HAL
open science

An improved group contribution method for PC-SAFT applied to branched alkanes: Data analysis and parameterization

M. Jaber, W. Babe, E. Sauer, J. Gross, R. Lugo, J.C. C de Hemptinne

To cite this version:

M. Jaber, W. Babe, E. Sauer, J. Gross, R. Lugo, et al.. An improved group contribution method for PC-SAFT applied to branched alkanes: Data analysis and parameterization. Fluid Phase Equilibria, 2018, 473, pp.183 - 191. 10.1016/j.fluid.2018.06.014 . hal-01905718

HAL Id: hal-01905718

https://ifp.hal.science/hal-01905718v1

Submitted on 26 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An improved group contribution method for PC-SAFT applied to branched alkanes: Data Analysis and parameterization

M. Jaber[1], W.Babe[1], E. Sauer[2], J. Gross[2], R. Lugo[1], JC de Hemptinne[1],

[1] IFP Energies Nouvelles, 1-4 avenue de Bois-Préau 92852 Rueil-Malmaison, France
[2] Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Abstract

In this work, the heterosegmented GC-PC-SAFT model is parameterized for branched alkanes. Branched alkanes can be regarded as the skeleton of more complex molecules. Analyzing experimental data of branched alkanes in comparison to the linear analogous gives rationale to decisions about what substances require a more refined or individualized parameterization.
A neighborhood-correction of the segment-number parameter is then proposed for branched alkanes as function of the distance between the functional groups. This correction leads to better agreement with experimental data in the prediction of vapor pressure (27 % AAD compared to 39 %) and allows distinguishing isomers.

1. Introduction

It is desirable in process engineering to replace measurements of some properties by predictive methods. This is specially the case for the first steps of process design, when the feasibility of the process has to be evaluated with the data available in literature and without willingness to conduct additional experiments. Increasingly complex molecules, i.e. substances with multiple functional groups and branched structures are targeted in process design, which necessitates predictive methods with strong physical basis. Before analyzing complex multifunctional molecules and mixtures, it is essential that pure and less complex compounds such as branched alkanes are adequately modeled. To this end, many approaches exist [1], but the most well-known ones are based on group contributions.

In group-contribution (GC) methods, the molecular structure is decomposed into building blocks referring to functional groups and any property of a considered molecule is estimated as a function of the contributions of these building blocks. The assumption is that the electron structure of a functional group is not (much) altered by neighboring groups. As a consequence, the parameter value of any group is considered to have the same contribution to the property of a considered species, irrespective of the other groups present in that molecule. GC methods are thus predictive by definition.

GC-methods can be used to estimate pure component properties. For example, such methods have been developed to determine critical properties of pure components in a predictive

manner [2, 3, 4]. The estimated values of critical temperature and critical pressure can subsequently be used as parameters in cubic equations of state.

In process engineering, it is often required to develop a thermodynamic model for the prediction of fluid properties and phase equilibria (such as fugacities) for mixtures that might involve multifunctional branched molecules. The predictive power of cubic equations of state (EoS) remains limited for molecules involving polar and/or associative interactions leading to pronounced thermodynamic non-ideality. Equations of state (EoS) derived from statistical mechanical approaches have meanwhile grown to powerful engineering models [5].

An important representative is the Statistical Associating Fluid Theory (SAFT) that is based on Wertheim's perturbation theory [6, 7, 8, 9, 10]. Here, the fluid is first assumed to consist of hard spheres. Attractive forces between the spheres are taken into account through a dispersive potential that is added (e.g. square-well or Lennard-Jones potential). Then, chain sites on each sphere enable the formation of chains. Finally, the chain can then associate through short-ranged attractive interactions as a model for hydrogen bonds. All of these steps contribute to the residual Helmholtz energy.

Many versions of SAFT equations have been developed such as the original SAFT [11], Chen−Kreglewski SAFT (CK-SAFT) [12], simplified SAFT [13], Lennard-Jones SAFT (LJ-SAFT) [14, 15] , soft-SAFT [16], variable-range SAFT (SAFT-VR) [17], perturbed-chain SAFT (PC-SAFT) [18], SAFTD-LJ-Branch [19] . Some authors propose additional terms. In the context of molecular models, the polar term is probably the most common one, as in Polar PC-SAFT [20, 21, 22].

This study applies the PC-SAFT framework [18]. An important feature of SAFT models is the underlying molecular model. The molecular model, although rather coarse-grained, is sufficiently detailed to distinguish various interaction types, which is a prerequisite for adequately correlating or predicting mixtures with substances that have asymmetric interaction-types.

In this work, an improved heterosegmented GC method for PC-SAFT applied to branched alkanes is proposed. The proposed model captures the effect of neighboring functional groups (i.e. GC) on a considered functional group by a neighborhood-correction term, which leads to a substantially improved description of some branched structures.

This study is organized as follows. A bibliographical review of GC approaches and their applicability for branched alkanes. Next, experimental data from literature that is used to adjust group-contribution parameter is analyzed. Finally, a group-contribution concept for PC-SAFT is proposed where neighborhood-corrections are incorporated.

2. Bibliographical review of GC-SAFT methods

A determining aspect for using an equation on a complex mixture is the parameterization. The usual and most accurate parameterization method is regression on experimental data. Yet, when no data are available, a predictive method is required.

The pure component parameters defining the molecular model of SAFT equations are the energy parameter ε, defining the van der Waals attraction, the segment size parameter σ and m as the number of spherical segments in chain-like molecules. More parameters can be added depending on the polarity and the association strength of a molecule. Those parameters are used to calculate the different contributions to the molar residual Helmholtz energy so that

$$a^{res} = a^{hs} + a^{disp} + a^{chain} \qquad (1)$$

Where a^{hs} is the Helmholtz energy of the hard-sphere reference term and a^{disp} that of the dispersion term, a^{chain} is the contribution from chain formation. The association and polar contribution are not discussed here because they are zero for alkanes.
Depending on the way the groups are introduced, the GC SAFT methods can be classified in homosegmented and heterosegmented GC approaches.

2.1 Homosegmented GC-SAFT approaches

In homosegmented GC methods, molecular chains are considered to be composed of identical segments (the segments representing the various chemical functional groups) and mixing rules are used for determining molecular parameters from functional group parameters.

Vijande et al. [23] were the first to implement a group-contribution approach within the SAFT framework. They developed a homosegmented GC approach for hydrofluoroethers and n-alkanes using the PC-SAFT EoS by considering a linear dependence of pure component parameters of a homologous series on the molecular mass. Vapor pressures and liquid densities were calculated and compared to experimental data showing deviations lower than 1% for both calculated properties on hydrofluoroethers. The mixing rules are

$$\boldsymbol{m_i} = \sum_{\alpha} \boldsymbol{n_{\alpha_i}}\ \boldsymbol{m_\alpha} \qquad (2)$$

$$\boldsymbol{m_i \varepsilon_i} = \sum_{\alpha} \boldsymbol{n_{\alpha_i}}\ \boldsymbol{m_\alpha \varepsilon_\alpha} \qquad (3)$$

$$\boldsymbol{m_i \sigma_i^3} = \sum_{\alpha} \boldsymbol{n_{\alpha_i}}\ \boldsymbol{m_\alpha \sigma^3_\alpha} \qquad (4)$$

where $n_{\alpha i}$ denotes the number of groups of type α within molecule i.

In subsequent work [24, 25], mutual perturbation parameters were added to consider proximity effects between functional groups.

The group of Passarello and de Hemptinne [26, 27] developed a GC approach by using the Lorentz Berthelot combining rules

$$\boldsymbol{m_i} = \sum_\alpha \boldsymbol{n_{\alpha_i}}\ \boldsymbol{R_\alpha} \qquad (5)$$

$$\boldsymbol{\varepsilon_i} = \sqrt[\left(\sum_\alpha n_{\alpha_i}\right)]{\prod_\alpha \boldsymbol{\varepsilon_\alpha}^{\boldsymbol{n_{\alpha_i}}}} \qquad (6)$$

$$\boldsymbol{\sigma_i} = \frac{\sum_\alpha \boldsymbol{n_{\alpha_i}}\,\boldsymbol{\sigma_\alpha}}{\sum_\alpha \boldsymbol{n_{\alpha_i}}} \qquad (7)$$

The approach has been applied to the SAFT-VR [17, 28] and to the PC-SAFT [18, 29] models. It appeared that the accuracy of the method was good, with deviations generally within 3% on vapor pressure and 2% on saturated liquid volumes of various hydrocarbon series [26]. Every type of functional group α, with n_{α_i} groups per molecule *i* contributes with a chain fraction $n_{\alpha_i}\ R_\alpha$ to the overall chain parameter m_i. In order to account for the branching of molecules, the contribution of the group to the segment-number parameter m_i is adjusted depending on the position of the group with respect to the end of the chain, as shown in Figure 1. The parameters of each group have been obtained by simultaneously regressing on vapor pressure and liquid density data of linear alkanes first (for CH_2 and CH_3) and branched alkanes second (for CH).

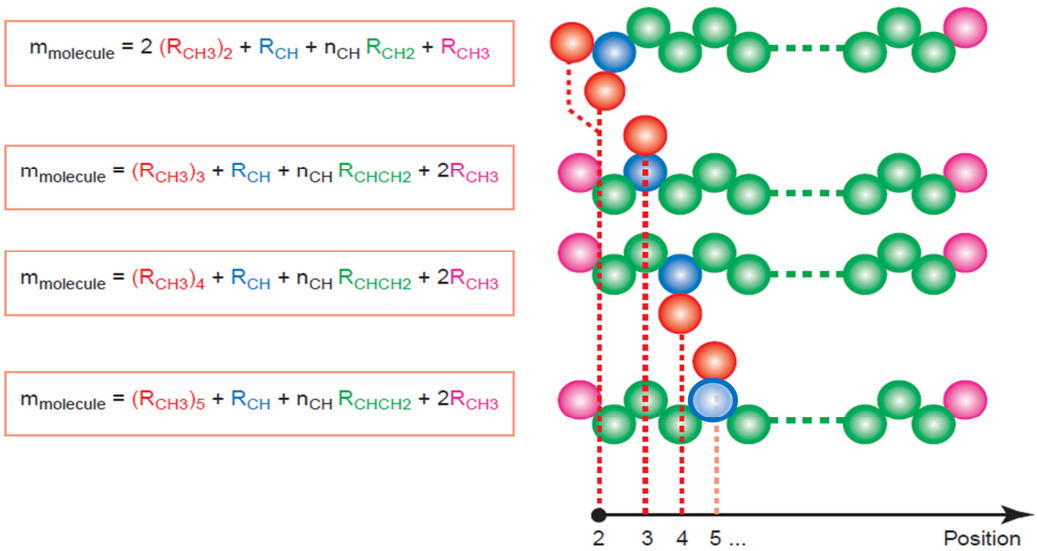

Figure 1: Distinction between isomers possible by taking into account the position of methyl groups through a different contribution R that is function of the position. Different parameters are used for end of chain CH_3 in magenta as compared to middle of chain CH_3 in red.

NguyenHuynh [30] proposed to improve the GC method by Tamouza et al. [26] by adding an extra parameter (hereby changing the name to mg-SAFT). The improvement is the "exclusion" of the over-accounting of dispersion energy between intra-molecular segments through an empirical correlation parameter. Three families (n-alkanes, n-alkyl-cycloalkanes and n-alkyl-benzenes were investigated). It allowed to improve significantly the prediction capability of the model for heavy compounds, particularly the prediction of vapor pressure, liquid density, and enthalpy of vaporization. The overall average absolute deviation obtained on the regression database is found to be 1.76% for vapor pressure and 0.85% for saturated liquid density for n-alkanes.

Tihic et al. [31] presented yet another GC approach applied to the PC-SAFT [32] EoS. Their originality was to consider a decomposition following the Constantinou-Gani GC method [4] thus introducing the distinction between first order groups (FOG: CH3, CH2, CH…) and second order groups (SOG: larger groups) so that

$$\boldsymbol{m_i} = \sum_{\alpha}(\boldsymbol{n_{\alpha_i} m_\alpha})_{FOG} + \sum_{\beta}(\boldsymbol{n_{\beta_i} R_\beta})_{SOG} \quad (8)$$

$$\boldsymbol{m_i \sigma_i^3} = \sum_{\alpha}(\boldsymbol{n_{\alpha_i} m_\alpha \sigma^3}_{\alpha})_{FOG} + \sum_{\beta}(\boldsymbol{n_{\beta_i} m_\beta \sigma^3}_{\beta})_{SOG} \quad (9)$$

$$\boldsymbol{m_i \varepsilon_i / k} = \sum_{\alpha}(\boldsymbol{n_{i\alpha} m_\alpha \varepsilon_\alpha / k})_{FOG} + \sum_{\beta}(\boldsymbol{n_{i\beta} m_\beta \varepsilon_\beta / k})_{SOG} \quad (10)$$

The concept of second order groups accounts for proximity effects and structural isomers. All group parameters (first and second order) were adjusted to a data-base of pure component {ε_i, σ_i, m_i} values, rather than directly to experimental data or correlated experimental data. For n-alkanes from n-pentane to n-eicosane the model showed an average error of 2.4% in vapor pressure and of 0.9% in liquid density towards experimental data.

Also, another study [33] presented a GC method for calculating PC-SAFT parameters by extending the group contribution model of Tihic and co-workers [31], for alkanes, aromatics, and cycloalkanes. The parameter values are a function of contributions from the various functional groups present and the nature of the various carbon atoms (aliphatic, aromatic, and naphthenic) comprising the molecule. Two different sets of GC parameters were proposed: one set for prediction of PC-SAFT parameters at high pressures, and another set for obtaining an accurate parameter set at low pressures.

Using the homo GC approach for mixtures may not be enough, since binary interaction parameters are generally necessary. Nguyen-Huynh et al. have proposed several papers in order to include such additional parameter in the GC framework [34, 35, 36, 37, 38].

2.2 Heterosegmented GC-SAFT approaches

SAFT models consider molecules as chains of tangent spherical segments. In the homosegmented approaches, all spherical segments are identical, with average parameters as shown in the previous section. It is also possible to consider a molecule consisting of (spherical) segments of different type, for example different segment size parameters, or each with an specific van der Waals (dispersive) energy parameter. In these heterosegmented approaches, the Helmholtz energy is explicitly expressed in terms of the involved segment types. The equations use the group parameters directly. This implies that in principle, more structural isomers can be distinguished.

Banaszak et al [39] and Shukla and Chapman [40] developed the basic procedure for treating heterosegmented molecules in the context of copolymers. Later, a heterosegmented PC-SAFT model was presented by Gross el al. [41] and Domanska and Paduszynski [42] using different chain contributions to the Helmholtz energy. Also, Peters et al. [43] developed a heterosegmented GC approach for polymer systems based on the PC-SAFT EoS. Table 1 summarizes the Helmholtz energy expressions according to the homosegmented and the heterosegmented GC-PC-SAFT model. One can see how for the heterosegmented approach,

group parameters are implemented directly into the equations whereas mixing rules are required for the homosegmented approach.

McCabe and co-workers [44] combined a GC approach to a hetero- SAFT-VR [45, 46] model leading to a GC-SAFT-VR. Because of the consideration of a specified chain connectivity, the GC-SAFT-VR EoS can distinguish between various structural isomers. Several chemical families such as alkanes, esters, alkylbenzenes, and associating chemical families have been studied. Furthermore, the GC-SAFT-VR EoS has been applied to describe phase equilibria of small molecules in polymer systems. For methyl-alkanes we note that the agreement with experimental data is good, with an overall deviation from experimental data for the liquid densities of 3.96% and 6.46% for the vapor pressures. As a test of the theory, several branched alkanes (dimethylalkanes) not included in the process of parameter identification were studied but show deviations higher than 10% in vapor pressures for all compounds and an overall deviation of 5% in liquid density.

Lymperiadis et al. [47, 48] developed a GC EoS based on SAFT–VR, called the SAFT–γ EoS, assuming that the molecule is constructed by fused heteronuclear united-atom groups. Within this formalism, an extra parameter (shape parameter S_k) is introduced per group for non-associating components, which essentially characterizes the portion of the group that contributes to the overall molecular properties. The parameter is equivalent to parameter $m_{\alpha i}$ of ref [33]. The authors present optimized group parameters for CH_3, CH_2, CH_3CH, with average deviations of 3.6% for vapor pressure and 0.9% for liquid densities for various families of organic compounds (n-alkanes and other hydrocarbons as well as alcohols).

Sauer et al. [49] compared the homosegmented approach and the heterosegmented approach for the PC-SAFT model using the same concept of parameter regression and the same training data. Their work showed that the heterosegmented GC approach leads to significantly better agreement with experimental data for the calculation of liquid density and vapor pressure for various chemical families including alkanes. The percentage average absolute deviations (%AAD) in liquid density and vapor pressure are respectively 1.140% and 4.639% for n-alkanes, 3.291% and 2.773 % for methylalkanes, 1.236% and 4.270% for ethylalkanes, 7.583% and 7.220 % for dimethylalkanes, 5.128% and 6.804% for trimethylalkanes.

However, a closer look at the results suggests that the results of Sauer et al. may still need further improvement to provide average deviations below 5% for dimethylalkanes and trimethylalkanes. In this work, we thus evaluate the performance of the heterosegmented GC-PC-SAFT as proposed by Sauer et al. [41, 49] for branched alkanes.

Table 1: Main differences between the homosegmented and heterosegmented GC-PPC-SAFT for non-polar and non-associative molecules

	Homosegmented GC-PC-SAFT	Heterosegmented GC-PC-SAFT
SAFT equation	$a^{res} = a^{hs} + a^{disp} + a^{chain}$	
parameters	$m_i = \sum_\alpha n_{\alpha_i}\, m_\alpha$ $\boldsymbol{\varepsilon_i} = \sqrt[(\sum_\alpha n_{\alpha_i})]{\prod_{\boldsymbol{\alpha}} \boldsymbol{\varepsilon_\alpha}^{n_{\alpha_i}}}$ $\sigma_i = \frac{\sum_\alpha n_{\alpha_i}\, \sigma_\alpha}{\sum_\alpha n_{\alpha_i}}$	m_α ε_α σ_α
Hard chain contribution	$\frac{A^{chain}}{RT} = \sum_{i=1}^{n} n_i (1 - m_i) \ln(g_{ii}^{hs})$ $g_{ij}^{hs} = \frac{1}{1-\zeta_3} + 3\frac{d_{ij}}{(1-\zeta_3)^2} + 2\frac{(d_{ij}\zeta_2)^2}{(1-\zeta_3)^3}$	$\frac{A^{chain}}{RT} = -\sum_i^n n_i \sum_\alpha \sum_\beta b_{\alpha\beta} \ln g_{\alpha\beta}^{hs}(d_{\alpha\beta})$ $g_{\alpha\beta}^{hs}(d_{\alpha\beta}) = \frac{1}{1-\zeta_3} + 3\frac{d_{\alpha\beta}}{(1-\zeta_3)^2} + 2\frac{(d_{\alpha\beta}\zeta_2)^2}{(1-\zeta_3)^3}$
Hard sphere contribution	$\zeta_l = \frac{\pi N}{6V} \sum_i x_i m_i d_i^l$ $d_i = \sigma_i \left[1 - 0.12 \exp\left(-3\frac{\varepsilon_i}{kT}\right)\right]$	$\zeta_l = \frac{\pi N}{6V} * \sum_i x_i \sum_\alpha n_{i\alpha}\, m_\alpha\, d^l{}_\alpha$ $d_\alpha = \sigma_\alpha[1 - 0.12\exp(-3\frac{\varepsilon_\alpha}{kT})]$
Dispersive contribution	$\overline{m^2\varepsilon\sigma^3} = \sum_i \sum_j x_i x_j m_i\, m_j \frac{\varepsilon_{ij}}{kT}\, \sigma_{ij}^{\mathbf{3}}$ $\overline{m^2\varepsilon^2\sigma^3} = \sum_i \sum_j x_i x_j m_i\, m_j \frac{\varepsilon_{ij}{}^2}{kT}\, \sigma_{ij}{}^3$	$\overline{m^2\varepsilon\sigma^3} = \sum_i \sum_j x_i x_j \sum_\alpha \sum_\beta n_{i\alpha}\, m_\alpha\, n_{j\beta}\, m_\beta (\frac{\varepsilon_{\alpha\beta}}{kT}) \sigma_{\alpha\beta}{}^3$ $\overline{m^2\varepsilon^2\sigma^3} = \sum_i \sum_j x_i x_j \sum_\alpha \sum_\beta n_{i\alpha}\, m_\alpha\, n_{j\beta}\, m_\beta \frac{\varepsilon_{\alpha\beta}{}^2}{kT} \sigma_{\alpha\beta}{}^3$

In those equations, ζ_l is the reduced density, d is the temperature dependent segment diameter, g^{hs} is the radial distribution function, $\overline{m^2\varepsilon\sigma^3}$ and $\overline{m^2\varepsilon^2\sigma^3}$ are abbreviations used in the calculation of the dispersive term, the subscripts α and β correspond to groups, the subscripts i and j correspond to molecules. $n_{i\alpha}$ is the number of groups of type α in the molecule i.

3. Analyzing experimental data for rationalizing GC-parameterization strategies

We here stress the importance of critically evaluating the available experimental data. The compounds investigated in this work are listed in table 1 of the supporting information.

The DIPPR [50] database generally provides uncertainties for their correlations, which are obtained from their methodology based on an internal consistency assessment (i.e. evaluating the thermodynamic consistency between different type of pure component properties, as for example the Clapeyron rule) [51]. We propose using data analysis, consisting in evaluating patterns in fluid phase behavior, as also proposed by Mathias [52]. We have used the approach by Rozmus [53] to further evaluate the quality of the vapor pressure correlations (i.e. consistency within a family of molecules). This analysis is very similar to that proposed by Mathias et al [52].

Rozmus et al [53] proposed to assess experimental data using a method based on the Clausius-Clapeyron equation:

$$\Delta h^{vap} = -R \left. \frac{\partial \ln P^{\sigma}}{\partial \left(1/T\right)} \right|_{T\,=\,T_b} \qquad (11)$$

where R is the gas constant, P^{σ} the vapor pressure, Δh^{vap} the enthalpy of vaporization at the normal boiling temperature and T the temperature.

With the Clausius-Clapeyron assumptions (constant enthalpy of vaporization), the logarithm of vapor pressure is linear with respect to the inverse of temperature so that the vapor pressure is approximately known if one point and the slope are known. For many substances, experimental data exist at the normal boiling point and we thus use this point for data analysis.

It is possible to represent each component with a single point on a diagram showing Δh^{vap} versus T_b and observe the trends within a given family (thus possibly identifying outliers). This is done in Figure 2, using quasi-experimental data taken from the DIPPR database for various branched alkanes. It appears that, as already observed by Mathias et al [54], all substances lie approximately on a line that goes from low enthalpies of vaporization and low boiling temperatures for low molecular weight compounds, to high values for substances of high molecular weight.

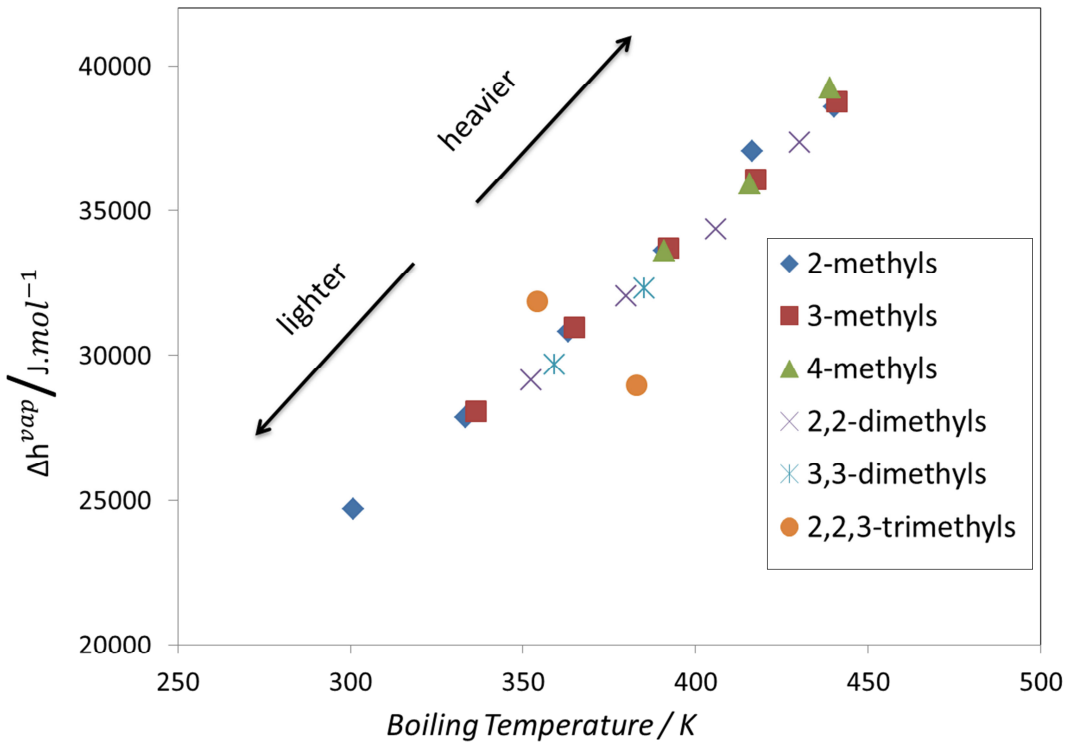

Figure 2: enthalpy of vaporization (Δh^{vap}) vs boiling temperature (T_b) from experimental data of alkanes.

Yet, the graph, as it is, does not clearly reveal the effect of the molecular topology. For better isolating the effect of branching, we find it advantageous to take the linear n-alkane with the same number of carbons as a reference and graphically represent the difference with its branched isomers for both enthalpy of vaporization and normal boiling temperature, according to

$$\Delta \mathbf{h}_{\mathbf{vap}_{\mathbf{isomer}}(\boldsymbol{shift})} = \Delta \mathbf{h}_{\mathbf{vap}_{\mathbf{isomer}}} - \Delta \mathbf{h}_{\mathbf{vap}_{\mathbf{n-alkane}}} \quad (12)$$

$$\mathbf{T}_{\mathbf{b}_{\mathbf{isomer}}(\boldsymbol{shift})} = \mathbf{T}_{\mathbf{b}_{\mathbf{isomer}}} - \mathbf{T}_{\mathbf{b}_{\mathbf{n-alkane}}} \quad (13)$$

As an example, for 2-methyloctane , we get

$$\Delta h_{vap_{2-methyloctane}(shift)} = \Delta h_{vap_{2-methyloctane}} - \Delta h_{vap_{nonane}}$$

This correction removes the effect of the molecular mass, which otherwise shifts heavier substances to the upper right corner of the diagram (figure 2). The location of a compound in the proposed diagram then better shows the effect of molecular topology.

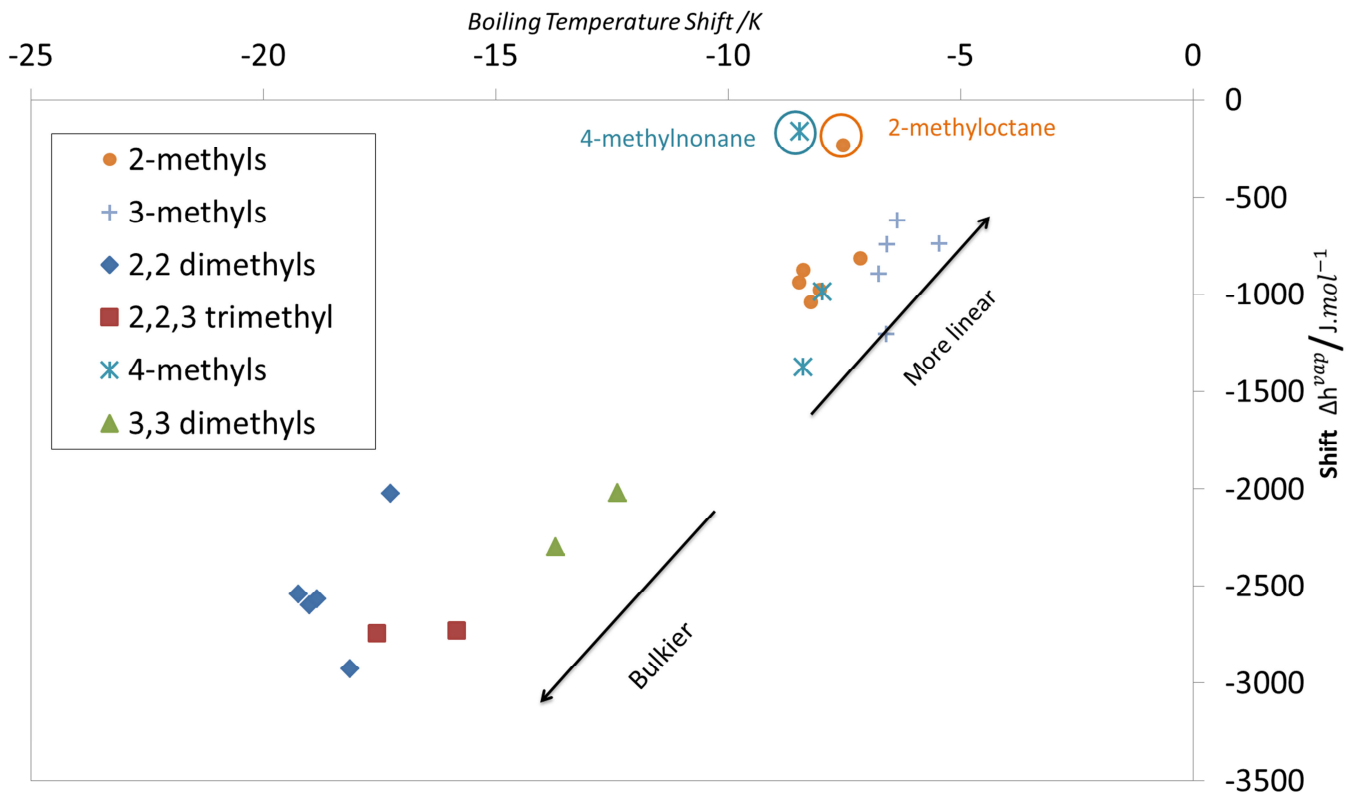

***Figure 3*: Shift in the enthalpy of vaporization (Δh^{vap}) vs shift in the boiling temperature (T_b) according to equations 12 and 13: the 4-methylnonane and 2-methyloctane do not follow the general trend**

Figure 3 thus shows the results as presented using equations 12 and 13, for all compounds investigated in this work (see table 1 of supporting information). Linear n-alkanes fall onto the zero-zero position. It can be observed that all branched alkanes have negative values for $\Delta h_{vap_{isomer}(shift)}$ and $T_{b_{isomer}(shift)}$: they are always more volatile compared to the corresponding n-alkane. This phenomenon increases as the number of branches increases. It may be related to an increased bulkiness of the molecules. Considering only single branches (mono-methyls), the trends are somewhat blurry, but it seems that 2-methyls behave as if they were more volatile (more linear or less bulky) compared to 3-methyls. Yet, looking at 4-methyls, the trend is opposite: the farther away the methyl group lies from the chain end, the bulkier it behaves. Although the number of available data is lower for dimethyls, the trend observed for 2,2-dimethyls and 3,3-dimethyls is similar as that observed for 2- and 3-monomethyls: the 3,3-dimethyls behave as less 'bulky' than 2,2-dimethyls.

Diagrams such as figure 3 make it possible to zoom and visualize the effect of molecular topology. If a trend can be made visible on such a plot, then our goal is to ensure that the group contribution method is capable to capture this trend. It can give rationale to decisions about what substances require a more refined or individualized parameterization, and where neighboring groups are particularly relevant.

For example, if 2-methyl-, 3-methyl-, or 4-methyl-nonane do not fall approximately within the same location in the diagram, one cannot expect a simple GC-approach (i.e. unable to distinguish among isomers) to be successful.

In addition, if a substance falls out of the expected corridor in the Δh^{vap} - T_b diagram one can speculate on whether the experimental data is reliable. In figure 3, we observe two such cases, which are 2-methyloctane and 4-methylnonane, which we have discarded from our experimental database. We recognize that this logic implies that the phenomenon that is considered (here vapor pressure) can be entirely described by considering only the number of groups. Examples can be found where such trend does not exist (example is melting point of di-acids). The reason for this is that other phenomena, that cannot be described by a simple GC scheme (as for example steric issues related to the molecular conformation), have a significant effect on the property at hand.

4. Improved GC method for the heterosegmented PC-SAFT approach

4.1 Principle

In this work, we suggest to further improve the heterosegmented SAFT approach by considering that group parameters may be affected by neighboring groups. Such considerations have already been investigated by Wu & Sandler [55] who used ab initio calculations to help GC developers define molecular fractions that truly can be considered transferable. Unfortunately, their work leads to the conclusions that rather large groups should be used, thus leading to the need of a large amount of data and reducing the scope of GC methods.

We therefore considered the method proposed by Vijande [24] who suggests using a correction term on group parameters that is inversely proportional to the distance with other groups, according to

$$\Pi_{\boldsymbol{\alpha}} = \Pi_{\boldsymbol{\alpha}}{}^{\mathbf{0}} + \sum_{\boldsymbol{\beta}\neq\boldsymbol{\alpha}} \frac{\Lambda_{\boldsymbol{\alpha\beta}}}{\boldsymbol{P}_{\boldsymbol{\alpha\beta}}} \qquad (14)$$

Where $\Pi_{\boldsymbol{\alpha}}$ can be any parameter (such as ε, σ, m) of the considered group α, $\Pi_{\boldsymbol{\alpha}}{}^{\mathbf{0}}$ is the initial parameter-value of a simple GC-PC-SAFT EoS (without correction), further, index β runs through all groups of the molecule and $\Lambda_{\alpha\beta}$ is the correction parameter for each pair of groups (indices are not interchangeable). The quantity $P_{\alpha\beta}$ is the (integer-valued) 'distance' (in terms of bonds) between two functional groups. As an example, consider n-butane. At first, each individual group is labeled (shown here as superscripts):
$CH_3^{(1)} – CH_2^{(2)}$ - $CH_2^{(3)}$ - $CH_3^{(4)}$
Then, the distance between the leftmost CH_3 (α = 1) and the first CH_2 (β = 2) is $P_{\alpha\beta}$=1.

To decide on which parameters (ε, σ, m) the neighborhood-correction should be applied, we use the assumption initially used by Tamouza at al. [26] which is to consider that the branching should be captured by the segment number parameter m_α. It is then observed that

the effect of the closest neighbor (i.e. when $P_{\alpha\beta}$ =1), is often quite different from that of groups that are farther away. Hence, it was decided to use a specific neighborhood parameter for the closest neighbors, thus modifying the original equation as follows:

$$\boldsymbol{m_\alpha} = \boldsymbol{m_\alpha}^{\mathbf{0}} + \sum_{\boldsymbol{\beta}(\boldsymbol{P_{\alpha\beta}=1})} \left(\frac{\Lambda_{(\alpha\beta)1}}{\boldsymbol{P_{\alpha\beta}=1}}\right) + \sum_{\boldsymbol{\beta}(\boldsymbol{P_{\alpha\beta}>1})} \left(\frac{\Lambda_{(\alpha\beta)N}}{\boldsymbol{P_{\alpha\beta}}}\right) \qquad (15)$$

Where the two parameters $\Lambda_{(\alpha\beta)1}$ and $\Lambda_{(\alpha\beta)N}$ refer to the neighborhood-parameter for immediate neighboring groups or for more distant groups, respectively.

As an example, we apply equation 15 to the 2,2,3,5-tetramethylhexane figure 4 and correct the segment number parameter m of the >C< group (number 2) by considering proximity effects with CH groups. There are two CH groups in the molecule. The first one (number 5) is at a distance $P_{\alpha\beta}$=1, so that the correction parameter to be used is $\Lambda_{(>C<-\,CH)_1}$ the second one (group number 8) is located at $P_{\alpha\beta}$=3 from the group we consider, so that the correction parameter to be used is $\Lambda_{(>C<-\,CH)_N}$. The equation is then :

$$m_{>C<^{(2)}} = m_{>C<}{}^{0} + \left(\frac{\Lambda_{\left(>C<^{(2)}-\,CH^{(5)}\right)_1}}{1}\right) + \left(\frac{\Lambda_{\left(>C<^{(2)}-\,CH^{(8)}\right)_N}}{3}\right)$$

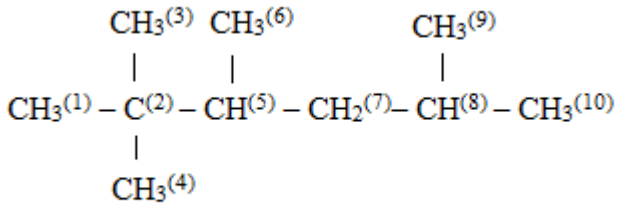

Figure 4: example on the use of the neighborhood parameter for the correction of a >C< parameter as a function of the distance with CH goups.

4.2 First regression: without neighborhood-correction, to investigate which families require correction

All the group parameters (ε, σ or m) are first regressed simultaneously to the entire database listed in table 1 of the supporting information. We use vapor pressure and liquid density correlations from DIPPR choosing 11 points for each alkane from the temperature range containing the accepted experimental data and excluding the most uncertain data identified earlier (section 3). Unlike Sauer et al [49], only correlated values rather than true experimental data are used and consequently, the number of points (Np) is 11 in our case for each alkane, independently of the number of available experimental points for that same alkane. All compounds contribute therefore with the same weight to the objective function.

The resulting regression deviations and parameters are given in table 2 and 3 where the percentage average absolute deviation (%AAD) is defined as

$$\%\mathbf{AAD} = \frac{1}{N_p}\sum_{i=1}^{Np}\left|\frac{w_i^{exp}-w_i^{calc}}{w_i^{exp}}\right| \quad (16)$$

with w either representing vapor pressure, or liquid molar volume, and N_p denoting the number of data-points for each type of data.

***Table 2: Adjusted Alkane Group Parameters of the Heterosegmented GC-PC-SAFT approach* . The parameters for the CH3, CH2 and CH are final; those of >C< are further refined in section 4.3 and table 4.**

Group	m_α	σ_α,Å	ε_α/k,K
CH3	0.749403	3.70393	185.918
CH2	0.794548	3.02763	156.784
>CH	0.564269	0.870882	252.16
>C<	-0.737247	-0.456943	0

Table 3: regression deviations obtained from this work without neighborhood-correction (detailed results available in table 1, supporting information)

	n-alkanes	monomethyls	Dimethyls with >C<	Dimethyls with no >C<	Trimethyls with 1 **>C<**	multimethyls
% AAD in Vapor pressure	4.6	2.8	10.9	7.2	7.1	11.2
% AAD in liquid density	1.1	4.5	6.5	8.1	2.6	8.6

In table 3, the average percent absolute deviation in vapor pressure represented for each family of alkanes shows that uncertainties on dimethyls without quaternary carbon, dimethyls with quaternary carbon, trimethyls and multimethyls are still higher than the expected uncertainties (of ≃ 5%). In addition, one notices that the parameters of the quaternary carbon have no physical sense (negative values of m_α and σ_α). The value of ε_α/k was set to zero because this gave the best results. The quaternary carbon has no external surface and therefore does not contribute to the dispersive energy. Improvement should thus be proposed to try to meet those experimental uncertainties on the identified families.

4.3 Second regression: using neighborhood-correction, to improve the results

The same method of regression as in section 4.2 was used but applying the neighborhood correction, as shown in equation 15. The fitted parameters are now the three group parameters

of the quaternary C, as well as $\Lambda_{(\alpha-\beta)1}$ and $\Lambda_{(\alpha-\beta)N}$ between the quaternary group and some other groups as shown in table 4. Note that the $\Lambda_{(\alpha-\beta)}$ are not symmetric, but that it was decided, in view of reducing the number of adjustable parameters, to set the correction parameters between >CH and >C< equal to those between >C< and >CH. A total of 9 parameters were thus regressed. The resulting values are given in table 4 and deviations are improved as shown in table 5. All detailed deviations are detailed in table 1 of the supporting information. Only the quaternary group parameters are modified in table 4 as compared to table 2.

Table 4: Adjusted Alkane group parameters of the Heterosegmented GC-PC-SAFT approach by neighborhood-correction (the used database is listed in table 1 of the supporting information)

Group	m_α	σ_α,Å	ε_α/k,K
>C<	0.131697	0.923746	0

Group α	**Group β**	$\Lambda_{(\alpha-\beta)1}$	$\Lambda_{(\alpha-\beta)N}$
>C<	CH3	0	-0.16937
>CH >C<	>C< >CH	-0.0505477	0.0606485
>C<	>C<	-0.0213632	0.0793068

Table 5:* % AAD in Vapor pressure *obtained with and without introducing neighborhood-corrections.

	Dimethyls with >C<	Trimethyls with 1 >C<	multimethyls
Without nighborhood-corrections	10.9	7.1	11.2
With neighborhood-corrections	2.4	5.9	8.7

Table 5 shows that the heterosegmented GC-PC-SAFT with neighborhood-corrections improves the deviations for all the families of alkanes where proximity effects have been taken into account (dimethyls with >C<, trimethyls with 1 >C<, and multimethyls).

We have also investigated the possible improvement brought by the proximity correction (equation 15) to other compounds (e.g. containing two CH groups), but no significant improvement could be achieved.

The results with and without neighborhood-corrections for all alkanes used in the regression process are presented in table 1 of the supporting information. The liquid densities were not

discussed here because the improvements seem to mainly affect vapor pressure calculations without worsening the prediction of saturated liquid densities.

5. Analysis and discussion of results

The improvement can now also be illustrated with a Δh^{vap} - T_b diagram as shown in figure 5. The blue diamonds show experimental data obtained from the DIPPR correlations, red squares are computed using parameters from the first regression section 4.2 while green triangles are results using the neighborhood-parameters from this work (section 4.3). The first regression, is an approach similar to the original approach of Sauer et al. [49] and made it difficult to distinguish isomers (such as the 2,2-dimethylhexane and the 3,3-dimethylhexane) compared to the method provided in this work that uses the neighborhood correction.

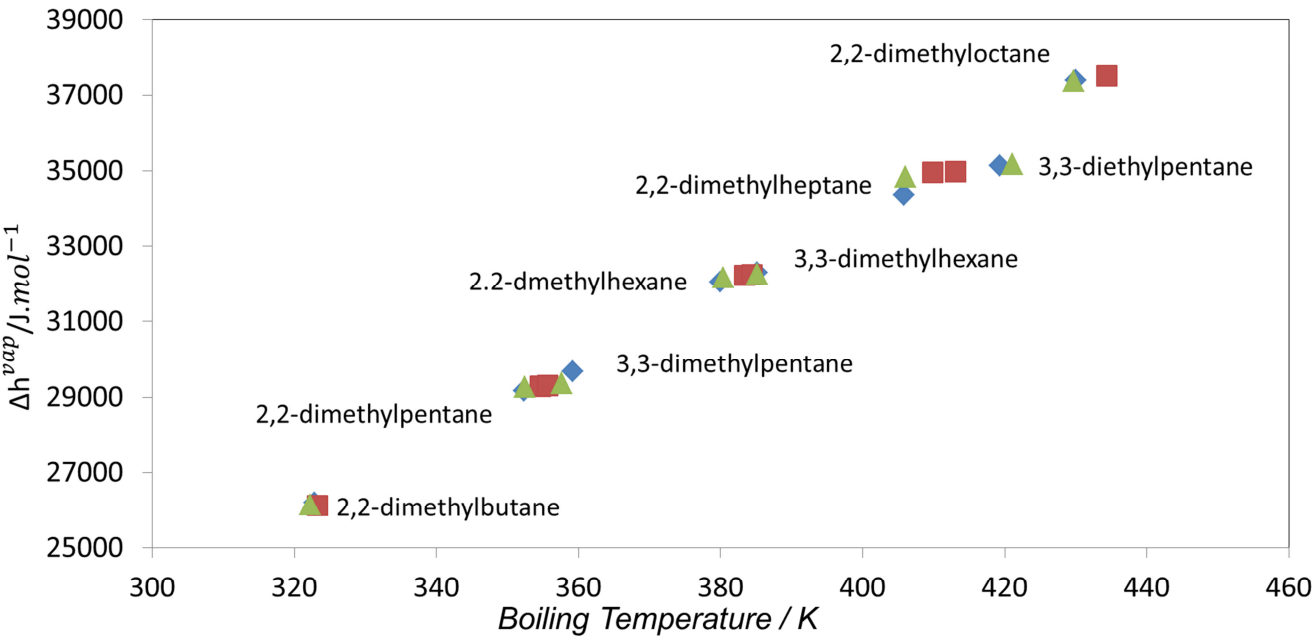

Figure 5: experimental data from DIPPR (blue diamonds) compared to calculated data using the first regression (red squares) and the second regression (green triangles).

The predicted enthalpy of vaporization of alkanes at their normal boiling point using the heterosegmented GC-PC-SAFT EoS are very satisfactory compared to other GC methods [56] specialized for this particular property: 0.47% relative deviation from experimental data in our work compared to 2.7 % in the cited work.

It is also worth stressing the quality of the results obtained for the normal boiling point calculations compared to a number of well-known group contribution methods. Table 6 shows that the GC-SAFT approach predicts this property with an accuracy below 0.7% for all families, which cannot be attained by the selected methods. Detailed results are available in table 2 of the supporting information.

Table 6: Deviations in the calculation of normal boiling temperatures per family using different group contribution methods as compared to experimental data from DIPPR [43]

	Alkanes	Methyls	Dimethyls with >C<	Dimethyls with no >C<	Trimethyls with 1 >C<	multimethyls
% AAD heterosegmented GC-PC-SAFT with neighborhood correction	0.3	0.3	0.1	0.6	0.2	0.6
% AAD Marrero et Gani [3]	3.1	3.0	1.5	1.2	2.0	1.5
% AAD Constantinou [4]	3.2	2.5	1.0	0.7	1.8	2.3

6. Conclusion

The heterosegmented GC-PC-SAFT model was shown by Sauer et al. [49] to be superior to the homosegmented approach. Results for branched alkanes, however, show higher deviations than expected when analyzing vapor pressure experimental data. This can be explained by the fact that neighboring groups may perturb the electron density of a given group. To alleviate the problem, this study proposes neighborhood-corrections, where group-contribution parameters are incremented depending on the distance to other significant groups. The average deviations in vapor pressure calculations on the branched families containing a quaternary carbon have thus decreased from 39 % to 27 %.

When considering complex molecules, which is the aim of our study, it is expected that the regular group contribution approach will not work. The proposed modification can then be extended to adjust other parameters or other groups.

The liquid densities were not discussed here because the improvements seem to mainly affect vapor pressure calculations without worsening the prediction of saturated liquid densities. Other properties will probably also be affected by the proposed approach but were not investigated here.

Références

[1] Poling B.E., Prausnitz J.M., O'connell J.P., *The properties of gases and liquids*, Mcgraw-hill New York, **2001**.

[2] Joback K.G., Reid R.C., estimation of pure-component properties from group-contribution, *Chemical Engineering Communications*, **1987**, *57(1-6)*, p. 233.

[3] Marrero J., Gani R., Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria*, **2001**, *183-184*, p. 183.

[4] Constantinou L., Gani R., New group contribution method for estimating properties of pure compounds, *AIChE J.*, **1994**, *40(10)*, p. 1697.

[5] De Hemptinne J.C, Mougin P., Barreau A., Ruffine L., Tamouza S., Inchekel R., Application to Petroleum Engineering of Statistical Thermodynamics – Based Equations of State, *Oil & Gas Science and Technology - Rev. IFP*, **2006**, *61(3)*, p. 363.

[6] Wertheim M.S., Fluids with highly directional attractive forces. I. Statistical thermodynamics, *J Stat Phys*, **1984**, *35(1-2)*, p. 19.

[7] Wertheim M.S., Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations, *J Stat Phys*, **1984**, *35(1-2)*, p. 35.

[8] Wertheim M.S., Fluids with highly directional attractive forces. III. Multiple attraction sites, *J Stat Phys*, **1986**, *42(3-4)*, p. 459.

[9] Wertheim M.S., Fluids with highly directional attractive forces. IV. Equilibrium polymerization, *J Stat Phys*, **1986**, *42(3-4)*, p. 477.

[10] Wertheim M.S., Fluids of dimerizing hard spheres, and fluid mixtures of hard spheres and dispheres, *The Journal of chemical Physics*, **1986**, *85(5)*, p. 2929.

[11] Huang S.H., Radosz M., Equation of state for small, large, polydisperse, and associating molecules, *Ind. Eng. Chem. Res.*, **1990**, *29(11)*, p. 2284.

[12] Chen S.S., Kreglewski A., Applications of the Augmented van der Waals Theory of Fluids: I. Pure Fluids, *Berichte der Bunsengesellschaft für physikalische Chemie*, **1977**, *81(10)*, p. 1048.

[13] Fu Y.-H., Sandler S.I., A Simplified SAFT Equation of State for Associating Compounds and Mixtures, *Ind. Eng. Chem. Res.*, **1995**, *34(5)*, p. 1897.

[14] Kraska T., Gubbins K.E., Phase Equilibria Calculations with a Modified SAFT Equation of State. 1. Pure Alkanes, Alkanols, and Water, *Ind. Eng. Chem. Res.*, **1996**, *35(12)*, p. 4727.

[15] Kraska T., Gubbins K.E., Phase Equilibria Calculations with a Modified SAFT Equation of State. 1. Pure Alkanes, Alkanols, and Water, *Ind. Eng. Chem. Res.*, **1996**, *35(12)*, p. 4727.

[16] Blas F.J., Vega L.F., Prediction of Binary and Ternary Diagrams Using the Statistical Associating Fluid Theory (SAFT) Equation of State, *Ind. Eng. Chem. Res.*, **1998**, *37(2)*, p. 660.

[17] Gil-Villegas A., Galindo A., Whitehead P.J., Mills S.J., Jackson G., Burgess A.N., Statistical associating fluid theory for chain molecules with attractive potentials of variable range, *The Journal of chemical Physics*, **1997**, *106(10)*, p. 4168.

[18] Gross J., Sadowski G., Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules, *Ind. Eng. Chem. Res.*, **2001**, *40(4)*, p. 1244.

[19] Zhang Y., Chapman W.G., Modeling thermodynamic properties of isomeric alkanes with a new branched equation of state, *Ind. Eng. Chem. Res.*, **2018**.

[20] NguyenHuynh D., Passarello J.-P., Tobaly P., de Hemptinne J.C, Application of GC-SAFT EOS to polar systems using a segment approach, *Fluid Phase Equilibria*, **2008**, *264(1-2)*, p. 62.

[21] Gross J., Vrabec J., An equation-of-state contribution for polar components: Dipolar molecules, *AIChE J.*, **2006**, *52(3)*, p. 1194.

[22] Huynh D.N., Benamira M., Passarello J.-P., Tobaly P., de Hemptinne J.C, Application of GC-SAFT EOS to polycyclic aromatic hydrocarbons, *Fluid Phase Equilibria*, **2007**, *254(1-2)*, p. 60.

[23] Vijande J., Piñeiro M.M., Bessières D., Saint-Guirons H., Legido J.L., Description of PVT behaviour of hydrofluoroethers using the PC-SAFT EOS, *Phys. Chem. Chem. Phys.*, **2004**, *6(4)*, p. 766.
[24] Vijande J., Piñeiro M.M., Legido J.L., Bessières D., Group-Contribution Method for the Molecular Parameters of the PC-SAFT Equation of State Taking into Account the Proximity Effect. Application to Nonassociated Compounds, *Ind. Eng. Chem. Res.*, **2010**, *49(19)*, p. 9394.
[25] Vijande J., Piñeiro M.M., Legido J.L., Group-Contribution Method with Proximity Effect for PC-SAFT Molecular Parameters. 2. Application to Association Parameters: Primary Alcohols and Amines, *Ind. Eng. Chem. Res.*, **2014**, *53(2)*, p. 909.
[26] Tamouza S., Passarello J.-P., Tobaly P., de Hemptinne J.C, Group contribution method with SAFT EOS applied to vapor liquid equilibria of various hydrocarbon series, *Fluid Phase Equilibria*, **2004**, *222-223*, p. 67.
[27] Nguyen Thi T.X., Tamouza S., Tobaly P., Passarello J.-P., de Hemptinne J.C, Application of group contribution SAFT equation of state (GC-SAFT) to model phase behaviour of light and heavy esters, *Fluid Phase Equilibria*, **2005**, *238(2)*, p. 254.
[28] Alej A.G.L.A.D., The thermodynamics of mixtures and the corresponding mixing rules in the SAFT-VR approach for potentials of variable range, *Molecular Physics*, **1998**, *93(2)*, p. 241.
[29] Gross J., Sadowski G., Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains, *Fluid Phase Equilibria*, **2000**, *168(2)*, p. 183.
[30] NguyenHuynh D., A modified group-contribution PC-SAFT equation of state for prediction of phase equilibria, *Fluid Phase Equilibria*, **2016**, *430*, p. 33.
[31] Tihic A., Kontogeorgis G.M., Solms N. von, Michelsen M.L., Constantinou L., A Predictive Group-Contribution Simplified PC-SAFT Equation of State: Application to Polymer Systems, *Ind. Eng. Chem. Res.*, **2008**, *47(15)*, p. 5092.
[32] Solms N. von, Michelsen M.L., Kontogeorgis G.M., Computational and Physical Performance of a Modified PC-SAFT Equation of State for Highly Asymmetric and Associating Mixtures, *Ind. Eng. Chem. Res.*, **2003**, *42(5)*, p. 1098.
[33] Burgess W.A., Tapriyal D., Gamwo I.K., Wu Y., McHugh M.A., Enick R.M., New Group-Contribution Parameters for the Calculation of PC-SAFT Parameters for Use at Pressures to 276 MPa and Temperatures to 533 K, *Ind. Eng. Chem. Res.*, **2014**, *53(6)*, p. 2520.
[34] Nguyen-Huynh D., Passarello J.-P., Tobaly P., de Hemptinne J.C, Modeling Phase Equilibria of Asymmetric Mixtures Using a Group-Contribution SAFT (GC-SAFT) with a kij Correlation Method Based on London's Theory. 1. Application to CO 2 + n -Alkane, Methane + n -Alkane, and Ethane + n -Alkane Systems, *Ind. Eng. Chem. Res.*, **2008**, *47(22)*, p. 8847.
[35] Nguyen-Huynh D., Tran T.K.S., Tamouza S., Passarello J.-P., Tobaly P., de Hemptinne J.C, Modeling Phase Equilibria of Asymmetric Mixtures Using a Group-Contribution SAFT (GC-SAFT) with a kij Correlation Method Based on London's Theory. 2. Application to Binary Mixtures Containing Aromatic Hydrocarbons, n -Alkanes, CO 2 N 2 and H 2 S, *Ind. Eng. Chem. Res.*, **2008**, *47(22)*, p. 8859.
[36] NguyenHuynh D., Passarello J.-P., Tobaly P., In Situ Determination of Phase Equilibria of Methyl Benzoate + Alkane Mixtures Using an Infrared Absorption Method. Comparison with Polar GC-SAFT Predictions †, *J. Chem. Eng. Data*, **2009**, *54(6)*, p. 1685.
[37] Tran T.K.S., NguyenHuynh D., Ferrando N., Passarello J.-P., de Hemptinne J.C, Tobaly P., Modeling VLE of H 2 + Hydrocarbon Mixtures Using a Group Contribution SAFT with a kij Correlation Method Based on London's Theory, *Energy Fuels*, **2009**, *23(5)*, p. 2658.
[38] Mourah M., NguyenHuynh D., Passarello J.P., de Hemptinne J.C, Tobaly P., Modelling LLE and VLE of methanol+n-alkane series using GC-PC-SAFT with a group contribution kij, *Fluid Phase Equilibria*, **2010**, *298(1)*, p. 154.

[39] Banaszak M., Chen C.K., Radosz M., Copolymer SAFT Equation of State. Thermodynamic Perturbation Theory Extended to Heterobonded Chains, *Macromolecules*, **1996**, *29(20)*, p. 6481.
[40] Shuklak.P., Chapman W.G., SAFT equation of state for fluid mixtures of hard chain copolymers, *Molecular Physics*, **1997**, *91(6)*, p. 1075.
[41] Gross J., Spuhl O., Tumakaka F., Sadowski G., Modeling Copolymer Systems Using the Perturbed-Chain SAFT Equation of State, *Ind. Eng. Chem. Res.*, **2003**, *42(6)*, p. 1266.
[42] Paduszyński K., Domańska U., Heterosegmented Perturbed-Chain Statistical Associating Fluid Theory as a Robust and Accurate Tool for Modeling of Various Alkanes. 1. Pure Fluids, *Ind. Eng. Chem. Res.*, **2012**, *51(39)*, p. 12967.
[43] Peters F.T., Laube F.S., Sadowski G., Development of a group contribution method for polymers within the PC-SAFT model, *Fluid Phase Equilibria*, **2012**, *324*, p. 70.
[44] Peng Y., Goff K.D., dos Ramos, M. Carolina, McCabe C., Developing a predictive group-contribution-based SAFT-VR equation of state, *Fluid Phase Equilibria*, **2009**, *277(2)*, p. 131.
[45] McCabe C., Gil-Villegas A., Jackson G., DEL RIO F., The thermodynamics of heteronuclear molecules formed from bonded square-well (BSW) segments using the SAFT-VR approach, *Molecular Physics*, **1999**, *97(4)*, p. 551.
[46] Peng Y., Zhao H., McCabe C., On the thermodynamics of diblock chain fluids from simulation and heteronuclear statistical associating fluid theory for potentials of variable range, *Molecular Physics*, **2006**, *104(4)*, p. 571.
[47] Lymperiadis A., Adjiman C.S., Galindo A., Jackson G., A group contribution method for associating chain molecules based on the statistical associating fluid theory (SAFT-gamma), *The Journal of chemical Physics*, **2007**, *127(23)*, p. 234903.
[48] Lymperiadis A., Adjiman C.S., Jackson G., Galindo A., A generalisation of the SAFT- group contribution method for groups comprising multiple spherical segments, *Fluid Phase Equilibria*, **2008**, *274(1-2)*, p. 85.
[49] Sauer E., Stavrou M., Gross J., Comparison between a Homo- and a Heterosegmented Group Contribution Approach Based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State, *Ind. Eng. Chem. Res.*, **2014**, *53(38)*, p. 14854.
[50] Rowley R. L., Wilding W. V., Oscarson J. L., and Giles N. F., "DIPPR® Data Compilation of Pure Chemical Properties", Design Institute for Physical Properties, *AIChE J.*, **2016**.
[51] Cox K.R., Chapman W.G., The Properties of Gases and Liquids, 5th Edition *J. Am. Chem. Soc.*, **2001**, *123(27)*, p. 6745.
[52] Mathias P.M., Watanasiri S., Some Patterns of Fluid Phase Behavior, *J. Chem. Eng. Data*, **2011**, *56(4)*, p. 1658.
[53] Rozmus J., de Hemptinne J.C, *Equation d'état électrolyte prédictive pour le captage du CO2*, **2012**.
[54] Mathias P.M., Jacobs G., Cabrera J., Modified Trouton's Rule for the Estimation, Correlation, and Evaluation of Pure-Component Vapor Pressure, *J. Chem. Eng. Data*, **2018**, *63(4)*, p. 943.
[55] Wu H.S., Sandler S.I., Use of ab initio quantum mechanics calculations in group contribution methods. 1. Theory and the basis for group identifications, *Ind. Eng. Chem. Res.*, **1991**, *30(5)*, p. 881.
[56] Jia Q., Wang Q., Ma P., Prediction of the Enthalpy of Vaporization of Organic Compounds at Their Normal Boiling Point with the Positional Distributive Contribution Method, *J. Chem. Eng. Data*, **2010**, *55(12)*, p. 5614.

Table 1: Supporting information

			%AAD Vapor pressures				%AAD liquid densities	
	NP	T range (K)	without neighborhood-correction	with neighborhood-correction	NP	T range (K)	without neighborhood-correction	with neighborhood-correction
n-butane	11	140-420	12.6	12.6	11	311-394	2.4	2.4
n-pentane	11	223-470	11.2	11.2	11	143-443	1.5	1.5
n-hexane	11	196-503	10.9	10.9	11	193-478	1.1	1.1
n-heptane	11	194-540	7.9	7.9	11	183-513	0.9	0.9
n-octane	11	228-569	7.6	7.6	11	333-533	0.7	0.7
n-nonane	11	311-511	6.6	6.6	11	223-533	0.5	0.5
n-decane	11	268-490	3.2	3.2	11	244-583	0.6	0.6
n-undecane	11	348-499	3.6	3.6	11	253-573	1.0	1.0
n-dodecane	11	289-520	2.5	2.5	11	263-623	1.6	1.6
n-tridecane	11	290-540	2.4	2.4	11	273-573	1.2	1.2
n-tetradecane	11	280-559	1.7	1.7	11	283-523	1.1	1.1
n-pentadecane	11	346-577	1.4	1.4	11	283-543	1.4	1.4
n-hexadecane	11	295-594	1.2	1.2	11	291-563	1.7	1.7
n-heptadecane	11	390-609	3.0	3.0	11	295-613	0.9	0.9
n-octadecane	11	318-624	2.3	2.3	11	302-593	2.3	2.3
n-nonadecane	11	306-638	2.8	2.8	11	305-603	2.5	2.5
n-eicosane	11	344-652	3.8	3.8	11	310-613	2.4	2.4
n-henicosane	11	318-630	3.7	3.7	11	314-372	0.3	0.3
n-docosane	11	318-668	1.3	1.3	11	318-373	0.3	0.3
n-tricosane	11	418-693	4.0	4.0	11	321-372	0.3	0.3
n-tetracosane	11	379-588	4.1	4.1	11	343-343	0.2	0.2
n-pentacosane	11	379-674	3.0	3.0	11	293-373	0.5	0.5
n-hexacosane	11	379-684	4.3	4.3	11	330-431	0.7	0.7
n-heptacosane	11	351-543	11.1	11.1	11	333-372	0.2	0.2
n-octacosane	11	418-588	3.9	3.9	11	335-573	2.4	2.4
n-nonacosane	11	456-714	0.3	0.3	11	343-343	0.3	0.3

n-triacontane	11	463-722	1.3	1.3	11	343-363	0.2	0.2
n-dotriacontane	11	475-739	4.1	4.1	11	348-373	0.2	0.2
n-hexatriacontane	11	497-769	8.7	8.7	11	353-573	2.7	2.7
mean alkanes			4.6	4.6			1.1	1.1
2-methylpropane	11	239-407	2.2	2.2	11	213-378	9.1	9.1
2-methyl-butane	11	217-448	6.3	6.3	11	150-433	6.4	6.4
2-methylpentane	11	241-498	2.1	2.1	11	258-333	5.3	5.3
3-methylpentane	11	243-504	7.7	7.7	11	266-339	3.7	3.7
2-methylhexane	11	271-530	0.6	0.6	11	272-503	5.7	5.7
3-methylhexane	11	272-535	1.5	1.5	11	233-503	3.2	3.2
2-methylheptane	11	222-560	1.8	1.8	11	263-523	4.4	4.4
3-methylheptane	11	223-564	1.0	1.0	11	263-533	3.5	3.5
4-methylheptane	11	293-562	2.8	2.8	11	263-533	3.5	3.5
3-methyloctane	11	306-445	0.9	0.9	11	283-313	3.2	3.2
4-methyloctane	11	305-443	6.1	6.1	11	283-313	3.3	3.3
2-methylnonane	11	325-469	2.8	2.8	11	273-313	3.8	3.8
3-methylnonane	11	325-470	0.8	0.8	11	273-313	3.0	3.0
mean methyls			2.8	2.8			4.5	4.5
2,2-dimethylpropane	11	258-434	26.6	4.5	11	253-403	4.0	4.6
2,2-dimethylbutane	11	199-489	8.7	4.7	11	269-320	6.8	6.3
2,2-dimethylpentane	11	261-513	11.5	0.7	11	150-493	5.3	5.4
3,3-dimethylpentane	11	213-536	4.9	3.7	11	272-503	8.1	6.7
2,2-dimethylhexane	11	272-513	11.6	1.3	11	276-550	4.9	5.2
3,3-dimethylhexane	11	279-562	2.7	0.8	11	272-533	10.2	9.3
2,2-dimethylheptane	11	303-433	11.7	2.7	11	293-298	3.6	3.7
2,2-dimethyloctane	11	316-458	9.4	1.2	11	293-298	3.2	3.3
mean dimethyls with >C<			10.9	2.4			6.5	6.1
2,3-dimethylbutane	11	238-500	12.1	12.1	11	266-473	11.4	11.4
2,3-dimethylpentane	11	205-523	9.2	9.2	11	272-503	8.1	8.1
2,4-dimethylpentane	11	262-520	8.2	8.2	11	272-493	10.6	10.6
2,3-dimethylhexane	11	283-563	2.5	2.5	11	272-533	7.9	7.9
2,4-dimethylhexane	11	278-553	9.2	9.2	11	272-523	9.5	9.5

2,5-dimethylhexane	11	278-550	10.2	10.2	11	255-513	9.4	9.4
3,4-dimethylhexane	11	272-533	5.1	5.1	11	284-569	6.7	6.7
2,6-dimethylheptane	11	299-435	6.1	6.1	11	286-360	8.1	8.1
2,3-dimethyloctane	11	321-466	4.5	4.5	11	293-298	6.1	6.1
2,4-dimethyloctane	11	316-457	8.3	8.3	11	281-298	7.1	7.1
2,5-dimethyloctane	11	317-460	4.2	4.2	11	293-298	6.6	6.6
2,6-dimethyloctane	11	319-462	5.7	5.7	11	293-298	7.0	7.0
2,7-dimethyloctane	11	271-461	7.6	7.6	11	273-393	7.3	7.3
mean dimethyls with no >C<			7.2	7.2			8.1	8.1
2,2,3-trimethylbutane	11	254-528	3.3	3.8	11	261-503	2.5	1.4
2,2,3-trimethylpentane	11	284-564	2.8	1.7	11	272-533	3.9	3.2
2,2,4-trimethylpentane	11	199-544	9.1	12.4	11	166-513	1.1	1.1
2,3,3-trimethylpentane	11	280-573	9.8	3.8	11	272-543	5.6	3.3
2,2,5-trimethylhexane	11	206-424	14.8	10.7	11	283-523	0.4	0.6
2,4,4-trimethylhexane	11	300-432	5.6	4.5	11	283-323	2.0	1.5
3,3,5-trimethylheptane	11	313-458	4.0	4.2	11	293-298	2.3	2.0
mean trimethyls with 1 >C<			7.1	5.9			2.6	1.9
2,2,3,3-tetramethylbutane	11	375-406	5.4	5.8	11	374-510	16.8	12.5
2,2,3,3-tetramethylpentane	11	307-442	20.0	7.2	11	288-323	14.7	12.8
2,2,3,4-tetramethylpentane	11	301-434	3.8	4.1	11	288-298	1.0	1.5
2,2,4,4-tetramethylpentane	11	293-423	2.4	10.8	11	288-303	10.5	9.5
2,3,3,4-tetramethylpentane	11	308-443	16.3	1.9	11	283-323	0.9	0.2
2,2,3,3-tetramethylhexane	11	314-463	3.3	6.0	11	293-298	12.4	11.1
2,2,5,5-tetramethylhexane	11	300-438	20.6	11.8	11	293-298	7.2	6.7
2,2,4,4,6,8,8-heptamethylnonane	11	424-545	6.1	7.6	11	298-313	6.6	5.8
2,6,10,15,19,23-hexamethyltetracosane	11	468-536	22.6	22.6	11	273-562	7.1	7.1
mean multimethyls			11.2	8.7			8.6	7.5
3-ethylpentane	11	266-541	2.8	2.8	11	155-513	3.2	3.2
3-ethylhexane	11	286-565	4.7	4.7	11	272-533	1.8	1.8
3-ethylheptane	11	305-384	9.8	9.8	11	293-298	2.5	2.5
3-ethyl-2-methylpentane	11	283-560	3.3	3.3	11	261-533	6.1	6.1
2,2-dimethyl-3-ethylpentane	11	296-435	12.0	12.0	11	293-298	2.5	2.3

2,4-dimethyl-3-ethylpentane	11	298-438	12.9	12.9	11	293-298	9.3	9.3
2,3,4-trimethylpentane	11	223-566	6.1	6.0	11	272-533	12.0	12.0
3,3-diethylpentane	11	304-448	12.6	5.6	11	288-298	9.5	8.3
3-ethyl-3-methylpentane	11	249-576	10.0	2.9	11	273-323	9.0	8.0
others			8.2	6.7			6.2	5.9

Table 2: Supporting information

	heterosegmented GC-PC-SAFT with neighborhood correction	Marrero et Gani [3] (MG)	Constantinou [4] (Const)	Experimental (Exp)[43]	% AAD heterosegmented GC-PC-SAFT with neighborhood correction / Exp	% AAD MG/ Exp	% AAD CONST/ Exp
	T_b (K)	T_b (K)	T_b (K)	T_b (K)			
n-butane	275.5	253.7	263.1	272.65	1.0	7.0	3.5
n-pentane	312.1	299.5	309.5	309.2	0.9	3.1	0.1
n-hexane	345.0	337.4	347.2	341.9	0.9	1.3	1.6
n-heptane	374.7	369.8	379.1	371.6	0.8	0.5	2.0
n-octane	401.8	398.1	406.6	398.9	0.7	0.2	1.9
n-nonane	426.7	423.2	430.9	424.0	0.6	0.2	1.6
n-decane	449.8	445.7	452.6	447.3	0.6	0.3	1.2
n-undecane	471.2	466.2	472.2	469.1	0.5	0.6	0.7
n-dodecane	491.3	485.0	490.1	489.5	0.4	0.9	0.1
n-tridecane	510.1	502.3	506.6	508.6	0.3	1.2	0.4
n-tetradecane	527.8	518.3	521.8	526.7	0.2	1.6	0.9
n-pentadecane	544.6	533.3	535.9	543.8	0.1	1.9	1.5
n-hexadecane	560.5	547.3	549.2	560.0	0.1	2.3	1.9
n-heptadecane	575.6	560.5	561.6	575.3	0.0	2.6	2.4
n-octadecane	590.0	572.9	573.4	589.9	0.0	2.9	2.8
n-nonadecane	603.7	584.7	584.5	603.1	0.1	3.1	3.1
n-eicosane	616.8	595.9	595	616.9	0.0	3.4	3.6
n-henicosane	629.4	606.6	605	629.7	0.0	3.7	3.9
n-docosane	641.5	616.7	614.5	641.8	0.1	3.9	4.3
n-tricosane	653.1	626.5	623.6	653.4	0.0	4.1	4.6
n-tetracosane	664.2	635.8	632.4	664.5	0.0	4.3	4.8
n-pentacosane	675.0	644.7	640.7	675.1	0.0	4.5	5.1
n-hexacosane	685.4	653.3	648.8	685.4	0.0	4.7	5.3

n-heptacosane	695.4	661.6	656.5	685.3	1.5	3.5	4.2
n-octacosane	705.1	669.6	664	704.8	0.0	5.0	5.8
n-nonacosane	714.5	677.3	671.1	714.0	0.1	5.1	6.0
n-triacontane	723.6	684.8	678.1	722.9	0.1	5.3	6.2
n-dotriacontane	740.9	699.0	691.3	738.9	0.3	5.4	6.4
n-hexatriacontane	772.7	724.9	715.4	770.2	0.3	5.9	7.1
mean alkanes					0.3	3.1	3.2
2-methylpropane	260.2	232.3	219.3	261.4	0.5	11.1	16.1
2-methyl-butane	299.1	282.2	287.3	301.0	0.6	6.2	4.6
2-methylpentane	332.7	322.9	328.9	333.4	0.2	3.1	1.3
3-methylpentane	334.0	322.9	333.6	336.4	0.7	4.0	0.8
2-methylhexane	362.9	357.4	363.5	363.2	0.1	1.6	0.1
3-methylhexane	364.3	357.4	367.5	365.0	0.2	2.1	0.7
2-methylheptane	390.5	387.2	393.1	390.8	0.1	0.9	0.6
3-methylheptane	391.9	387.2	396.5	392.1	0.0	1.2	1.1
3-methyloctane	417.2	413.5	422	417.4	0.0	0.9	1.1
2-methylnonane	439.1	437.0	441.9	440.2	0.3	0.7	0.4
3-methylnonane	440.6	437.0	444.6	441.0	0.1	0.9	0.8
mean methyls					0.3	3.0	2.5
2,2-dimethylbutane	322.2	311.3	317.1	322.9	0.2	3.6	1.8
2,2-dimethylpentane	352.5	347.4	353.6	352.3	0.1	1.4	0.4
3,3-dimethylpentane	357.6	347.1	355.3	359.2	0.4	3.4	1.1
2,2-dimethylhexane	380.3	378.5	384.5	380.0	0.1	0.4	1.2
3,3-dimethylhexane	385.1	378.2	386	385.1	0.0	1.8	0.2
2,2-dimethylheptane	406.0	405.8	411.4	405.8	0.1	0.0	1.4
2,2-dimethyloctane	429.7	430.1	435.1	430.1	0.1	0.0	1.2
mean dimethyls with >C<					0.1	1.5	1.0
2,3-dimethylbutane	327.7	324.1	339	333.1	1.6	2.7	1.8
2,3-dimethylpentane	360.2	358.5	357.5	362.9	0.7	1.2	1.5

2,4-dimethylpentane	351.9	347.8	351	353.6	0.5	1.6	0.7
2,3-dimethylhexane	388.4	388.2	387.9	388.8	0.1	0.2	0.2
2,4-dimethylhexane	381.3	375.5	382.3	382.6	0.3	1.9	0.1
2,5-dimethylhexane	379.9	375.4	378.6	382.3	0.6	1.8	1.0
3,4-dimethylhexane	389.8	388.3	391.4	390.9	0.3	0.7	0.1
2,6-dimethylheptane	405.5	403.0	406.2	408.4	0.7	1.3	0.5
2,3-dimethyloctane	437.8	437.8	437.8	437.5	0.1	0.1	0.1
2,4-dimethyloctane	430.5	427.7	433.4	429.1	0.3	0.3	1.0
2,7-dimethyloctane	429.1	427.6	430.6	433.0	0.9	1.2	0.6
mean dimethyls with no >C<					0.6	1.2	0.7
2,2,3-trimethylbutane	352.6	333.6	335.9	354.0	0.4	5.8	5.1
2,2,3-trimethylpentane	382.4	379.7	373.3	383.0	0.2	0.9	2.5
2,2,4-trimethylpentane	372.3	366.4	369.4	372.4	0.0	1.6	0.8
2,3,3-trimethylpentane	386.4	384.7	391.2	387.9	0.4	0.8	0.9
2,2,5-trimethylhexane	397.6	395.1	398.2	397.2	0.1	0.5	0.2
2,4,4-trimethylhexane	403.1	394.8	399.6	403.8	0.2	2.2	1
mean trimethyls with 1 >C<					0.2	2.0	1.8
2,2,3,3-tetramethylbutane	412.3	403.5	431.0	413.4	0.3	2.4	4.3
2,2,3,3-tetramethylpentane	408.8	399.3	403.6	406.2	0.7	1.7	0.6
2,2,3,4-tetramethylpentane	391.0	386.9	389.8	395.4	1.1	2.2	1.4
2,2,4,4-tetramethylpentane	415.0	416.7	407.7	414.7	0.1	0.5	1.7
2,3,3,4-tetramethylpentane	435.9	428.0	452.7	433.5	0.6	1.3	4.4
2,2,3,3-tetramethylhexane	414.0	413.2	416.1	410.6	0.8	0.6	1.3
mean multimethyls					0.6	1.5	2.3