Improved Genetic Algorithms for Deterministic Optimization and Optimization under Uncertainty. Part II. Solvent Selection under Uncertainty

W. Xu and U. M. Diwekar*

Center for Uncertain Systems, Tools for Optimization and Management (CUSTOM) Vishwamitra Research Institute, Westmont, Illinois 60559

The existence of a combinatorial search space in molecular design poses a challenge for traditional deterministic optimization methods. This paper presents an innovative approach based on improved genetic algorithms to optimally design solvents or extracting agents to separate the binary mixture of acetic acid and water in the presence of separation and physical property constraints. The UNIFAC−VLE model and Hansen's solubility parameter model are used to estimate the mixture properties. Since the mixture properties are predicted using group contribution methods, in which the group properties are derived from the regression of experimental data, uncertainties are inherent in the predictions. To account for these uncertainties, an "uncertainty factor" is introduced, which is defined as the ratio of experimental value to the value computed by the model. This uncertainty factor then is propagated through the model. A deterministic optimization framework is developed to compare the performance of two algorithms: (1) the improved genetic algorithm and (2) efficient stochastic annealing. Uncertainties are propagated through the stochastic framework. The results of the stochastic framework, with respect to the two algorithms, have also been analyzed and compared. As a stochastic optimization technique, the improved genetic algorithm outperforms its counterpart, the stochastic annealing technique. Using this method, new solvents with better targeted properties are found with less computational time.

1. Introduction

With the advancement in chemical processes and the growing need to substitute existing solvents to adhere to tighter environmental regulations, the development of newer solvents with optimal performance has gained prominence and has initiated considerable research. The traditional trial-and-error method of knowledge-based experimental screening is relatively accurate, but time-consuming, and does not guarantee an optimal solvent. Thus, a guiding procedure is required to augment advantages of knowledge-based experimental screening while avoiding its defects. Computer-aided molecular design (CAMD), which is the reverse use of a group contribution method (GCM), is one such technique. The successful use of GCM in predicting physicochemical properties for both pure compounds and mixtures ensures the reliability and applicability of CAMD, in which functional groups are selected to satisfy chemical feasibility and property constraints. The CAMD problem can be formulated as a mixed integer linear programming (MILP) problem[1,2] or a mixed integer nonlinear programming (MINLP) problem.[3] In this work, the framework has been formulated as an MINLP, because of the existence of nonlinear components inherited from the nonlinear mixture property prediction model such as UNIFAC and the solubility-parameter-based model.

In the application of CAMD, both deterministic and stochastic optimization procedures are used. Deterministic optimization techniques include the branch and bound (BB), outer approximation (OA), and generalized bender's decomposition (GBD). These techniques may face problems in the case of a combinatorial explosion or a discontinuous search space. One strategy to address the aforementioned problems is to linearize nonlinear components and reformulate the original MINLP problem to a MILP problem.[4] Another tactic is to reduce the search space by exploiting the problem structure. A reduced BB method based on this idea has been developed[5] to find globally optimal solvents. Stochastic optimization techniques, which include simulated annealing and genetic algorithm, can handle combinatorial explosion and a discontinuous search space, because they are essentially combinatorial in nature. Simulated annealing and genetic algorithm have been successfully used in refrigerants, heating media, solvents, and polymer design.[6−8]

Solvent properties such as distribution coefficient, selectivity, and solvent loss are the key criteria in determining the solvent performance. One approach to evaluate these properties is the group-contribution-based UNIFAC method, from which infinite dilution activity coefficients are calculated. UNIFAC method was first introduced by Fredenslund et al.[9] There are two sets of UNIFAC parameter tables: liquid−liquid equilibrium (LLE) and vapor−liquid equilibrium (VLE), of which the LLE set has not been revised and extended since the original publication, whereas the VLE set has been revised and extended. VLE is more commonly used in mixture property prediction.[10] Improvements in the UNIFAC model have primarily been focused on the following areas: (1) introduction of new groups, (2) revision and extension of data tables,[11] (3) modification of a combinatorial part and the addition of temperature-dependent parameters in the UNIFAC model.[12,13] An-

* To whom correspondence should be addressed. Tel.: +1 630 515 8773. E-mail: urmila@vri-custom.org.

10.1021/ie049126l CCC: $30.25
Published on Web 08/09/2005

Table 1: Functional Groups in the UNIFAC Group Contribution Method (GCM)

i	N_2^i	i	N_2^i	i	N_2^i	i	N_2^i
1	CH_3	7	CH_2=C	13	CH_3CO	19	CH_3O
2	CH_2	8	CH=C	14	CH_2CO	20	CH_2O
3	CH	9	C=C	15	CHO	21	CH−O−
4	C	10	OH	16	CH_3COO	22	COOH
5	CH_2=CH	11	CH_3OH	17	CH_2COO	23	HCOOH
6	CH=CH	12	H_2O	18	HCOO	24	COO

Table 2: Functional Groups in Hansen's Solubility Parameter Model

i	N_2^i	i	N_2^i	i	N_2^i	i	N_2^i
1	CH_3	6	CH_2=C	11	O	16	NH
2	CH_2	7	CH=CH	12	C=O	17	CN
3	CH	8	CH=C	13	O=CH		
4	C	9	C=C	14	COOH		
5	CH_2=CH	10	OH	15	COO		

other approach is the solubility-parameter-based models, which typically use Hansen's three-dimensional solubility parameters.[14] These parameters are generally used for the solvent selection of polymer binders and also can be incorporated in the Flory−Huggins model to predict solvent activity coefficients at infinite dilution in several acrylate and acetate polymers.[15] Compared to the infinite dilution activity coefficients model for predicting solvent properties, Hansen's solubility-parameter-based model lacks accuracy, although it is simple to implement and is computationally less intensive than the UNIFAC method. Table 1 lists the functional groups of UNIFAC, and Table 2 lists the functional groups of Hansen's solubility parameter model.

Pure component properties such as the "boiling point" are also important in solvent selection, especially for determination of a suitable separation strategy. Three types of estimation methods[16−18] have been developed from the GCMs, in which estimated properties are the sum of multiplication of frequency and contribution of groups. The method proposed by Marrero and Gani[17] uses three sets of groups: first-order, second-order, and third-order. The first-order groups serve as building blocks, the second-order groups account for isomers and proximity effects, and the third-order groups provide structural information about molecular fragments. The method proposed by Marrero-Morejón and Pardillo-Fontdevila[18] use contributions of interactions between bonding groups in the molecule instead of contributions by simple groups. The method proposed by Joback and Reid[16] simplifies the molecular structure but cannot distinguish between isomers, although it provides a quick estimation method suitable for the pre-design. Another important issue is prediction error in the infinite dilution activity coefficients and Hansen's solubility parameters. Infinite dilution activity coefficients are derived from the UNIFAC model, in which all of the group contributions are regressed from the experimental data using different mathematical approaches, with the objective of minimizing the discrepancy between predicted values and experimental values. Thus, the estimation error is unavoidable, and sometimes this relative error can range from 6.6% up to 72.2%.[19] The extent of these errors will nullify the efforts spent on developing more efficient and robust algorithms unless uncertainties are considered in CAMD.

In this paper, two improved genetic algorithms named the efficient genetic algorithm (EGA) and the Hammersley stochastic genetic algorithm (HSGA), developed by applying the Hammersley sequence sampling (HSS) technique[20] and the HSS confidence interval,[21,22] are utilized in the optimal solvent design for the liquid−liquid extraction process for the dilute mixture of acetic acid in water. The chemicals knowledge-based selection operation, which originated from Deb's constraint handling method,[23] has also been implemented to facilitate the evolution process. Three solvent properties are used as criteria of selection: (1) distribution coefficient, (2) selectivity, and (3) solvent loss, which are evaluated separately using an infinite dilution activity coefficient model, as well as Hansen's solubility parameter model. A pure compound physical property—the normal boiling point—is evaluated by Joback's boiling point estimation method. Section 2 of this paper presents the problem formulation for solvent selection in liquid−liquid extraction of the acetic acid−water system. Section 3 introduces EGA, its application to the deterministic CAMD, and the comparison between EGA and efficient simulated annealing (ESA). Section 4 applies the Hammersley stochastic genetic algorithm (HSGA) to the CAMD under uncertainty and compares results to its counterpart, Hammersley stochastic simulated annealing (HTSA). The final section is the conclusion.

2. Problem Formalization for Solvent Selection in Liquid−Liquid Extraction of the Acetic Acid−Water System

Acetic acid is a commonly used, valuable in-process solvent, as well as a pollutant, if released to the environment. Therefore, it is desirable to recycle acetic acid from its aqueous solutions. Direct separation can be used for this purpose, but it is avoided for its high energy consumption. Liquid−liquid extraction, as a substitute, is commonly used by adding another solvent to extract acetic acid from water to decrease the amount of water to be taken overhead in the distillation. The most important properties of the added agents are the solute distribution coefficient (m), solvent loss (S_L), selectivity (β), and boiling point (T_{BP}). In this formulation, the variable m, which is the key factor in determining extractor size and recycled solvent amount, was selected as the main objective. Other properties are treated as constraints. The general formulation is presented below.

$$\begin{aligned} &\max_{N_1,N_2^i} && m \\ &\text{s.t.} && \beta \geq \beta_{\min} \\ & && S_L \leq S_{L,\max} \\ & && T_{bp,\min} \leq T_{bp} \leq T_{bp,\max} \\ & && N_{1,\min} \leq N_1 \leq N_{1,\max} \\ & && N_{2,\min}^i \leq N_2^i \leq N_{2,\max}^i \\ & && \text{chemical is feasible} \\ &N_1,N_2^i && \in \text{integer.} \end{aligned} \tag{1}$$

where N_1 is the number of functional groups in a molecule, and N_2^i is the index of total available functional groups (where $i = 1, ..., N_1$). The lower and upper bounds of other constraints are generally taken from available literature resources for specific problems.

The CAMD formulation for the acetic acid−water system is taken from Kim's formulation.[19] However, to keep the uniformity of the cross-over operation for two molecules with different length, the length of a molecule is fixed to the largest allowable number of functional groups and the lower-bound index is relaxed to zero instead of one in the original form.

3. EGA and Deterministic CAMD

In the first part of these series of papers, the efficient genetic algorithm (EGA), the efficient stochastic genetic algorithm (ESGA), and the Hammersley stochastic genetic algorithm (HSGA) are presented. In this section, EGA is used to obtain optimal solvents for the acetic acid−water separation.

3.1. Genetic Operators and Molecular Representation. The operators used in EGA are tournament selection, uniform crossover, and jump mutation. In the application to CAMD, the genetic operators not only affect the genetic algorithm search performance, but also the molecular representation. The two design variables identified for the optimization process are (1) the number of functional groups in a molecular structure (N_1) and (2) the functional group index (N_2^i), which represents the molecular configurations whose length is determined by N_1. The change in N_2^i induces the change of solution string length. Although this length change has no effect on the selection and mutation operations, it disturbs uniform crossover, which involves exchange of two solution strings bit by bit with cross-over probability. When uniform crossover is applied to two strings with different lengths, part of the longer string is left without a change, which restricts the exploration in the genetic algorithm search. To counter this drawback, a strategy is developed, in which N_1 is transformed to an implicit design variable; in other words, we do not treat N_1 as a design variable anymore. The length of a molecular structure is pre-assigned, which is the maximal number of allowable building groups, $N_{1,\max}$. At the same time, the lower bound of N_2^i is also relaxed to zero, which means that some slots can be empty. After this transformation, we can simultaneously keep the uniform cross-over property and satisfy the constraints on N_1. The molecular structure is ultimately represented as a vector with $N_{1,\max}$ positions. Each position holds an index of the functional group ranges from zero to $N_{1,\max}$. For example, acetic acid can be expressed in the following format using the groups in Table 1, where $N_{1,\max}$ is assumed to be 10.

Acetic acid: (1, 22, 0,0,0,0,0,0,0,0)

3.2. Constraint Handling. The penalty function method is widely used in regard to constrained optimization. The most difficult part of applying penalty functions is to determine the penalty parameter that guides the search to the constrained optimum. In genetic algorithm population size, the cross-over probability and mutation probability must be adjusted to obtain the best performance. The introduction of another parameter that has considerable impact on the genetic algorithm performance would make it less flexible and would also increase the difficulty of finding the optimal parameter set. Another issue that needs to be addressed is that, in CAMD applications, there are two types of constraints: (1) VIOL_1, which are the solvent loss and selectivity constraints (where the violations can be quantified), and (2) VIOL_2, which are the chemical feasibility constraints (where the violations cannot be quantified). The fact that the violations cannot be quantified is case of VIOL_2 constraints, poses another challenge to the application of the penalty function method. Considering these difficulties, an approach based on ranking using pairwise comparison in tournament selection was developed from the constraint handling method.[23] The salient points of the algorithm are as follows: (a) feasible solutions have the priority over infeasible solutions; (b) between two feasible solutions, the one that has the better objective function value is preferred; and (c) between two infeasible solutions, three conditions are considered:

(i) If both violations are from VIOL_1, then the solution with the smaller constraint violation is preferred.

(ii) If one violation is from VIOL_1, whereas another violation is from VIOL_2, then the solution with violation from VIOL_1 is preferred.

(iii) If both violations are from VIOL_2, then the solution with the smaller VIOL_1 constraint violation is preferred.

3.3. Deterministic CAMD Based on the Infinite Dilution Activity Coefficient Model. The distribution coefficient (m), which is a measure of solvent capacity, represents the solute distribution between the solvent and the raffinate phases. A high value of m reduces the size of extraction equipment and the amount of recycling solvent. Solvent selectivity (β) is the ratio between the distribution coefficients of solute and raffinate. It estimates the ability of the solvent to selectively dissolve a solute. A high solvent selectivity value thus can reduce the cost of solute recovery. Solvent loss (S_L) is a measure of the concentration of solvent in raffinate phase. A low solvent loss value means high selectivity toward solute and determines the immiscibility between the solvent and the raffinate. The mathematical definitions of the aforementioned three terms derived from infinite dilution activity coefficients are defined in eqs 2, 3, and 4.

$$m = \frac{\text{concentration of solute in extractive phase}}{\text{concentration of solute in raffinate phase}} = \frac{x_{B,S}}{x_{B,A}}\frac{MW_A}{MW_S} \cong \frac{\gamma^{\infty}_{B,A}}{\gamma^{\infty}_{B,S}}\frac{MW_A}{MW_S} \quad (2)$$

$$\beta = \frac{\text{distribution coefficient of solute}}{\text{distribution coefficient of solvent}} \cong \frac{m_B}{m_A} \cong \frac{\gamma^{\infty}_{A,S}}{\gamma^{\infty}_{B,S}}\frac{MW_B}{MW_A} \quad (3)$$

$$S_L = \text{concentration of solvent in raffinate phase} \cong \frac{1}{\gamma^{\infty}_{S,A}}\frac{MW_S}{MW_A} \quad (4)$$

where the symbols A, B, and S represent the raffinate, solute, and solvent phases. MW denotes the molecular weight, and x represents the mole fraction. The infinite dilution activity coefficients are calculated by the fifth-revised original UNIFAC model,[11] in which the original UNIFAC model is used with new groups and a revised

data table. The UNIFAC model is shown as depicted below.

$$\ln \gamma_i = \ln \gamma_i^{\mathrm{C}} + \ln \gamma_i^{\mathrm{R}} \tag{5}$$

$$\ln \gamma_i^{\mathrm{C}} = \ln\left(\frac{\Phi_i}{x_i}\right) + \frac{z}{2} q_i \ln\left(\frac{\theta_i}{\Phi_i}\right) + l_i - \frac{\Phi_i}{x_i}\sum_j x_j l_j \tag{6}$$

$$\ln \gamma_i^{\mathrm{R}} = \sum_k \nu_k^i (\ln \Gamma_k - \ln \Gamma_k^i) \tag{7}$$

$$\ln \Gamma_k = Q_k\left[1 - \ln\left(\sum_m \theta_m \psi_{mk}\right) - \sum_m \frac{\theta_m \psi_{km}}{\sum_n \theta_n \psi_{nm}}\right] \tag{8}$$

$$\theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n}, \quad \psi_{mn} = \exp\left(-\frac{a_{mn}}{T}\right),$$

$$\theta_i = \frac{q_i x_i}{\sum_j q_j x_j}, \quad \Phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \tag{9}$$

In these equations, x_i is the mole fraction of component i, θ_i is the area fraction, Φ_i is the segment fraction, and r_i and q_i are measures of molecular van der Waals volumes and molecular surface area. θ_m is the area fraction of group m, X_m the mole fraction of group m in the mixture, and a_{mn} the group interaction parameter. The combinatorial part, $\ln \gamma_i^{\mathrm{C}}$, is dependent on the sizes and shapes of the molecules, whereas the residual part, $\ln \gamma_i^{\mathrm{R}}$, is dependent on group areas and group interactions.

To estimate normal boiling point (T_{BP}), the following linear prediction model[16] is used:

$$T_{\mathrm{BP}}\,(^\circ\mathrm{C}) = \sum_{i=1}^{N_1} t_a(N_2^{(i)}) + t_b \tag{10}$$

The octet rule is used for acyclic groups in this paper to determine whether a chemical is feasible.

$$\sum_i^{N_1} b_i = 2(N_1 - 1) \tag{11}$$

Here b_i is the number of free attachments in a group index i. Both high-boiling-point and low-boiling-point solvents are included in our configuration. All the constraints are listed in Table 3. The optimal result and other local optima computed in the search for high-boiling-point solvents are listed in Table 4. Table 5 shows the candidate solvents found by solving the same configurations using ESA.[19] From Table 4, we can see the best solvent found by EGA with the configuration ($4CH_2$, CH, CH_2=CH, CH_3CO, HCOO) has values of $m = 1.236$ and $\beta = 26.80$, which exceeds the optimal solvent (CH_3, $6CH_2$, OH) found by ESA (with $m = 0.6074$ and $\beta = 19.77$). It is also worth mentioning that EGA has also found a set of local optima with better distribution coefficients than the optimal set found by ESA. There are two reasons for this phenomenon: (1) the genetic algorithm starts from a set of initial points instead of one as simulated annealing, and this property decreases the dependence on the initial guess, giving a better chance to find the global optimum; and (2) a different strategy is used in constraint handling. In simulated annealing, an infeasible solution is accepted or discarded randomly according to the Metropolis criterion. The disadvantage of this constraint violation handling strategy is that there is no standard for accepting or discarding solutions. Therefore, some infeasible solutions that have more chances to reach the optimum may be discarded, and, consequently, some good patterns are lost in the process. At each step, the set of different configurations found by EGA makes the comparison between infeasible solutions possible and decreases the randomness of acceptance. Through this method, some infeasible solutions that have a better chance to attain optimality would be preserved. These two properties of EGA improve the efficiency of the search process and make it less susceptible to being trapped in local optima. Table 6 lists the low-boiling-point candidate solvents found by EGA, and Table 7 lists the low-boiling-point candidate solvents found by ESA. From the comparison of these two tables, we can see that both EGA and ESA find the same optimal solvent configuration ($2CH_3$, CH_2, CH, HCOO) with $m = 0.8658$ and $\beta = 24.11$. Another observation is that EGA does not find as many top potential solvents as ESA. This can be explained by the fact that it does not traverse as many local optima as ESA, which gives EGA a quicker convergence and a lesser probability of getting trapped in local optima. In addition, the driving force in the case of EGA is larger than ESA, because it combines good patterns from different populations leading to better solutions. To increase diversity, one strategy that can be used is increasing the population size of EGA to cover more search space. The major limitation in this case is when the population size becomes large enough to cover the discrete search space, thereby becoming a generation-and-test problem. Consequently, all the feasible solutions would be found, but the computational burden would be considerable, because of the explosive number of combinations. Another strategy is to employ sharing, to avoid having most of the population cluster in one subdomain as evolution proceeds. The basic idea of sharing is to penalize clustered solutions, and, consequently, the less-fit populations in other less-clustered subdomains would also have a chance to be selected and contribute to the next generation.[24] By applying these two strategies, we would expect that more-feasible solutions are traversed before reaching the optimum. Table 8 lists the potential solvents found with increased population size by EGA. Comparing Table 8 to Table 7,

Table 3. Constraints of Both High-Boiling-Point and Low-Boiling-Point Solvents for the Acetic Acid−Water System, Using UNIFAC

parameter	bounds
High-Boiling-Point Solvent Constraints	
number of groups in a molecule, N_1	2−10
$\beta_{\min}$	7
$S_{\mathrm{L,max}}$	0.01
boiling point temperature, T_{BP} (°C)	148−268
N_2	0−24
Low-Boiling-Point Solvent Constraints	
number of groups in a molecule, N_1	2−10
$\beta_{\min}$	7
$S_{\mathrm{L,max}}$	0.058
boiling point temperature, T_{BP} (°C)	47−108
N_2	0−24

Table 4. Optimal High-Boiling-Point Solvents and Other Top Feasible Solvents Found by Efficient Genetic Algorithm (EGA)

index	solvents	chemical name	m	β
1	**$4CH_2$, CH, CH_2=CH, CH_3CO, HCOO**	**formic acid 1-acetyl-hept-6-enyl ester**	**1.23**	**26.8**
2	$2CH_3$, $2CH_2$, 2CH, CH_3COO, HCOO	1-isopropenyloxy-3-methyl-4-vinyloxy-penetane	1.12	47.7
3	$6CH_2$, CH_3COO, HCOO	1-isopropenyloxy-6-vinyloxy-hexane	1.12	47.6
4	$2CH_3$, $3CH_2$, CH, CH_2COO, HCOO	formyloxy-acetic acid 1-methyl-pentyl ester	0.97	47.1
5	$2CH_3$, CH_2, CH, CH=C, CH_3COO, HCOO	5-isopropenyloxy-4-methyl-2-vinyloxy-pent-2-ene	0.96	36.1
6	CH_3, $3CH_2$, CH=C, CH_3COO, HCOO	1-isopropenyloxy-2-methyl-5-vinyloxy-pent-1-ene	0.96	36.05
7	CH_3, $2CH_2$, CH, CH=CH, CH_3COO, HCOO	5-isopropenyloxy-3-methyl-1-vinyloxy-pent-1-ene	0.93	35.18
8	CH_3, $2CH_2$, CH, CH_2=C, CH_3COO, HCOO	5-isopropenyloxy-2-methyl-3-vinyloxy-pent-1-ene	0.9	32.5
9	$4CH_2$, CH_2=C, CH_3COO, HCOO	1-isopropenyloxy-3-methylene-5-vinyloxy-pentane	0.89	32.4
10	CH_3, CH_2, CH, CH_2=CH, CH_2COO, HCOO	3-ethyl-pent-4-enoic acid methyl ester	0.88	30.63

Table 5. Top 10 High-Boiling-Point Candidate Solvents Found by Efficient Stochastic Annealing (ESA)

index	solvents	chemical name	m	β
1	**CH_3, $6CH_2$, OH**	**hexan-1-ol**	**0.6074**	**19.77**
2	$2CH_3$, $4CH_2$, CH, OH	heptan-2-ol	0.6073	19.78
3	CH_3, $7CH_2$, OH	octan-1-ol	0.5478	21.73
4	$2CH_3$, $5CH_2$, CH, OH	octan-2-ol	0.5477	21.74
5	$2CH_3$, $4CH_2$, CH, HCOO	2-vinyloxy-heptane	0.5311	30.8
6	CH_3, $8CH_2$, OH	nonan-1-ol	0.4998	23.66
7	$2CH_3$, $6CH_2$, CH, OH	nonan-2-ol	0.4998	23.68
8	$2CH_3$, $5CH_2$, CH, OH	6-methyl-heptan-1-ol	0.4698	32.77
9	$2CH_3$, $7CH_2$, CH, OH	2-methyl-nonan-1-ol	0.4603	25.6
10	$2CH_3$, $6CH_2$, CH, HCOO	2-vinyloxy-nonane	0.4194	34.65

Table 6. Optimal Low-Boiling-Point Solvents and Other Top Feasible Solvents Found by EGA

index	solvents	chemical name	m	β
1	**$2CH_3$, CH_2, CH, HCOO**	**formic acid *sec*-butyl ester**	**0.8658**	**24.11**
2	CH_3, $3CH_2$, HCOO	formic acid butyl ester	0.8656	24.08
3	$2CH_3$, CH_2, CH_2CO	pentan-2-one	0.5083	41.67
4	$2CH_3$, CH_2, CH, CH_2=C, CH_3O	4-methoxy-2-methyl-pent-1-ene	0.3305	7.1802
5	CH_3, CH=CH, CH_3COO	2-methyl-hexa-1,4-diene	0.2995	13.46

Table 7. Top 10 Low-Boiling-Point Candidate Solvents Found by ESA

index	solvents	chemical name	m	β
1	**$2CH_3$, CH_2, CH, HCOO**	**formic acid *sec*-butyl ester**	**0.87**	**24.1**
2	CH_3, $3CH_2$, HCOO	formic acid butyl ester	0.87	24.1
3	CH_3, CH_2, CH=CH, HCOO	formic acid but-1-enyl ester	0.76	16.7
4	CH_3, CH_2, CH_2=C, HCOO	(1-methylene-propoxy)-acetaldehyde	0.75	15.2
5	$2CH_2$, CH_2=CH, HCOO	formic acid but-3-enyl ester	0.72	15
6	CH_3, CH,CH_2=CH, HCOO	3-vinyloxy-but-1-ene	0.72	15
7	CH_3, CH_2=CH, CH_3O, CH−O	3-methylperoxy-but-1-ene	0.66	7.6
8	CH_3, CH, CH_2=CH, CH_2O, CH_3O	3-methyl-4-methylperoxy-but-1-ene	0.63	7.21
9	$2CH_2$, CH_2=CH, CH_2O, CH_3O	5-methylperoxy-pent-1-ene	0.63	7.21
10	CH_3, CH_2=C, CH_3CO	3-methyl-but-3-en-2-one	0.61	14.6

Table 8. Optimal Low-Boiling-Point Solvents and Other Top Feasible Solvents Found by EGA with Increased Population Size

index	solvents	chemical name	m	β
1	**$2CH_3$, CH_2, CH, HCOO**	**formic acid *sec*-butyl ester**	**0.8658**	**24.11**
2	CH_3, $3CH_2$, HCOO	formic acid butyl ester	0.8656	24.08
3	CH_3, CH_2, CH=CH, HCOO	formic acid but-1-enyl ester	0.7632	16.65
4	CH_3, CH_2, CH_2=C, HCOO	(1-methylene-propoxy)-acetaldehyde	0.7481	15.23
5	$2CH_2$, CH_2=CH, HCOO	formic acid but-3-enyl ester	0.7194	14.97
6	CH_3, CH,CH_2=CH, HCOO	3-vinyloxy-but-1-ene	0.7191	14.99
7	CH_3, CH_2=CH, $2CH_2O$	3-ethylperoxy-propene	0.6659	7.5832
8	CH_3, CH_2=C, CH_3CO	3-methyl-but-3-en-2-one	0.6123	14.59
9	CH_3, CH=CH, CH_3CO	2-methyl-penta-1,3-diene	0.6103	16.1
10	CH_2, CH_2=CH, CH_3CO	Pent-4-en-2-one	0.6025	14.09

one can see that EGA finds a solvent configuration (CH_3, CH_2=CH, $2CH_2O$) with $m = 0.6659$ and $\beta = 7.5831$, but it does not find a configuration (CH_3, CH_2=CH, CH_3O, CH−O) shown in the ESA candidate solvents pool with $m = 0.66$ and $\beta = 7.60$. This shows that both EGA and ESA cannot find all of the local optima. As mentioned in the generation-and-test[25] approach, the candidate solvent pool is used to decrease the effect of estimation uncertainties. However, in the optimization approach, this strategy is questionable, because it is natural that the search does not traverse all the local optima.

3.4. Deterministic CAMD Based on Hansen's Solubility Parameter Model. Hansen's three-dimensional solubility parameter model is another important method that is widely used in solvent selection. It can be expressed in the following equation:

$$\delta\ (\mathrm{MPa}^{1/2}) = \sqrt{{\delta_d}^2 + {\delta_p}^2 + {\delta_h}^2} \qquad (12)$$

Here δ_d is a dispersive term, δ_p is a polar term, and δ_h is a hydrogen-bonding term. The solubility parameter and its three terms can be determined by a semi-

Table 9. Optimal Low-Boiling-Point Solvents and Other Top Feasible Solvents Found by EGA

index	solvents	chemical name	m	β
1	**CH_3, CH_2, OH**	**ethanol**	**17.22**	**190.6**
2	CH_3, CH=CH, OH	propen-1-ol	12.62	175.9
3	CH_3, $2CH_2$, OH	propan-1-ol	9.5	144.4
4	CH_2=CH, OH	ethenol	9.34	105.87
5	$2CH_3$, CH, OH	propan-2-ol	6.06	99.04
6	CH_2, CH_2=CH, OH	prop-2-en-1-ol	5.82	91.39
7	CH_3, CH_2=C, OH	propen-2-ol	5.19	85.33
8	CH_3, CH_2, CH_2=C, OH	2-methyl-prop-2-en-1-ol	3.4	72.84
9	$2CH_3$, C=O	acetone	2.05	42.3
10	CH_3, $2CH_2$, O=CH	butanal	1.83	45.36

Table 10. Top 10 Low-Boiling-Point Candidate Solvents Found by ESA

index	solvents	chemical name	m	β
1	**CH_3, CH_2, OH**	**ethanol**	**17.2**	**190.6**
2	CH_3, CH=CH, OH	propen-1-ol	12.6	175.9
3	CH_3, $2CH_2$, OH	propan-1-ol	9.5	144.4
4	CH_2=CH, OH	ethenol	9.34	105.9
5	$2CH_3$, CH, OH	propan-2-ol	6.06	99
6	$2CH_2$, CH_2=CH, OH	3-buten-1-ol	5.83	91.4
7	CH_3, CH_2=C, OH	propen-2-ol	5.2	85.3
8	CH_2=CH, $2CH_2$, OH	3-buten-1-ol	3.86	72.8
9	CH_3, CH_2=C, CH_2, OH	2-methyl-prop-2-en-1-ol	3.41	60.7
10	CH_2, CH_2=CH, CH, OH	prop-2-en-1-ol	2.78	47.7

empirical method for most common liquids.[26] The solute distribution coefficient (m), solvent loss (S_L), and selectivity (β) using Hansen's three-dimensional solubility parameters can be expressed in the following equations:

$$m \propto \left(\frac{r_{BA}}{r_{BS}}\right)^2 \frac{MW_A}{MW_S} \tag{13}$$

$$\beta \propto \left(\frac{r_{AS}}{r_{BS}}\right)^2 \frac{MW_B}{MW_A} \tag{14}$$

$$S_L \propto \left(\frac{1}{r_{SA}}\right)^2 \frac{MW_S}{MW_A} \tag{15}$$

Here, the subscripts A, B, and S represent the raffinate, solute and solvent, respectively. r_{ij} is the Euclidean distance metric between two molecules i and j in the three-dimensional space, as shown in the following equation:

$$r_{ij} = [(\delta_d^i - \delta_d^j)^2 + (\delta_p^i - \delta_p^j)^2 + (\delta_h^i - \delta_h^j)^2]^{1/2} \tag{16}$$

A linear group contribution model is developed[27] for the three solubility parameter terms, using the least-squares method from the literature data.[26] There are a total of 17 functional groups that are listed in Table 2. The solubility parameters are estimated by linearly adding group properties and intercept values.

$$\delta_{d,p,h}\ (\mathrm{MPa}^{1/2}) = \sum n_i \delta_{(d,p,h)i} + a \tag{17}$$

Here, the subscripts d, p, and h denote dispersive, polar, and hydrogen-bonding terms, respectively. The term n_i is the occurrence frequency of group i, and a is the intercept term that is constant. One thing that must be mentioned is that the groups in Table 2 are designed for linear or branched hydrocarbons and not for the aromatic, cyclic, and/or halogenated compounds, because of environmental concerns. In the Hansen's solubility parameter model, only low-boiling-point solvents are considered and the candidate solvents found in EGA and ESA are listed in Tables 9 and 10, respectively. Both EGA and ESA find the same optimal molecular configuration (CH_3, CH_2, OH) with $m = 17.2$ and $\beta = 190.6$. EGA, as expected, does not find as many near-optimal solutions as ESA. Another observation is that most candidate solvents are alcohols, which can be explained by the fact that the alcohol functional groups have large hydrogen-bonding and polar terms, which result in the increase of the distribution coefficient m.

4. Implementation of the HSGA to Stochastic CAMD

4.1. Uncertainty Analysis for Original UNIFAC Model and Its Implementation. The original UNIFAC model used in this paper has three input parameters: the surface area (R_K), volume (Q_K), and interaction parameter between groups m and n (a_{mn}). The surface area and volume are calculated from atomic and molecular structural data, so it is assumed that there are no estimation errors in them. Interaction parameters are regressed from experimental data and, thus, are subject to uncertainty, because of experimental and regression errors. The omission of such prediction errors would lower the credibility of optimal design results. Therefore, to analyze the effect of uncertainties on design, an "uncertainty factor" term is introduced, which is defined as the ratio of the experimental value to the calculated value. In UNIFAC, because of the large size of the interaction parameters and their utilization in the lower level, the uncertainties of predicted infinite dilution activity coefficients (γ^∞) are considered directly instead of uncertainty factors for interaction parameters (a_{mn}). Because water has very different properties from those of organic chemicals, γ^∞ is divided into three categories, based on the type of family: organic/water, water/organic, and organic/organic families. The uncertainty factor of γ^∞ is defined as

$$UF = \frac{\gamma^\infty_{exp}}{\gamma^\infty_{cal}} \tag{18}$$

The ideal case is when $UF = 1$, which means there are no estimation errors. The uncertainty factors of three types of γ^∞ have been identified and quantified by Kim.[19]

The *UF* of the organic/water family has a log-normal distribution, with an arithmetic mean of 2.92 and a standard deviation of 5.94. The *UF* of water/organic has a normal distribution, with a mean of 1.08 and a standard deviation of 0.37. The *UF* of organic/organic has a log-normal distribution, with a mean of 1.42, and a standard deviation of 1.14. Based on the three distributions, one can see that the UNIFAC prediction model has a tendency to underestimate the γ^{∞} values and the effect on the three categories is different. The uncertainty factors in the distribution coefficient are both log-normal, and the effect of uncertainty on the organic/water family is larger than the effect on the organic/organic family; thus, one would expect the stochastic distribution coefficient to be larger, when compared to the deterministic case, as well as closer to the real value. After applying these uncertainties, the stochastic solvent selection model is formulated for the low-boiling solvents selection as follows.

$$\max \frac{1}{N_{\text{samp}}} \sum_{j=1}^{N_{\text{samp}}} \left[\frac{\xi_1^j \gamma_{\text{B,A}}^{\infty}}{\xi_3^j \gamma_{\text{B,S}}^{\infty}} \right] \frac{MW_{\text{A}}}{MW_{\text{S}}}$$

s.t.

ξ_1: $\log N(2.92, 5.94)$ for the organic/water family

ξ_2: $N(1.08, 0.37)$ for the water/organic family

ξ_3: $\log N(1.42, 1.14)$ for the organic/organic family

$$\beta = \frac{1}{N_{\text{samp}}} \sum_{j=1}^{N_{\text{samp}}} \left[\frac{\xi_2^j \gamma_{\text{A,S}}^{\infty}}{\xi_3^j \gamma_{\text{B,S}}^{\infty}} \right] \frac{MW_{\text{B}}}{MW_{\text{A}}} \geq 7$$

$$S_{\text{L}} = \frac{1}{N_{\text{samp}}} \sum_{j=1}^{N_{\text{samp}}} \left[\frac{1}{\xi_1^j \gamma_{\text{S,A}}^{\infty}} \right] \frac{MW_{\text{S}}}{MW_{\text{A}}} \leq 0.058$$

$$47\ ^{\circ}\text{C} \leq T_{\text{BP}} \leq 108\ ^{\circ}\text{C}$$

$$2 \leq N_1 \leq 10; \quad 1 \leq N_2^i \leq 24, \forall\, i \in N_1; \quad \text{chemical is feasible} \qquad (19)$$

Here, A represents water, B represents acetic acid, and C represents the solvent. ξ_i is an uncertain parameter of UF_i and is imposed on γ^{∞}. N_1 and N_2^i are the discrete variables. This optimization problem under uncertainty is solved using the HSGA. Similar to the HSTA,[19] the number of samples with uncertain variables is determined in an outer optimization loop and the information about the distribution functions, expressed as 0.1% and 99.9% quantiles, is supplied to an inner sampling loop. The sampling technique used here is Hammersley sequence sampling (HSS). Both HSGA and HSTA find the same optimal configuration ($2CH_3$, CH_2, CH, HCOO), but the value of distribution coefficients is different; in HSGA, $m = 2.3405$, whereas, in HTSA, $m = 2.95$. From the stochastic model (eq 19), one can see that the number of samples has an impact on the distribution coefficient. The greater the number of samples, the closer the expected distribution coefficient is to the real value. In both HSGA and HTSA, the number of samples is determined in the outer optimization loop. Initially, the number of samples is small, and as optimization proceeds, the number of samples increases to get a more accurate effect of the uncertainties. Because HSGA and HTSA have different strategies to find the optimal solution, the convergence path is different, and, thus, the final number of samples is also different. A limitation of HSGA is that sometimes the optimum solution is found within several generations and, consequently, the number of samples taken becomes very small, resulting in an error in the calculation of the objective function. Therefore, we use a higher value as the initial number of samples, i.e., we consider 50 samples instead of 5 in HTSA, without sacrificing much efficiency in HSGA.

Table 11. Constraints for Solvent Selection under Uncertainty Based on Hansen's Solubility Parameters

parameter	bounds
number of groups in a molecule, N_1	2−12
β_{min}	17.43
upper bound of S_{L}, $S_{\text{L,max}}$	0.0045
boiling point temperature, T_{BP} (°C)	47−108
N_2	0−17

4.2. Uncertainty Analysis for Hansen's Solubility Parameter Model. There are 17 groups (Table 2) in Hansen's solubility parameter model, and each group has three solubility parameter terms, which are regressed from experimental data. Thus, there are a total of 51 uncertainty distributions, and individual representation of each distribution would be time-consuming and computationally expensive. Another factor that must be taken into account is that, for some chemicals (such as di-isobutyl ketone), each solubility parameter term has a large discrepancy, but the resulting total solubility parameter is closer to the value reported in the literature. Based on the above two considerations, such as the solvent selection model based on γ^{∞}, the uncertainties of the Hansen's solubility parameters (δ_{d}, δ_{p}, and δ_{h}) of the solute/solvent systems are characterized and quantified in terms of the uncertainty factor *UF*. The uncertainty factor defined here is as follows:

$$UF = \frac{\delta_{\text{lit}}}{\delta_{\text{cal}}} \qquad (20)$$

The solubility parameters of 66 noncyclic and nonaromatic compounds are collected from the literature, and the uncertainty is characterized using *UF* and is quantified in terms of probability distributions. The *UF* of the dispersive term (δ_{d}) is normally distributed with a mean of 1.05 and a standard deviation of 0.08. The *UF* of the polar term (δ_{p}) is normally distributed with a mean of 1.21 and a standard deviation of 0.85. The *UF* of the hydrogen-bonding term (δ_{h}) is normally distributed with a mean of 1.28 and a standard deviation of 0.96.[19] The constraints for solvent selection under uncertainty based on Hansen's solubility parameters are listed in Table 11. The optimal and top feasible solvents, as determined by HSGA, are listed in Table 12. The top 11 solvents, as determined by HTSA, are listed in Table 13. The optimal solvent determined in HSGA has the configuration (CH_3, CH_2, OH), with $m = 3.7313$ and $\beta = 36.82$, which is the same as the deterministic case. The best solvent determined by HTSA has the configuration ($3CH_3$, CH, COO) and has much larger m and β values, which are 161 and 835.1, respectively. However, the same solvent configuration is also found in the feasible solvent pool of HSGA in Table 12, with much smaller m and β values (0.799 and 27.17, respectively). As mentioned previously, the objective values are de-

Table 12. Optimal Low-Boiling-Point Solvents and Other Top Feasible Solvents Found Using the Hammersley Stochastic Genetic Algorithm (HSGA)

index	solvents	chemical name	m	β
1	**CH_3, CH_2, OH**	**ethanol**	**3.7313**	**36.82**
2	CH_3, $2CH_2$, OH	propan-1-ol	2.7799	36.76
3	$2CH_3$, CO	acetone	2.7493	39.56
4	CH_2=CH, OH	ethenol	2.6873	27.1
5	$2CH_3$, CH, OH	propan-2-ol	2.5517	34.77
6	CH_3, CH=CH, OH	propen-1-ol	2.3725	30.18
7	CH_3, $2CH_2$, O=CH	butanal	2.2701	39.93
8	CH_3, CH_2=C, OH	propen-2-ol	2.201	29.7
9	$3CH_3$, CH, COO	acetic acid isopropyl ester	0.799	27.17
10	$2CH_3$, CH_2=C, CO	3-methyl-but-3-en-2-one	0.753	22.96
11	CH_3, CH_2, CH_2=CH, COO	but-3-enoic acid methyl ester	0.737	25.23

Table 13. Top 10 Low-Boiling-Point Candidate Solvents Found Using Hammersley Stochastic Simulated Annealing Algorithm (HTSA)

index	solvents	chemical name	m	β
1	**$3CH_3$, CH, COO**	**acetic acid isopropyl ester**	**161**	**835.1**
2	$2CH_3$, CH_2=C, CO	3-buten-2-one	20.1	78.3
3	$2CH_3$, CH, OH	isopropyl alcohol	5.95	22.2
4	CH_3, CH_2, CH_2=CH, COO	1-penten-3-one	4.91	30.9
5	$2CH_3$, COO	methyl ester	4.26	18.2
6	CH_3, $2CH_2$, OH	propan-1-ol	3.9	15.5
7	CH_3, CH_2, OH	ethanol	3.53	11.4
8	$2CH_3$, CO	acetone	3.51	11.4
9	$2CH_2$, CH_2=CH, OH	3-buten-1-ol	3.38	12.9
10	CH_2, CH_2=CH, OH	prop-2-en-1-ol	3.34	12.2

pendent on the number of samples used. In HTSA, the starting number of samples is 5. For the case where the best molecular structure is found very early in the optimization process, a small number of samples is used. Consequently, this could result in an abnormal value, which would be preserved throughout the optimization process if that same configuration is not found again and the simulated annealing algorithm does not converge until reaching the freeze temperature. Hence, the configuration found by HTSA would not be optimal. In HSGA, we use two approaches to avoid this situation. One approach is to use a larger initial number of samples. Another approach is to use an elitism strategy. At each generation, the best configuration found in the previous generation would be kept in the current generation and re-evaluated with the new samples. The number of samples is increased as the optimization process continues, and the property values (such as distribution coefficient) would be rectified at each generation with the new uncertainty sample number. With this elitism strategy, the property values of the previous set of configurations will be rectified in the later generation and the deviation from the real value would be eliminated.

5. Conclusion

In this paper, the efficient genetic algorithm (EGA) and the Hammersley stochastic genetic algorithm (HSGA) are employed for deterministic and stochastic computer-aided solvent design problems. The optimal results are compared with the results of a simulated annealing algorithm. In the deterministic case, EGA finds a better solvent with larger distribution coefficient and solvent selectivity. These improvements result from employing the multiple points search technique and incorporating infeasible solutions based on ranking. In the stochastic case, HSGA uses two approaches to compute more-accurate estimates of property values. The first approach is to use a higher number of initial random samples. Because of HSGA's fast convergence property, the increasing number of random samples does not affect the efficiency but avoids the drawbacks resulting from small sample size. Elistism strategy is used in the other approach. In this strategy, the best population is passed on directly to the next generation and re-evaluated with new random samples as one gets closer and closer to an optimum solution. In the stochastic UNIFAC model, both HSGA and HTSA find the same optimal configuration but the property values are different. These differences are mainly due to the different sample sizes induced by the different convergence paths of HSGA and HTSA. In the stochastic Hansen's solubility parameter model, HSGA finds the different optimal solvent configuration, compared to the configuration found by HTSA, and the optimal configuration in HTSA has a much lower distribution coefficient and solvent selectivity than that computed by HSGA.

Literature Cited

(1) Jaksland, C. A.; Gani, R.; Lien, K. M. Separation process design and synthesis based on the thermodynamic insights. *Chem. Eng. Sci.* **1995**, *50*, 511−530.

(2) Sahinidis, N. V.; Tawarmalani, M.; Yu, M. Design of alternative refrigerants via global optimization. *AIChE J.* **2003**, *49*, 1761−1775.

(3) Duvedi, A. P.; Achenie, L. E. K. Designing environmentally safe refrigerants using mathematical programming. *Chem. Eng. Sci.* **1996**, *51*, 3727−3739.

(4) Maranas, C. Optimal computer-aided molecular design: A polymer design case study. *Ind. Eng. Chem. Res.* **1996**, *35*, 3403−3414.

(5) Sinaha, M.; Achenie, L. E. K.; Ostrovsky, G. M. Environmentally benign solvent design by global optimization. *Comput. Chem. Eng.* **1999**, *23*, 1381−1394.

(6) Venkatasubramanian, V.; Chan, K.; Caruthers, J. M. Computer-aided molecular design using genetic algorithms. *Comput. Chem. Eng.* **1994**, *18*, 833−844.

(7) Marcoulaki, E. C.; Kokossis, A. C. On the development of novel chemicals using a systemic synthesis approach. Part II. Solvent design. *Chem. Eng. Sci.* **2000**, *55*, 2547−2561.

(8) Kim, K. J.; Diwekar, U. M. Efficient combinatorial optimization under uncertainty. 2. Application to stochastic solvent selection. *Ind. Eng. Chem. Res.* **2002**, *41*, 1285−1296.

(9) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, 1086−1099.

(10) Pretel, E. J.; López, P. A.; Bottini, S. B.; Brignole, E. A. Computer-aided molecular design of solvents for separation processes. *AIChE J.* **1994**, *40*, 1349−1360.

(11) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352−2355.

(12) Gemhling, J.; Li, J.; Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 178−193.

(13) Larsen, B. L.; Rasmussen, P.; Fredenslund, A. Modified UNIFAC group-contribution model for the prediction of phase equilibria and heats of mixing. *Ind. Eng. Chem. Res.* **1987**, *26*, 2274−2286.

(14) Hansen, C. The universality of the solubility parameter. *Ind. Eng. Chem. Prod. Res. Dev.* **1969**, 2−11.

(15) Lindvig, T.; Michelsen, M. L.; Kontogeorgis, G. M. A Flory−Huggins model based on the Hansen solubility parameters. *Fluid Phase Equilib.* **2002**, *203*, 247−260.

(16) Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **1987**, *57*, 233−243.

(17) Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, *183*−*184*, 183−208.

(18) Marrero-Morejón, J.; Pardillo-Fontdevila, E. Estimating of pure compound properties using group-interaction contributions. *AIChE J.* **1999**, *45*, 615−621.

(19) Kim, K. J.; Diwekar, U. M. Efficient combinatorial optimization under uncertainty. 2. Application to stochastic solvent selection. *Ind. Eng. Chem. Res.* **2002**, *41*, 1285−1296.

(20) Kalagnanam, J. R.; Diwekar, U. M. An efficient sampling technique for off-line quality control. *Technometrics* **1997**, *39*, 308−319.

(21) Chaudhuri, P.; Diwekar, U. M. Synthesis approach to the determination of optimal waste blends under uncertainty. *AIChE J.* **1999**, *45*, 1671−1687.

(22) Kim K. J.; Diwekar, U. M. Characterizing Sampling Error for Optimization Under Uncertainty: A Fractal Geometry Approach. submitted to *Operations Research*, 2004.

(23) Deb, K. An efficient constraint handling method for genetic algorithms. *Comput. Methods Appl. Mech. Eng.* **2000**, *186*, 311−338.

(24) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison−Wesley: Reading, MA, 1989.

(25) Constantinou, L.; Bagherpour, K.; Gani, R.; Klein, J. A.; Wu, D. T. Computer aided product design: Problem formulations, methodology and applications. *Comput. Chem. Eng.* **1996**, *20*, 685−702.

(26) Barton, A. *CRC Handbook of Solubility Parameters and Other Cohesion Parameters*; CRC Press: Boca Raton, FL, 1983.

(27) Joback, K. G. Solvent substitution for pollution preventation. In *Pollution Prevention via Process and Product Modifications*; AIChE Symposium Series, Vol. 90; American Institute of Chemical Engineers: New York, 1994; pp 98−104.

Received for review September 9, 2004
Revised manuscript received June 8, 2005
Accepted June 24, 2005

IE049126L