol and cyclohexane shown in Fig. 10.

A regression on all the data in Figure 10 and 11 with the non-homogeneous double-binary COSMOSPACE model yields at a temperature of 293.15 K $\tau_{AB} = 0.1093$, which is equivalent to an interaction energy $u_{12} - \frac{1}{2}(u_{11} + u_{22}) = 1.289$ kcal/mol, and $f = 0.4363$. These values are similar to those of previous regression. Again, no liquid-liquid phase separation was observed (as should be the case, of course). The limiting activity coefficients at 293.15 K also

improve slightly: $\gamma_1^\infty = 43.54$ and $\gamma_2^\infty = 10.00$. The temperature dependence of the limiting activity coefficient is quite well described by the COSMOSPACE model and above parameters as shown in Fig. 12. For comparison Fig 12 also shows the activity coefficients predicted by the group contribution model UNIFAC (Fredenslund et al., 1975, 1977). The results obtained with UNIQUAC are not as good as those obtained with UNIFAC in this case (e.g. the limiting activity coefficient for ethanol in cyclohexane and that of cyclohexane in ethanol were higher and lower respectively).

Figure 12 also shows that the COSMOSPACE model and the original UNIFAC model yield almost the same results. However, the description by the original UNIFAC model requires two interaction energies, while COSMOSPACE needs only one. Moreover, UNIFAC predicts liquid-liquid phase separation below 73°C where, in reality none exists. COSMOSPACE does not predict phase separation.

The Dortmund modified UNIFAC model (Weidlich et al. 1987, Gmehling et al. 1993) gives a better description of the limiting activity coefficients for the above data and predicts no liquid-liquid equilibrium (see Fig. 13). This is primarily a result of the 2 additional pairs of parameters, which make the interaction energy temperature dependent (i.e.

$u_{ij} - u_{ii} = a_{ij} + b_{ij}T + c_{ij}T^2$). Thus, the modified UNIFAC model contains 6 adjustable parameters for the interaction energy and one for the division of ethanol into two main groups. The same approach for COSMOSPACE would require only 4 parameters; 3 parameters for the interaction energy and one for the division of the ethanol surface into two types of surfaces.

From Figs. 10 and 11 we see that a linear relationship should suffice to fit the interaction energy in order to give as good a fit to the data as is given by the Dortmund modified

UNIFAC model. For the description in Fig. 13 the COSMOSPACE fraction of alkyl-surface is $f = 0.442$ and the COSMOSPACE interaction energy in kcal/mol is found to be

$$u_{12} - \frac{1}{2}(u_{11} + u_{22}) = 2.935 - 5.0132 \times 10^{-3} T$$

Now, the limiting activity coefficients at 293.15 K are $\gamma_1^\infty = 73.1$ and $\gamma_2^\infty = 10.25$. These values are within the experimental error.

Figure 13 makes clear that the non-homogeneous double-binary COSMOSPACE description has at least the same potential as the Dortmund modified UNIFAC model, but uses fewer parameters.

5.2 Benzene - Acetonitrile

Activity coefficients for the benzene – acetonitrile system at 20°C are shown in Fig. 14. The experimental values are computed from the data of Werner and Schuberth (1966). Here we try to describe the system as a homogeneous double binary system, yielding:

$$n_1^A = 2q_1 = 4.8, \quad n_1^B = 0, \quad n_2^A = 0, \quad n_2^B = 2q_2 = 3.448$$

The best fit we could achieve with this assumption yields a COSMOSPACE interaction parameter of $\tau_{AB} = 0.745$ (see the broken lines shown in Fig. 14). Apparently, this simple model is incapable of describing the asymmetry of the activity coefficients. This failure of the simplest COSMOSPACE model results from the fact, that in this model the ratio of the logarithmic infinite dilution activity coefficients is just the ratio of the surface areas of the components (see Eq. 27). In this case compound 2 is smaller than compound 1, but it has the higher activity coefficient. Thus, it is not possible to obtain a good fit to the data using the homogeneous double binary model (in which each species consists of its own surface segments only) and one single binary interaction parameter. It is, however, possible to obtain significantly improved fits if, in addition to τ_{AB} , we also consider the numbers of segments as

adjustable parameters. That is, we estimate the segment numbers from $n_i^v = 2s_i q_i$, where s_i are segment scaling parameters. Also shown in Fig. 14 (by the solid lines) are the activity coefficients from COSMOSPACE obtained in this way using the segment number parameters of $s_1=1$ for benzene (indicating no adjustment), and $s_2= 1.5$ for acetonitrile. It can be seen that the fit is significantly improved (although we recognize that an extension towards a multi-surface kind model would be the physically more consistent way of improvement). We have observed that this simple extension of the COSMOSPACE approach allows us to obtain good fits to the activity coefficients for many binary systems using the simple, homogeneous double binary model of Table 1.

5.3 Chloroform - Methanol

The chloroform - methanol mixture is an example of a binary system in which the activity coefficients of one of the species (chloroform) appears to pass through a maximum. Not all conventional activity coefficient models are able to represent data from such systems (see Walas, 1985, page 189). Indeed, neither is COSMOSPACE if we stay within the analytic double-binary approach that we have employed in our previous examples. Figure 15 shows the activity coefficients for this system. The points are from the data of Kireev and Sitnikov (1941) at 49.3°C. The broken lines were obtained with COSMOSPACE after optimising the segment scaling parameters and a single binary interaction parameter in a homogeneous double binary model ($n_1^A = 2q_1 = 4.68$, $n_1^B = 0$, $n_2^A = 0$, $n_2^B = 6.6q_2 = 9.42$, $\tau = 0.74$). The fit over most of the concentration range is quite good, but fails in the region where the activity coefficient of chloroform flattens ($x_1 < 0.2$ - approximately).

To improve the fit with COSMOSPACE we turned first to the *a-priori* predictive model COSMO-RS, which suggested that the chloroform-methanol system should, perhaps be modelled by five different kinds of segment. The results of this approach also are shown in

Fig. 6 and it can be seen that COSMOSPACE is indeed able to model the flattening of the activity coefficient of chloroform, with a small, but noticeable, improvement in its ability to represent the rest of the activity coefficient profile.

The surface areas and volumes needed for this calculation were obtained using COSMO-RS with the following results (in Ångstrom units):

Species	Area	Volume
Chloroform	117.52	105.73
Methanol	66.36	47.52

The surface kinds, their characteristics, and the numbers of each kind of surface per molecule, also from COSMO-RS, are provided in Table 3.

The segment-segment interaction parameters were computed from

$$\tau_{ij} = \exp(-a_{eff}(u_{ij} - \frac{1}{2}(u_{ii} + u_{jj}))/RT)$$

The segment interaction energies were estimated using parameters from COSMO-RS (Klamt and Eckert, 2000) and the mean values of σ given in Table 3. For example:

$$U_{1,4} = \frac{\alpha'}{2}(\sigma_{mean,1} + \sigma_{mean,4})^2 = \frac{6635}{2}(-0.015 + 0.005)^2 = 0.3$$

The complete set of interaction energies is provided in Table 4.

The empty spaces in Table 4 indicate that the 2-5 interaction energies, i.e. the interaction energies of the polar hydrogen of chloroform with the polar oxygen of methanol, were treated as adjustable parameters, because the 2-5 interaction is influenced most by the simplification that results from reducing the large number of surface types in COSMO-RS to just 5 in

COSMOSPACE. The value of $a_{eff} = 6.15$, to convert the number of COSMO-RS surface segments into the number of COSMOSPACE surface segments, was taken from COSMO-RS as well. Thus, there is, in fact, only 1 free parameter that has been fit to the experimental data instead of 10, as might have been anticipated from the fact that we have used 5 different segment kinds. The interaction parameters are given in Table 5.]

The flattening of the activity coefficient of chloroform at infinite dilution can mainly be attributed to the extremely high interaction of the segment pairs 1-5 and 2-5. We note that UNIFAC cannot describe this behaviour as it treats chloroform and methanol as compounds each with only one kind of surface. Hence, there is only one surface pair. But even in the approach of 5 segments UNIFAC will fail to describe the extrema correctly as it underestimates the surface segment activity coefficient between the strongly interacting surfaces 1-5 and 2-5 (See Fig. 14.).

5. Conclusions

An analytical solution to a statistical thermodynamic model of pair-wise interacting surface segments has been derived. The COSMOSPACE equations for homogeneous molecules are similar to equations of the quasi-chemical theory derived by Guggenheim (1952). Though similar the models have quite different starting points and they are used in a different way. The model of Guggenheim is that a central molecule is surrounded by a number of molecules (z) in a specific arrangement: e.g. when the liquid can be approximated by a lattice of cubes the coordination number is $z = 6$. In the approach developed here we divide the molecular surface into small pieces. The areas can be of different types, but this is not a requirement. The size of a surface piece is large enough to reflect differences in surface types on a heterogeneous molecule. In COSMO-RS the size of a surface piece is about 1 \AA^2 . In this way the COSMO-RS and COSMOSPACE models do not need a coordination number; the

interaction of surface areas is only in one dimension and is independent of most molecule structures (as long as the accessibility of the surface segments is not affected by the geometry of the molecule). Going from interacting surfaces to interacting molecules is the next step in our derivation of the equations for chemical potentials and activity coefficients, and this is the point of departure from quasi-chemical theory (although the final results appear similar for binary systems of homogeneous molecules). We note that our approach makes it easier to calculate activity coefficients in mixtures of compounds that would need different coordination numbers in a quasi-chemical model.

The widely used UNIQUAC equation for activity coefficients has been shown to be a first order approximation of the COSMOSPACE model if the two interaction parameters τ_{12} and τ_{21} are identical, and if some of the inconsistencies regarding the treatment of the coordination number in UNIQUAC are removed. If the same physical interaction energies per unit area are used, UNIQUAC and COSMOSPACE agree very well in the range of $x > 0.1$, but in the high dilution region $x < 0.1$ the COSMOSPACE equation shows a different and physically more consistent behavior.

COSMOSPACE is in extremely good agreement with Monte-Carlo simulations for cubic lattice fluids where the coordination number is 6. The differences between COSMOSPACE and UNIQUAC for these fluids can be quite considerable indeed. COSMOSPACE is also in good agreement with the lattice fluid model of Aranovich and Donohue as long as we are in the single fluid phase region. The phase boundaries predicted with COSMOSPACE are in good agreement with those estimated using Monte-Carlo methods. COSMOSPACE is unable to distinguish between 3D and 2D lattices with the same coordination number, and it performs worse for the 2D case. However, a 2D-hexagonal lattice with coordination number 6 is a less

realistic model of a real liquid system than is a 3D-cubic lattice with the same coordination number.

The experimental data of ethanol-cyclohexane can be described quite well with a one parameter COSMOSPACE model. UNIQUAC fails in the description of activity coefficient outside the finite concentration range, when the parameters are fitted to activity coefficients of finite concentrations. Moreover, UNIQUAC predicts a region of liquid-liquid phase separation that does not, in fact, exist. COSMOSPACE does not predict any phase separation. In order to describe the experimental data with a higher accuracy using the COSMOSPACE model, more detailed molecular information is required. The molecules could be divided into subgroups of a specific type of surface. This approach requires more interaction parameters. However, as shown here for the chloroform – methanol system, they can be estimated quite easily (using the *a priori* model COSMO-RS), and good fits to the data can be achieved using a very limited number of adjustable parameters. The flattening of the activity coefficient of chloroform at low concentrations of chloroform can be fit with COSMOSPACE using five kinds of surface segment, but only one adjustable interaction parameters. Another way to improve the results is by optimising the number of surface segments as shown in this paper for the benzene – acetonitrile system. This approach has also been used with UNIQUAC (Anderson and Prausnitz, 1978).

Due to its versatility and its inherent physical consistency the COSMOSPACE model could become a valuable tool for empirical activity coefficient fits in binary and multi-component systems and, hence, become an alternative to the much more heuristic and less consistent models that presently are used. It must be kept in mind that a homogeneous molecule description is a simplification of physical reality. In real systems there usually will be a number of different kinds of surface segment, each having different interaction patterns. This

wealth of interaction information presently can only be considered within the full COSMORS/COSMOtherm approach, which generates the interaction information based on quantum chemical calculations and finally combines them with the COSMOSPACE model.

Acknowledgements

The authors would like to thank Dr. Marc D. Donohue and Dr. Gregory Aranovich for their valuable comments during the preparation of this paper.

Nomenclature

a_{eff}	Size of surface segment
E_{tot}	Total interaction energy
G	Gibbs free energy
K	Boltzmann constant
n	Number of surface segments
N	Number of molecules
$p_{\mu\nu}(P)$	Total number of pairs of kind $\mu\nu$
r	Relative volume
R	Gas constant
q	Relative surface area
u	interaction energy
U	Internal energy
x	Mole fraction
z	Coordination number
Z	Partition sum of an ensemble

Greek letters

δ_{ij}	Kronecker delta (1 if $i = j$, 0 otherwise)
ε_{ij}	Interaction energy
γ	Activity coefficient
Φ	Volume fraction

Ψ^μ	Pseudo activity coefficients in Eq. 16
μ	Chemical potential or pseudo-chemical potential
Θ	Surface fraction
τ	Interaction parameter

Subscripts

i	Pertaining to species i
SG	Staverman – Guggenheim

Superscripts

C	Combinatorial part
R	Residual part
ν, μ	Segment identifiers
A, B	Segment identifiers

References

- Abrams, D.S. and 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**, 116-128 (1975).
- COSMOtherm software, COSMOlogic GmbH&CoKG, Leverkusen, Germany (1999)
- Anderson, T.F. and J.M. Prausnitz, "Application of the UNIQUAC equation to calculation of multicomponent phase equilibria. 1: Vapor-liquid equilibria; 2: Liquid-liquid equilibria," *Ind. Eng. Chem. Proc. Des. Dev.*, **17**, 552-567 (1978).
- Aranovich, G.L. and M.D. Donohue, "A new model for lattice systems," *J. Chem. Phys.*, **105**, 7059-7063 (1996).
- Asperion, N., H. Hasse, and G. Maurer, "Limiting activity coefficients in alcohol-containing organic solutions from headspace gas chromatography," *J. Chem. Eng. Data*, **43**, 74-80 (1998).
- Ben Naim, A., *Solvation Thermodynamics*, Plenum Press, New York, NY (1987).
- Bondi, A., "Van der Waals Volumes and Radii," *J. Chem. Phys.*, **48**, 441-451 (1964).
- Dortmund Databank Software Tools from DDBST Software and Separation Technology GmbH.
- Fredenslund, A., R.L. Jones, and J.M. Prausnitz, "Group-contribution estimation of activity coefficients in nonideal mixtures," *AIChE J.*, **21**, 1086-1099 (1975).
- Fredenslund, A., J. Gmehling, and P. Rasmussen, *Vapor-Liquid Equilibria using UNIFAC*; Elsevier: Amsterdam (1977).
- Gmehling, J., J. Li, and M. Schiller "A modified UNIFAC Model 2. Present parameter matrix and results for different thermodynamic properties," *Ind Eng. Chem. Res.*, **32**, 178-193 (1993).
- Gmehling, J., J. Mencke, and M. Schiller, *Activity coefficients at infinite dilution*, Chemistry Data Series Vol. 9, DECHEMA (1994).

- Gmehling, J. and U. Onken, Vapour-Liquid Equilibrium Data Collection, Organic Hydroxy compounds: Alcohols, DECHEMA Chemistry Data Series Vol. 1, part 2a, p443, DECHEMA (1977).
- Guggenheim, E.A., Mixtures, Oxford University Press, Oxford (1952).
- Kireev, V.A. and I.P. Sitnikov, Vol. I/2a, p18 of Gmehling and Onken (1977).
- Klamt, A., "Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena," *J. Phys. Chem.*, **99**, 2224-2235 (1995).
- Klamt, A., "COSMO and COSMO-RS," In Schleyer, P. v.R., and Allinger, L., (Eds.) Encyclopedia of Computational Chemistry, Vol. 2, Wiley, New York, NY, 604-615 (1998).
- Klamt, A., Eckert, F., "COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids," *Fluid Phase Equilibria*, **172**, 43-72 (2000).
- Klamt, A., F. Eckert, "Reconsideration of UNIQUAC r and q parameters based on COSMO-RS," manuscript in preparation (2002).
- Klamt, A., V. Jonas, T. Buerger, J.C.W. Lohrenz, "Refinement and Parametrization of COSMO-RS," *J. Phys. Chem.*, **102**, 5074-5085 (1998).
- Lipson, J.E.G., "Polymer blends and polymer solutions: A Born-Green-Yvon integral treatment," *Macromolecules*, **34**, 1334 (1991).
- Nagai, J., Ishii, N., *J. Soc. Chem. Ind. Jap.*, **3**, 86 (1935).
- Sandler, S.I., Chemical and Engineering Thermodynamics, 3rd Edition, Wiley, New York (1999).
- Walas, S.M. Phase Equilibria in Chemical Engineering, Butterworth, Boston (1985).
- Weidlich, U., J. Gmehling, "A modified UNIFAC Model 1. Prediction of VLE, h^E , γ^∞ ." *Ind. Eng. Chem. Res.*, **26**, 1372-1381 (1987)
- Werner, G. and H. Schuberth, Vol. I/7, p131-132 of Gmehling and Onken (1977).

Wu, D-W., Y. Cui, and M.D. Donohue, "Local composition model for lattice mixtures," *Ind. Eng. Chem. Res.*, **37**, 2936-2946 (1998).

Appendix A: Derivation of the COSMO-RS self-consistency equation

Consider an ensemble S of N objects, belonging to m classes (surface segment types in COSMOSPACE) of identical objects (surface segments in COSMOSPACE). Let N_i be the number of objects of class i , and $x_i = N_i / N$ is the relative concentration. Let there be N sites that can be occupied by the N objects. Each two of these sites form a close pair. Hence, occupying the N sites all objects are paired. The interaction energy of the pairs shall be described by a symmetric matrix, E_{ij} , where i and j denote two different classes of objects. Let Z be the partition sum of the ensemble. From statistical thermodynamics the chemical potential μ_i of objects of class i is given by:

$$\mu_i = -kT \frac{d \ln Z}{d N_i} = -kT \frac{\ln(Z) - \ln(Z_{(-i)})}{N - (N_i - 1)} = -kT \ln \left(\frac{Z}{Z_{(-i)}} \right) \quad (\text{A1})$$

where $Z_{(-i)}$ denotes the partition sum of the ensemble with N_i reduced by one. In the same way we find

$$\mu_i + \mu_j = -kT \ln \left(\frac{Z}{Z_{(-i)}} \frac{Z_{(-i)}}{Z_{(-i,-j)}} \right) = -kT \ln \left(\frac{Z}{Z_{(-i,-j)}} \right) \quad (\text{A2})$$

where $Z_{(-i,-j)}$ denotes the partition sum of the ensemble with N_i and N_j reduced by one. Let us now consider a second similar ensemble S' in which all objects may be distinguishable. Let us denote these objects by Greek indices α and β . Let the objects still fall into m classes. Let i be the class of object α and j the class of β . Let the multiplicities N_i of the classes be the same as in S . Let the pair interaction energies $E'_{\alpha\beta}$ depend only on the classes, i.e. $E'_{\alpha\beta} = E_{ij}$. Due to the permutations of the distinguishable objects in S' , each state of S corresponds to a number

$$N_{\text{deg}} = \prod_{i=1}^m N_i! \quad (\text{A3})$$

of states in S' which have the equivalent pairing of objects and hence have the same energy.

As a consequence, we have the relation $Z = Z' / N_{\text{deg}}$ for the partition sums of the two ensembles. Applying Eq. A1 to this relation we get

$$\begin{aligned}\mu_i &= -kT \ln \left(\frac{Z}{Z_{(-i)}} \right) = -kT \ln \left(\frac{Z' N_{\text{deg}(-i)}}{N_{\text{deg}} Z'_{(-\alpha)}} \right) \\ &= -kT \ln \left(\frac{Z'}{Z'_{(-\alpha)}} \right) - kT \ln \left(\frac{N_{\text{deg}(-i)}}{N_{\text{deg}}} \right) \\ &= \mu'_\alpha + kT \ln N_i\end{aligned}\tag{A4}$$

Here $N_{\text{deg}(-i)}$ is the degeneracy factor of the ensemble with one object of class i missing.

According to Eq. A3 this is by a factor N_i smaller than N_{deg} .

As a next step we derive a relation for the chemical potentials in ensemble S' . For that let us first choose any object α . We can sort all states of the ensemble with respect to the partner β of α . The partition sum of the system can be written as a sum over all partners β , taking into account the Boltzmann weight $\exp(-E'_{\alpha\beta})$ of the interaction of α with β multiplied by the partition sum $Z_{(-\alpha,-\beta)}$ of a system missing objects α and β system. A factor $N/2$ arises from the fact that a pair $\alpha\beta$ can be taken away from any of the $N/2$ pair positions of the ensemble, and a factor of 2 arises from the two possible orientations of that pair on each pair position. Hence we have

$$Z' = \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^N \frac{N}{2} 2 \exp\left(\frac{-E'_{\alpha\beta}}{kT}\right) Z'_{(-\alpha,-\beta)} = \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^N N \exp\left(\frac{-E'_{\alpha\beta} + \mu'_\alpha + \mu'_\beta}{kT}\right) Z'\tag{A5}$$

where we have made use of the equivalent of equation Eq. A2 for ensemble S' . Dividing both sides by Z' and $\exp(\mu'_\alpha / kT)$, and on taking the logarithm we have

$$\mu'_\alpha = -kT \ln \left\{ \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^N \exp\left(\frac{-E'_{\alpha\beta} + \mu'_\beta}{kT}\right) \right\} - kT \ln N\tag{A6}$$

The restriction $\beta \neq \alpha$, which is caused by the fact that the object α cannot be its own interaction partner, can be omitted in the thermodynamic limit of large N , because the resulting change is infinitesimally small. Doing this, replacing μ'_α and μ'_β by μ_i and μ_j according to Eq. A4, using the corresponding energies E_{ij} instead of $E'_{\alpha\beta}$, and replacing the summation with respect to β by a summation with respect to classes j of multiplicity N_j , we get:

$$\begin{aligned}\mu_i &= -kT \ln \left\{ \sum_{j=1}^m N_j \exp \left(\frac{-E_{ij} + \mu_j - kT \ln N_j}{kT} \right) \right\} + kT \ln \frac{N_i}{N} \\ &= -kT \ln \left\{ \sum_{j=1}^m \exp \left(\frac{-E_{ij} + \mu_j}{kT} \right) \right\} + kT \ln x_i\end{aligned}\tag{A7}$$

The activity coefficient γ_i of an object of type i in this ensemble is defined by

$$kT \ln \gamma_i = \mu_i - \mu_{ii} - kT \ln x_i\tag{A8}$$

where μ_{ii} denotes the chemical potential of class i in an ensemble with $x_i = 1$. In this case Eq.

A7 simply reduces to

$$\mu_{ii} = \frac{1}{2} E_{ii}\tag{A9}$$

An interaction parameter τ_{ij} , can be defined by

$$\tau_{ij} = \exp \left\{ - \frac{E_{ij} - \frac{1}{2}(E_{ii} + E_{jj})}{kT} \right\}\tag{A10}$$

Subtracting $kT \ln(x_i)$ and μ_{ii} on both sides of Eq. A7 yields, in combination with Eqs. A8-

A10:

$$\begin{aligned}
kT \ln \gamma_i &= \mu_i - \mu_{ii} - kT \ln x_i \\
&= -kT \ln \left\{ \sum_{j=1}^m x_j \exp \left(\frac{-E_{ij} + \frac{1}{2}(E_{ii} + E_{jj}) + \mu_j - \frac{1}{2}E_{jj} - kT \ln x_j - \frac{1}{2}E_{ii}}{kT} \right) \right\} - \frac{E_{ii}}{2} \\
&= -kT \ln \left\{ \exp \left(\frac{-E_{ii}}{2kT} \right) \sum_{j=1}^m x_j \tau_{ij} \exp \left(\frac{kT \ln \gamma_j}{kT} \right) \right\} - \frac{E_{ii}}{2} \\
&= -kT \ln \left\{ \sum_{j=1}^m x_j \tau_{ij} \gamma_j \right\}
\end{aligned} \tag{A11}$$

Thus, we have the general result

$$\frac{1}{\gamma_i} = \sum_{j=1}^m x_j \tau_{ij} \gamma_j \tag{A12}$$

Appendix B: Proof of Gibbs-Duhem consistency

Here, we would like to provide an independent demonstration that activity coefficients obeying Eq. 12 satisfy the Gibbs-Duhem equation.

Multiplying Eq. A12 by $x_i \gamma_i$ followed by summation over the index i yields

$$1 = \sum_i x_i = \sum_i x_i \gamma_i \sum_j x_j \gamma_j \tau_{ij} \tag{B1}$$

Multiplying this equation by N^2 and replacing $N x_i$ by N_i yields

$$N^2 = \sum_{i,j} N_j \gamma_j \tau_{ij} N_i \gamma_i \tag{B2}$$

If we differentiate this equation with respect to any object number N_k we obtain

$$\frac{\partial}{\partial N_k} N^2 = 2N = \frac{\partial}{\partial N_k} \sum_{i,j} N_j \gamma_j \tau_{ij} N_i \gamma_i = 2 \sum_j N_j \gamma_j \frac{\partial}{\partial N_k} \sum_i N_i \gamma_i \tau_{ij} \tag{B3}$$

Division by $2N$ leads to

$$\begin{aligned}
1 &= \sum_j x_j \gamma_j \frac{\partial}{\partial N_k} \sum_i N_i \gamma_i \tau_{ij} = \sum_j x_j \gamma_j \left(\gamma_k \tau_{kj} + \sum_i N_i \tau_{ij} \frac{\partial \gamma_i}{\partial N_k} \right) \\
&= \gamma_k \sum_j x_j \gamma_j \tau_{kj} + \sum_i N_i \frac{\partial \gamma_i}{\partial N_k} \sum_j x_j \gamma_j \tau_{ij} = \frac{\gamma_k}{\gamma_k} + \sum_i \frac{N_i}{\gamma_i} \frac{\partial \gamma_i}{\partial N_k} = 1 + \sum_i x_i \frac{\partial \ln \gamma_i}{\partial x_k}
\end{aligned} \tag{B4}$$

where twice we have made use of Eq. A12. Subtracting 1 from both sides yields

$$0 = \sum_i x_i \frac{\partial \ln \gamma_i}{\partial x_k} \tag{B5}$$

which is the Gibbs-Duhem relation for activity coefficients. Hence activity coefficients that satisfy Eq. 12 must be thermodynamically consistent.

Appendix C: Algebraic Solution for the Binary Case

If we have only two kinds of segments the COSMOSPACE equations take the form

$$1 = \Theta^1 \gamma^{1^2} + \Theta^2 \gamma^2 \gamma^1 \tau \tag{C1a}$$

$$1 = \Theta^2 \gamma^{2^2} + \Theta^1 \gamma^1 \gamma^2 \tau \tag{C1b}$$

with $\tau = \tau_{12} = \tau_{21}$. Introducing the abbreviations $\eta_i = \Theta^i \gamma^i \tau^{0.5}$ we obtain

$$\tau \Theta^1 = \eta^{1^2} + \tau \eta^1 \eta^2 \tag{C2a}$$

$$\tau \Theta^2 = \eta^{2^2} + \tau \eta^1 \eta^2 \tag{C2b}$$

Upon substitution of $A = \eta^1 \eta^2$ and $B = \eta^1 / \eta^2$ the above expressions take the form

$$\tau \Theta^1 = A B + \tau A \tag{C3a}$$

$$\tau \Theta^2 = \frac{A}{B} + \tau A \tag{C3b}$$

By simple algebra using $\Theta^1 + \Theta^2 = 1$ we obtain a quadratic equation:

$$0 = A^2 \omega + A - \Theta^1 \Theta^2 \tag{C4}$$

where $\omega = \tau^{-2} - 1$. The solution is

$$A_{\pm} = \frac{-1 \pm \sqrt{1 + 4 \omega \Theta^1 \Theta^2}}{2 \omega} \tag{C5}$$

From limiting cases it can be shown that A- is the physically meaningful solution to this quadratic equation. The final expression for the activity coefficients of the different segments is

$$\gamma^i = \sqrt{\frac{1}{\Theta^i} + \frac{1 - \sqrt{1 + 4\Theta^1\Theta^2\omega}}{2\omega\Theta^i}} \quad (\text{C6})$$

Appendix D: Derivatives of the COSMOSPACE activity coefficients

For the general multi-component case, the set of equations represented by Eq. A12 can only be solved by an iterative method. While this calculation does not take much time, it could become quite expensive to calculate derivatives of the activity coefficients finite difference approximations. Since some process simulation packages make extensive use of these derivatives, it is worthwhile to present an efficient analytical formulation for them.

Let γ_i be a set of activity coefficients obeying Eq. A12 and let $\partial\gamma_i$ be the partial differential of γ_i . By multiplication of A12 with γ_i followed by differentiation we have

$$0 = \partial \left(\gamma_i \sum_j x_j \gamma_j \tau_{ij} \right) = \partial \gamma_i \sum_j x_j \gamma_j \tau_{ij} + \gamma_i \sum_j \partial x_j \gamma_j \tau_{ij} + \gamma_i \sum_j x_j \gamma_j \partial \tau_{ij} + \gamma_i \sum_j x_j \tau_{ij} \partial \gamma_j \quad (\text{D1})$$

Using A12 for the first term and making use of $\partial \ln \gamma = \partial \gamma / \gamma$ we get

$$0 = \partial \ln \gamma_i + \gamma_i \sum_j \partial x_j \gamma_j \tau_{ij} + \gamma_i \sum_j x_j \gamma_j \partial \tau_{ij} + \gamma_i \sum_j x_j \tau_{ij} \gamma_j \partial \ln \gamma_j \quad (\text{D2})$$

Multiplying D2 by $x_i^{0.5}$ we yield

$$0 = d_i + b_i + \sum_j C_{ij} d_j \quad (\text{D3})$$

where

$$d_i = \sqrt{x_i} \partial \ln \gamma_i \quad (\text{D4})$$

$$b_i = \sqrt{x_i} \gamma_i \sum_j \partial x_j \gamma_j \tau_{ij} + \sqrt{x_i} \gamma_i \sum_j x_j \gamma_j \partial \tau_{ij} \quad (\text{D5})$$

and

$$C_{ji} = \sqrt{x_i} \gamma_i \tau_{ij} \gamma_j \sqrt{x_j} \quad (\text{D6})$$

Note, that all derivatives in the b_i are directly available. If we now define the symmetric matrix C_{ij}^* by C_{ij} for $i \neq j$ and by $C_{ij} + 1$ for $i = j$, then D3 reduces to the simple linear equation system

$$0 = b_i + \sum_j C_{ij}^* d_j \quad (\text{D7})$$

Thus, after we have calculated all derivatives for segments with $x_i > 0$, we can calculate the remaining derivatives directly from equation D2.