When using equations of state (EOS) to predict the phase behavior of hydrocarbon mixtures, problems occur with the C7+ fraction that exists in such mixtures. Describing the C7+ fraction by a single C7+ fraction was found unsuitable for PVT reservoir fluids. Several methods for C7+ characterization were proposed. Those methods are grouped into two main categories: correlation and splitting and lumping. Correlation refers to the process of predicting C7+ properties solely from its specific gravity and molecular weight and/or true boiling point. However, a single C7+ fraction was found unsuitable for PVT predictions and phase behavior calculations. Splitting refers to the process of breaking down C7+ fraction into a number of pseudocomponents with a single carbon number; C7, C8, C9, Cn. Pseudocomponents are described by the same physical properties used for pure components; which were measured and compiled over the years. The exponential molar parameters or characterizing the C7+ fraction or both. Proper characterization of the plus fraction, however, reduces the need for extensive tuning of the EOS. Thus, C7+ characterization is considered the most important step associated with the description of reservoir fluids.

Several methods for C7+ characterization were proposed. Those methods are grouped into two main categories: correlation and splitting and lumping. Correlation refers to the process of predicting C7+ properties solely from its specific gravity and molecular weight and/or true boiling point. However, a single C7+ fraction was found unsuitable for PVT predictions and phase behavior calculations. Splitting refers to the process of breaking down C7+ fraction into a number of pseudocomponents with a single carbon number; C7, C8, C9, Cn. Pseudocomponents are described by the same physical properties used for pure components; which were measured and compiled over the years. The exponential molar parameters or characterizing the C7+ fraction or both. Proper characterization of the plus fraction, however, reduces the need for extensive tuning of the EOS. Thus, C7+ characterization is considered the most important step associated with the description of reservoir fluids.

Several methods for C7+ characterization were proposed. Those methods are grouped into two main categories: correlation and splitting and lumping. Correlation refers to the process of predicting C7+ properties solely from its specific gravity and molecular weight and/or true boiling point. However, a single C7+ fraction was found unsuitable for PVT predictions and phase behavior calculations. Splitting refers to the process of breaking down C7+ fraction into a number of pseudocomponents with a single carbon number; C7, C8, C9, Cn. Pseudocomponents are described by the same physical properties used for pure components; which were measured and compiled over the years. The exponential molar
distribution (mole fraction/molecular weight relation) is perhaps the simplest method for splitting the C₇₊ fraction into a number of pseudocomponents. The mole fraction of pseudocomponent n in the extended analysis is given by:

$$z_n = z_{C_7} A e^{-Bn}$$  \hspace{1cm} (1)$$

where A and B are fitting parameters. Katz (1983) proposed a simple graphical correlation; which is represented in a mathematical form as follows:

$$z_n = z_{C_7} (1.38205) e^{-0.25903n}$$  \hspace{1cm} (2)$$

Lohrenz et al. (1964) proposed that the C₇₊ fraction could be split into pseudocomponents with carbon numbers that range from 7 to 40. Their correlation is given by:

$$z_n = z_6 e^{4(n-6)^2 + B(n-6)}$$  \hspace{1cm} (3)$$

Ahmed et al. (1985) devised a simple method for splitting the C₇₊ fraction into a number of pseudocomponents. The only input data required are the mole fraction and molecular weight of C₇₊. The authors proposed the following expression for estimating pseudocomponent mole fractions:

$$z_n = z_{n+} \left( \frac{MW_{n+1} - MW_{n+}}{MW_{n+1} - MW_n} \right)$$  \hspace{1cm} (4)$$

Whitson (1980) proposed that the three-parameter gamma probability function can be used to model the molar distribution of the C₇₊ fraction. He expressed the mole fraction of each pseudocomponent in the extended analysis as follows:

$$z_n = \frac{z_{C_7}}{\Gamma(\alpha)} \sum_{j=0}^{\infty} \frac{y_{n+1} \alpha+y_n \alpha-y_{n+1} \alpha-y_n}{\Gamma(\alpha+j+1)}$$  \hspace{1cm} (5)$$

However, it is impractical to list all of the pseudocomponents of the extended analysis generated by splitting the C₇₊ fraction. This is due to the fact that the cost required for compositional reservoir simulation increases substantially with the number of components. Hence lumping is coupled all the way with splitting. It refers to the process of regrouping the pseudocomponents of the extended analysis into a more concise number; say three or four components which are believed to fully describe the C₇₊ fraction. Numerous lumping schemes were developed by different authors. The schemes are either trial-and-error methods, arbitrary rules, or trial-and-error methods within specific guidelines.

Whitson (1980) developed a regrouping scheme to reduce the extended analysis to only a few multiple-carbon-number (MCN) groups. The number of MCN groups needed for adequate description of the C₇₊ fraction is given by:

$$N_{G} = \int \left[ 1 + 3.3 \log \left( N - n \right) \right]$$  \hspace{1cm} (6)$$

The molecular weights separating each MCN group are given by:

$$MW_i = MW_n \left( \frac{MW_n}{MW} \right)^{\frac{i}{N}}$$  \hspace{1cm} (7)$$

2. THE CONTINUOUS DISTRIBUTION FUNCTION

Behrens and Sandler (1988) used a semicontinuous thermodynamic description to model the C₇₊ fraction for EOS calculations. In this description, a petroleum fluid mixture is considered to be composed of two categories of components: discrete components and other remaining hydrocarbon fractions. The discrete components include the light hydrocarbons C₁ to C₆ as well as inorganic gases such as N₂, CO₂ and H₂S. The remainder of the fluid include the heavy fractions and the identifiable components that are too numerous to be considered individually. The heavy fractions of the fluid mixture may be described by a continuous distribution. The authors proposed the distribution function F(I) to model the continuous distribution according to the index I that is chosen to be a property, such as boiling point, carbon number, or molecular weight. They stated that C₇₊ may be fully described by only two pseudocomponents, although the theory and procedure may extend to three or more components.

Whitson (1989) stated that the method is general and can be applied to any molar distribution model and for any number of C₇₊ groups. It gives surprisingly good results. The method uses the Gaussian quadrature with two points. The following equation to model the C₇₊ fraction was presented:

$$F(l) = A + B$$  \hspace{1cm} (8)$$

where F(l) is the distribution function. A and B are the lower limit and upper cutoff of the continuous distribution respectively. Behrens and Sandler pointed out that the parameter A can be effectively defaulted to 7 - ½ = 6½ and the parameter B is the ending carbon number in the extended analysis + ½.

Firoozabadi et al. (1978) presented a generalized set of physical properties for pure components C₆ through C₄₅. Whitson (1984) modified the original tabulated physical properties to make their use more consistent. This suggests that the upper cutoff of the Behrens and Sandler continuous distribution should be 45½. Behrens and Sandler selected the index of the distribution function, I, to be the carbon number, n. They proposed the following exponential form of F(I):

$$N_{G} = \int \left[ 1 + 3.3 \log \left( N - n \right) \right]$$  \hspace{1cm} (6)$$
\[ F(n) = e^{-\alpha n} D(n) \]  
(9)

Thus equation 8 is written as follows:

\[ z_{C7+} = \frac{\int_{A}^{B} F(n) \, dn}{\int_{A}^{B} e^{-\alpha n} D(n) \, dn} \]  
(10)

where \( n \) ranges between A and B. The parameter \( \alpha \) is the slope of the distribution, which is found from the average carbon number, which, in turn, is approximately related to the MW of the heavy fraction as:

\[ \overline{C}_n = \frac{MW_{C7+} + 4}{14} \]  
(11)

The slope of the distribution, \( \alpha \), is then found from equation 12, which calculates the slope of an exponential distribution from A to B to match the specified average carbon number, which is given by equation 11, Behrens and Sandler (1988):

\[ \frac{1}{\alpha} = \overline{C}_n - A + \left[ \frac{(B - A)}{e^{-A\alpha} - e^{-B\alpha}} \right] \]  
(12)

If the upper cutoff \( B \) were infinity, the term in the brackets would vanish and \( 1/\alpha \) could be found explicitly. Because the upper cutoff is finite, \( \alpha \) is evaluated iteratively by the solution of equation 12 which can be rewritten as follows:

\[ f(\alpha) = \frac{1}{\alpha} - \overline{C}_n + A + \left[ \frac{(B - A)}{e^{-A\alpha} - e^{-B\alpha}} \right] \]  
(13)

By transformation of variables and changing the limits of integration from A and B to 0 and c, the following distribution function is obtained:

\[ z_{C7+} = \int_{0}^{c} e^{-\alpha r} D(r) \, dr \]  
(14)

From which it is apparent that:

\[ c = \alpha(B - A) \]  
(15)

and

\[ r = \alpha(n - A) \]  
(16)

Solving equation 16 for \( n \) yields:

\[ n = A + \frac{r}{\alpha} \]  
(17)

The authors applied Gaussian quadrature numerical integration method with a two-point integration to evaluate equation (14). They expressed the integral as follows:

\[ z_{C7+} = \sum_{i=1}^{2} D(r_i)w_i \]  
(18)

Now with c calculated, the roots and weighting factors of Gaussian quadrature are looked up from the Behrens and Sandler roots and weights for the two-point integration. With roots and weights in hand, the pseudocomponent carbon numbers, \( n_i \), and mole fractions, \( z_i \), are calculated from the following expressions (see equation 17):

\[ n_1 = A + \frac{r_1}{\alpha} \]  
(19)

\[ n_2 = A + \frac{r_2}{\alpha} \]  

\[ z_1 = w_1 z_{C7+} \]  
(20)

\[ z_2 = w_2 z_{C7+} \]

3. POLYNOMIAL FIT PROCEDURE

The data required for the solution of the distribution function \( F(I) \) are the mole fraction \( z_{C7+} \), molecular weight \( MW_{C7+} \), and specific gravity \( \gamma_{C7+} \) of \( C_{7+} \). The solution sequence, however, starts with solving equation (13) for the parameter \( \alpha \) as a function of \( MW_{C7+} \). This equation is a nonlinear equation which requires some sort of trapping or bracketing a root and then hunting it down. Root finding invariably proceeds by iteration starting from some approximate trial solution. Frequently, the root-finding algorithm may converge, with ten-digit accuracy, to the wrong root, or it fails to converge to a root because the initial guess was not sufficiently close to it. Therefore, it is really crucial to use a good root-finding algorithm and to start with a good first guess to hunt a proper root.

Polynomials have the obvious advantage of being easy to evaluate directly. They are cheaper and more efficient than solving nonlinear equations. The polynomial fit procedure is summarized by the following steps:

1. Equation (13) was solved for the parameter \( \alpha \) for molecular weights ranging from 110 to 600. The set of calculated values was divided into 3 subsets of data: The first set includes \( \alpha \) values for molecular weights less than 150. The second set includes \( \alpha \) values for molecular weights from 150 to less than 230. The last set includes \( \alpha \) values for molecular weights greater than 230. Each set was fit with a cubic polynomial as given by equation 21. As a quality check, \( \alpha \) values, that were generated using equations 13 and 21, were plotted against molecular weights as shown in figures 1 to 3.

\[ a(MW_{C7+}) = \begin{cases} 
-5.60756(10^6)(MW_{C7+})^{-1} + 2.403665(10^5)(MW_{C7+})^{-2} \\
-3.48039(10^5)(MW_{C7+})^{-3} + 2.767812(10^2)(MW_{C7+})^{-4} \\
-1.58299(10^2)(MW_{C7+})^{-5} + 1.051902(10^1)(MW_{C7+})^{-6} \\
-2.404336(10^1)(MW_{C7+})^{-7} + 1.993476(150 \leq MW_{C7+} \leq 230) \\
-1.520498(10^0)(MW_{C7+})^{-8} + 2.354221(10^1)(MW_{C7+})^{-9} \\
-1.290899(10^0)(MW_{C7+})^{-10} + 0.2937318(230 \leq MW_{C7+}) \end{cases} \]  
(21)
2. Similarly, Behrens and Sandler roots and weighting factors of the two-point Gaussian quadrature \( r_1, r_2, w_1, \) and \( w_2 \) were calculated for \( \alpha \) values from 0.01 to 1.1. Each set was then fit with a cubic spline polynomial as given by equations 22 through 25. As a quality check, roots and weighting factors, that were looked up from the Behrens and Sandler roots and weights and those that were calculated using equations 22 through 25 were plotted against \( \alpha \) as shown in figures 4 to 7.

\[
\begin{align*}
\alpha &= -5.60755990 \times 10^{-6}w^3 + 2.40360487 \times 10^{-3}w^2 - 0.348038956w + 17.2678121, \\
\text{MW } &= 150 \\
\end{align*}
\]

\[
\begin{align*}
\alpha &= -1.58299201 \times 10^{-7}w^3 + 1.05190172 \times 10^{-4}w^2 - 0.0240433556w + 1.99547641, \\
\text{MW } &= 150 \leq \text{MW} < 230 \\
\end{align*}
\]

\[
\begin{align*}
\alpha &= -1.52049788 \times 10^{-9}w^3 + 2.35422084 \times 10^{-6}w^2 - 1.29098936 \times 10^{-3}w + 0.293373182, \\
\text{MW } &= \text{MW} > 230 \\
\end{align*}
\]

3. The pseudocomponent carbon numbers \( n_1 \) and \( n_2 \) and the mole fractions \( z_1 \) and \( z_2 \) are calculated from the following expressions:

\[
\begin{align*}
c_1 &= A + \frac{r_1}{\alpha} \Rightarrow n_1 = \text{int}(c_1) \\
c_2 &= A + \frac{r_2}{\alpha} \Rightarrow n_2 = \text{int}(c_2) \\
z_1 &= W_1 z_{C^7+} \\
z_2 &= W_2 z_{C^7+} \\
\end{align*}
\]

4. Properties of the two pseudocomponents are calculated as follows:

\[
\begin{align*}
\gamma_1 &= \gamma_{n_1} + (c_1 - n_1)(\gamma_{n_{1+1}} - \gamma_{n_1}) \\
\gamma_2 &= \gamma_{n_2} + (c_2 - n_2)(\gamma_{n_{2+1}} - \gamma_{n_2}) \\
MW_1 &= MW_{n_1} + (c_1 - n_1)(MW_{n_{1+1}} - MW_{n_1}) \\
MW_2 &= MW_{n_2} + (c_2 - n_2)(MW_{n_{2+1}} - MW_{n_2}) \\
\end{align*}
\]
A Polynomial Fit to the Continuous Distribution Function for C7+ Characterization

(34)

(35)

(36)

(37)

(38)

(39)

(40)

(41)

(42)

(43)

(44)

(45)

(46)

(47)

(48)

(49)

(50)

Table 1. Coats and Smart (1986) compositional data for four hydrocarbon systems

<table>
<thead>
<tr>
<th></th>
<th>Gas 1</th>
<th>Gas 2</th>
<th>Oil 2</th>
<th>Oil 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6 mol %</td>
<td>1.79</td>
<td>0.90</td>
<td>1.51</td>
<td>2.58</td>
</tr>
<tr>
<td>C7+ mol %</td>
<td>12.20</td>
<td>5.88</td>
<td>16.92</td>
<td>18.51</td>
</tr>
<tr>
<td>SGC7+</td>
<td>0.8115</td>
<td>0.8100</td>
<td>0.8364</td>
<td>0.8275</td>
</tr>
<tr>
<td>MWC7+</td>
<td>193</td>
<td>153</td>
<td>173</td>
<td>189</td>
</tr>
</tbody>
</table>

Table 2. Roland (1945), Hoffmann (1953), Donohoe (1981), and Firoozabadi (1978) Gas Condensates

<table>
<thead>
<tr>
<th></th>
<th>Roland</th>
<th>Hoffmann</th>
<th>Donohoe</th>
<th>Firoozabadi</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6 %</td>
<td>0.63</td>
<td>0.39</td>
<td>1.14</td>
<td>0.72</td>
</tr>
<tr>
<td>C7+ %</td>
<td>1.36</td>
<td>1.54</td>
<td>3.48</td>
<td>3.10</td>
</tr>
<tr>
<td>SGC7+</td>
<td>0.8268</td>
<td>0.7961</td>
<td>0.7763</td>
<td>0.7740</td>
</tr>
<tr>
<td>MWC7+</td>
<td>198</td>
<td>138.78</td>
<td>152.3</td>
<td>132</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSION

Eight data sets were used to check the validity of the polynomial approach compared with results obtained from Behrens and Sandler method of solution. Table 1 summarizes Coats and Smart (1986) compositional data for four hydrocarbon systems: two gas systems and two oil systems. Table 2 summarizes Roland (1945), Hoffmann (1953), Donohoe (1981), and Firoozabadi (1978) reported data sets for four gas condensate systems.

Table 3 summarizes the reported C7+ mole fractions for all data sets along with the calculated mole fractions for the two pseudocomponents generated with Behrens and Sandler method and with the polynomial fitting approach. Figure 8 is a plot of the calculated mole fractions for the two pseudocomponents versus the C7+ mole fractions.

Table 4 summarizes the reported C7+ molecular weights for all data sets along with the calculated molecular weights for the two pseudocomponents generated with Behrens and Sandler method and with the polynomial fitting approach. Figure 9 is a plot of the calculated molecular weights for the two pseudocomponents versus the C7+ molecular weights.
Table 3. Behrens-Sandler and the polynomial calculated mole fractions for the two pseudocomponents for all reported data sets

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behrens</td>
<td>Sandler</td>
</tr>
<tr>
<td>Polynomial</td>
<td>Polynomial</td>
</tr>
<tr>
<td>ZC7+</td>
<td>Z1</td>
</tr>
<tr>
<td>0.0588</td>
<td>0.09865089</td>
</tr>
<tr>
<td>0.1692</td>
<td>0.1401013</td>
</tr>
<tr>
<td>0.1851</td>
<td>0.1502283</td>
</tr>
<tr>
<td>0.0136</td>
<td>0.0109109</td>
</tr>
<tr>
<td>0.0154</td>
<td>0.0131143</td>
</tr>
<tr>
<td>0.0348</td>
<td>0.0294202</td>
</tr>
<tr>
<td>0.0311</td>
<td>0.0264422</td>
</tr>
</tbody>
</table>

Table 4. Behrens-Sandler and the polynomial calculated molecular weights for the two pseudocomponents for all reported data sets

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behrens</td>
<td>Sandler</td>
</tr>
<tr>
<td>Polynomial</td>
<td>Polynomial</td>
</tr>
<tr>
<td>MWc7+</td>
<td>MW1</td>
</tr>
<tr>
<td>193</td>
<td>138.446</td>
</tr>
<tr>
<td>153</td>
<td>123.559</td>
</tr>
<tr>
<td>173</td>
<td>131.917</td>
</tr>
<tr>
<td>189</td>
<td>137.265</td>
</tr>
<tr>
<td>198</td>
<td>139.836</td>
</tr>
<tr>
<td>138.78</td>
<td>116.173</td>
</tr>
<tr>
<td>152.3</td>
<td>123.233</td>
</tr>
<tr>
<td>132</td>
<td>112.315</td>
</tr>
</tbody>
</table>

Table 5 summarizes the reported C7+ specific gravities for all data sets along with the calculated specific gravities for the two pseudocomponents generated with Behrens and Sandler method and with the polynomial fitting approach. Figure 10 is a plot of the calculated specific gravities for the two pseudocomponents versus the C7+ specific gravities.

Table 5. Behrens-Sandler and the polynomial calculated specific gravities for the two pseudocomponents for all reported data sets

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behrens</td>
<td>Sandler</td>
</tr>
<tr>
<td>Polynomial</td>
<td>Polynomial</td>
</tr>
<tr>
<td>g7+</td>
<td>g1</td>
</tr>
<tr>
<td>0.8115</td>
<td>0.76576</td>
</tr>
<tr>
<td>0.81</td>
<td>0.77076</td>
</tr>
<tr>
<td>0.8364</td>
<td>0.77976</td>
</tr>
<tr>
<td>0.8275</td>
<td>0.78476</td>
</tr>
<tr>
<td>0.8268</td>
<td>0.78694</td>
</tr>
<tr>
<td>0.7961</td>
<td>0.76145</td>
</tr>
<tr>
<td>0.7763</td>
<td>0.7704</td>
</tr>
<tr>
<td>0.774</td>
<td>0.75621</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

A polynomial fit to the continuous distribution function F(I) for C7+ characterization was presented. This offers flexibility in programming and increases the computational efficiency, since polynomials have the obvious advantage of being easy to evaluate directly; in addition to the fact that the C7+ characterization represents a minor portion of the compositional simulation process.

REFERENCES

A Polynomial Fit to the Continuous Distribution Function for C7+, Characterization

presented at the 60th Annual Technical Conference of the SPE, Las Vegas, USA.


NOMENCLATURE

MW<sub>n</sub> molecular weight of the pseudo-component with n carbon atoms, lb/lb-mol

MW<sub>N</sub> molecular weight of last component in the extended system, lb/lb-mol

MW<sub>C7+</sub> molecular weight of the C7+ fraction in the hydrocarbon system, lb/lb-mol

I running index of Whitson's lumping scheme; i.e. 1, 2, 3... N<br>

n number of carbon atoms of the pseudo-component

N number of carbon atoms of last component in the extended system

NG number of MCN groups of Whitson's lumping scheme

SG<sub>n</sub> specific gravity of the pseudo-component with n carbon atoms

SG<sub>N</sub> specific gravity of last component in the extended system

SG<sub>C7+</sub> specific gravity of the C7+ fraction in the hydrocarbon system

Z<sub>6</sub> mole fraction of C6 component in the hydrocarbon system

Z<sub>n</sub> mole fraction of the pseudo-component with n carbon atoms

Z<sub>N</sub> mole fraction of last component in the extended system

Z<sub>C7+</sub> mole fraction of the C7+ fraction in the hydrocarbon system