JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

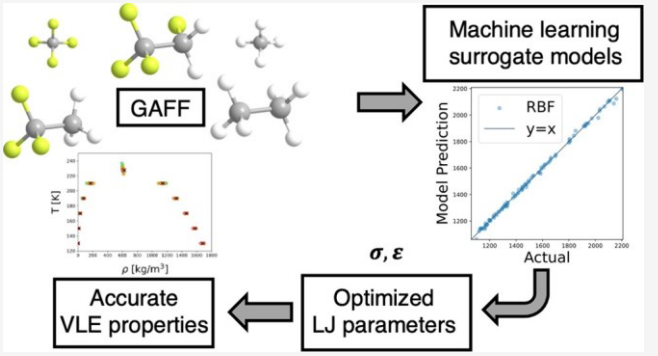
Machine Learning-Enabled Development of Accurate Force Fields for Refrigerants

Ning Wang, Montana N. Carlozo, Eliseo Marin-Rimoldi, Bridgette J. Befort, Alexander W. Dowling, and Edward J. Maginn*

Cite This: https://doi.org/10.1021/acs.jctc.3c00338 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Hydrofluorocarbon (HFC) refrigerants with zero ozone-depleting potential have replaced chlorofluorocarbons and are now ubiquitous. However, some HFCs have high global warming potential, which has led to calls by governments to phase out these HFCs. Technologies to recycle and repurpose these HFCs need to be developed. Therefore, thermophysical properties of HFCs are needed over a wide range of conditions. Molecular simulations can help understand and predict the thermophysical properties of HFCs. The prediction capability of a molecular simulation is directly tied to the accuracy of the force field. In this work, we applied and refined a machine learning-based workflow to optimize the Lennard-Jones parameters of classical HFC force fields for HFC-143a (CF_3CH_3), HFC-134a (CH_2FCF_3), R-50 (CH_4), R-170 (C_2H_6), and R-14 (CF_4). Our workflow involves liquid density iterations with molecular dynamics simulations and vapor−liquid equilibrium (VLE) iterations with Gibbs ensemble Monte Carlo simulations. Support vector machine classifiers and Gaussian process surrogate models save months of simulation time and can efficiently select optimal parameters from half a million distinct parameter sets. Excellent agreement as evidenced by low mean absolute percent errors (MAPEs) of simulated liquid density (ranging from 0.3% to 3.4%), vapor density (ranging from 1.4% to 2.6%), vapor pressure (ranging from 1.3% to 2.8%), and enthalpy of vaporization (ranging from 0.5% to 2.7%) relative to experiments was obtained for the recommended parameter set of each refrigerant. The performance of each new parameter set was superior or similar to the best force field in the literature.



INTRODUCTION

Hydrofluorocarbon (HFC) refrigerants are used widely in heating, ventilation, air conditioning, and refrigeration systems.[1] Because of the high global warming potentials (GWPs) of HFCs, recent regulation requires their gradual phase out over the next two decades.[2] Many HFCs in use are (near-)azeotropic mixtures, and novel processes must be developed to separate the components for subsequent recycling or repurposing. As such, technologies to recycle these HFCs are developing quickly. For example, recent work demonstrates how ionic liquids (ILs) can facilitate HFC separation.[3−7] Other works have used machine learning to search for new refrigerants and estimate HFC solubility in ILs.[8,9] However, all HFC separation endeavors require the often limited knowledge of the thermophysical properties of these HFC mixtures.[1,10] Computer-aided molecular design of HFC separations[1,10,11] has shown promise to accelerate the development of novel processes to meet the required goals. Accurate vapor−liquid equilibrium (VLE) data of HFC mixtures are desired, as is a microscopic understanding of the underlying physics that governs their physical properties.

Classical molecular dynamics (MD) and Monte Carlo (MC) simulations have played an important role to obtain thermophysical property predictions while providing microscopic insights of the various physical phenomena of practical interest. A key component of MD and MC simulations is an accurate description of molecular interactions, which has been traditionally achieved using classical molecular force fields. Force fields (FFs) make use of simple algebraic formulas parametrized to match the macroscopic properties of interest at the desired thermodynamic conditions. One recurring strategy to obtain FF parameters is to conduct quantum mechanical calculations of isolated molecules to obtain the parameters for covalent and electrostatic interactions. For dispersion interactions, parameters are often hand-tuned so that experimental properties are reproduced. Further manual

Special Issue: Machine Learning for Molecular Simulation

Received: March 24, 2023

refining might be required for both interaction types to improve the experimental agreement. Variants of this methodology have been used by general purpose FFs such as the Generalized Amber Force Field (GAFF)[12] or the Optimized Potentials for Liquid Simulations (OPLS) force field.[13] For the case of HFCs, molecule-specific FFs have been developed to target different properties. For instance, a semirigid all-atom model consisting of Halgren's Buf 14-7 function and Coulombic potential was introduced to describe HFC-125 (CF_3CHF_2) and HFC-134a (CH_2FCF_3).[14] Although fixed bond lengths and angles were used and only the torsion about the carbon−carbon bond was considered, the model demonstrated reasonable prediction of thermodynamic properties and showed great potential for later studies on ethane-type refrigerants. Lisal and Vacek[15] extended the semirigid all-atom model to study the conformational difference of fluoroethane isomers, including HFC-134 (CHF_2CHF_2), HFC-134a, HFC-143 (CHF_2CH_2F), HFC-143a (CF_3CH_3), HFC-152 (CH_2FCH_2F), and HFC-152a (CHF_2CH_3). Potter et al.[16] applied the 12-6 Lennard-Jones (LJ) potential and Coulombic interactions for fluoromethanes, like HFC-32, HFC-23 (CHF_3), and R-14 (CF_4). The latent enthalpy of HFC-32, HFC-23, and R-14 showed mean absolute percent errors (MAPEs) of 11%, 3%, and 12%, respectively, in comparison to experimental values. Also, the vapor pressure of HFC-32, HFC-23, and R-14 had MAPEs of 122%, 75%, and 32%, respectively. The model was able to accurately predict the orthobaric densities of all three refrigerants but was unable to replicate the structure of HFC-32, likely because of lack of polarization. Higashi and Takada also applied all-atom 12-6 LJ and Coulombic pair potentials to study various properties, including partial radial distribution functions, coordination numbers, pair potential energy distribution functions, the lifetime of clusters, and the liquid structure of HFC-32.[17] A semirigid all-atom force field was developed by Fermeglia et al.,[18] which includes a 9-6 LJ potential and Coulombic interactions for HFC-32, HFC-161 (CH_3CH_2F), HFC-152a, HFC-134, HFC-134a, HFC-143, HFC-143a, and HFC-125. Reasonable agreement was observed for intramolecular energetic and geometric properties, like bond lengths and angles, relative to experiments. The simulated saturated liquid and vapor densities showed an average deviation of approximately 2% from the corresponding experimental values. Four potential models with different LJ parameters and partial charges were proposed for HFC-134a by Peguin et al.[19] with the OPLS-AA functional form. The model with the best performance showed average deviations of 0.7%, 4.4%, 3.2%, 0.2%, 0.1%, 6.2%, 0%, and 2.2% when compared to the experimental values of liquid density, vapor density, vapor pressure, critical density, critical temperature, critical pressure, boiling temperature, and heat of vaporization, respectively, and was considered as the first all-atom model with 12-6 LJ potential to describe HFC-134a. Two all-atom FFs with the same intramolecular parameters but different LJ parameters and different partial charges were derived for HFC-152a by Yang et al.[20] based on the AMBER force field. The best FF showed mean absolute deviations of 0.89%, 2.32%, and 2.84% relative to experiments from 250 to 360 K for saturated liquid density, saturated vapor density, and vapor pressure, respectively, and average deviations of 0.49%, 0.38%, 3.80%, 0.34%, and 0.45% for critical density, critical temperature, critical pressure, boiling temperature, and heat of vaporization at 308.15 K of pure HFC-152a, respectively. Their study also provided justification for the capability of the aforementioned force fields to accurately predict phase equilibrium properties of the binary mixture of HFC-152a and HFC-32. An all-atom force field using the AMBER functional form was applied to predict the thermophysical properties of pure HFC-161 and VLE properties of the HFC-161 and HFO-1234yf mixture.[21] The average absolute relative deviations of vapor pressure, saturated liquid and vapor densities, critical temperature, critical pressure, and critical density of pure HFC-161 relative to experiments were 1.37%, 3.87%, 1.86%, 0.40%, 1.86%, and 1.47%, respectively. Raabe also hand-tuned an all-atom force field for HFC-32, which can accurately reproduce VLE properties, including saturated liquid density, saturated vapor density, vapor pressure, and enthalpy of vaporization with MAPEs of 1.45%, 7.17%, 4.31%, and 2.48%, respectively.[22,23] This hand-tuned FF is used as a comparison for the optimization of the HFC-32 FF[24] which serves as the basis of this work.

Manual fine-tuning of classical FFs is time consuming, computationally inefficient, and requires substantial human expertise. Recently, automated workflows have been proposed to accelerate the creation and refinement of classical FFs. For example, the CHARMM lipid force field[25] has been recently reparametrized using a semiautomated workflow that consistently included long-range dispersion through the LJ particle-mesh Ewald method. The OpenFF initiative[26−28] develops a software infrastructure that accelerates the creation, refinement, and validation of small organic molecule force fields.

Automated force field development opens the possibility of leveraging machine learning models to increase the accuracy and confidence of the predictions obtained from molecular FFs. For example, Madin and Shirts[29] showed that training a Gaussian process (GP) to reproduce the different physical properties of small molecules (e.g., alcohols, ketones, alkanes,) while tuning LJ parameters using differential evolution yielded significantly more accurate results than OpenFF. Multifidelity surrogate modeling strategies have shown similar results in FF optimization[30,31] and surrogate modeling, in general, and seen success in structure optimization.[32] In our previous work, we developed a semiautomated workflow that leverages machine learning models to accelerate the proposal of new parameters using the LJ potential energy function. Specifically, we used a GP to emulate the results of a molecular simulation, and these results were used to tune LJ parameters for HFC-32 and HFC-125 FFs. The results indicate that the parameters developed using this method offer better predictions of density, saturated vapor density, vapor pressure, and enthalpy of vaporization relative to Raabe's force field and to the popular GAFF model, which was parametrized for small organic molecules.[24]

Intramolecular parameters and partial charges are reliably calculated without egregious computational effort from quantum mechanical (QM) calculations, while the LJ parameters are relatively difficult to compute.[33] This is particularly important because the VLE properties we are interested in for HFC separation depend heavily on the LJ parameters and partial charges. At the same time, they are relatively insensitive to the intramolecular parameters.[34] Therefore, LJ parameters are often optimized by fitting experimental data for liquid phase properties such as heat of vaporization, molecular volume, or hydration free energy.[35]

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

This is most commonly achieved by manual tuning,[13,36,37] derivative-based optimization,[38−41] and stochastic heuristics such as genetic algorithms.[42−45] In this work, we improve the machine learning workflow from Befort et al.[24] to optimize the LJ parameters of the classical FFs of HFC-143a, HFC-134a, R-50, R-170, and R-14 with the structures shown in Figure 1.

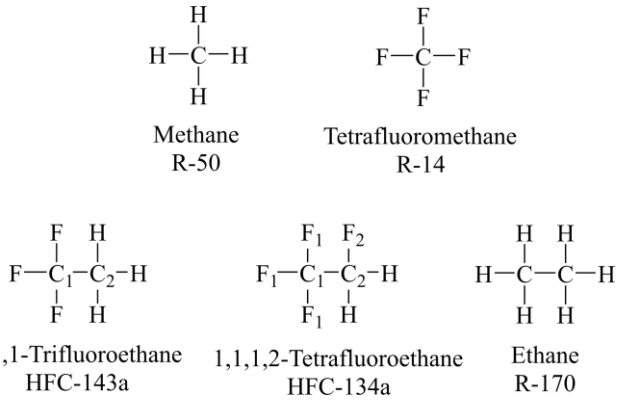


Figure 1. 2D chemical structures and shorthand names of studied refrigerants. Different notation reflects different atom types for LJ parameters, and there is no shared atom type between refrigerants. There are two atom types for R-50 (C and H), R-14 (C and F), and R-170 (C and H), five atom types for HFC-134a (C_1, C_2, F_1, F_2, and H), and four atom types for HFC-143a (C_1, C_2, F, and H).

The LJ parameters of these five refrigerants yield comparable or more accurate VLE property predictions than other available FFs in the literature. This workflow, consisting of a support vector machine (SVM) classifier, GP surrogate model, liquid density (LD) iterations using MD simulations, and VLE iterations using Gibbs ensemble Monte Carlo (GEMC) simulations, has been gradually refined and can be generalized to a broader family of refrigerants. The resulting parameters are easily accessible via public repositories, and the Foyer XML format facilitates the implementation in various simulation engines. The results of this work will establish a foundation for calibrating a generalized LJ parameter set for classical HFC FFs in future work.

METHODS

Problem Statement. The HFC FFs developed in this study use the GAFF[12] functional form:

$$\mathcal{V} = \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi[\cos(n\phi - \gamma) + 1] + \sum_i \sum_{j>i} \left\{ 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (1)$$

where $\mathcal{V}$ is the total potential energy, k is the force constant, r_0 is the nominal bond length, θ_0 is the nominal bond angle, γ is the nominal dihedral angle, n is the multiplicity, ε_{ij} and σ_{ij} are the LJ parameters, q is the partial charge, ϵ_0 is the vacuum permittivity, and r_{ij} is the distance between atom i and atom j. Intramolecular bonded parameters were directly taken from GAFF. Partial charges were obtained using density functional theory calculations at the B3LYP/6-311++g(d,p) level of theory using the RESP method.[46] The unlike interactions were calculated using the Lorentz−Berthelot combining rule. We chose to use the Lorentz−Berthelot combining rule partly for convenience and partly because it is widely used in the literature. We recognize, however, that relaxing this constraint might lead to more accurate force fields, and so we plan to investigate the combining rule in future studies. Following traditional Class I force field approaches, we assigned atom types based on their local environment. These atom types are defined in Figure 1. Other atom typing schemes could be used, such as the topology automated FF interactions (TAFFI) framework,[47] which leverages an automated atom-typing procedure along with a combination of hierarchical quantum chemical calculations and iterative MD simulations to parametrize transferable FF models.

The goal of this work is to estimate optimal parameters σ and ε using experimental liquid density and VLE data evaluated at different state points for five refrigerants.

General Workflow. The ε and σ parameters of the LJ potential were optimized using our machine learning-aided workflow described in Figure 2. First, the number of atom types and upper/lower parameter bounds were determined from domain knowledge. The initial parameter sets were generated using Latin hypercube sampling (LHS) for each refrigerant. MD simulations were conducted to compute the liquid density (LD) for the initial parameter sets. The simulation results from the first LD iteration were used to train the SVM classifier and GP surrogate model which then were applied to select 500,000 distinct LJ parameter sets for the next iteration. The selection criteria involved experimental root-mean-square error (RMSE) ranking and a space filling algorithm. LD iterations continued until 25 parameter sets with experimental RMSE less than 10 kg/m^3 were obtained. MC simulations were employed to calculate the VLE properties for the 25 parameter sets at different state points. Simulation results from MC were used to train different GP models for different VLE properties. The 25 well-separated parameter sets were selected using multiobjective optimization

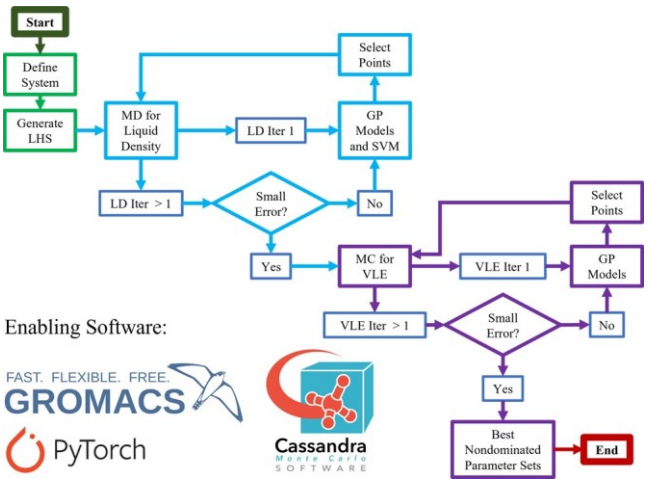


Figure 2. Overall workflow of machine learning-based optimization of LJ parameters of classical HFC FFs. Each color in the workflow represents a different step in the process. Green represents workflow initialization using Latin hypercube sampling (LHS). The blue loop uses Gaussian processes (GPs) and support vector machine (SVM) classifiers to estimate parameters using liquid density (LD) simulations and experimental data. Next, the purple loop uses Gibbs ensemble Monte Carlo (GEMC) simulations to predict vapor−liquid equilibrium (VLE) properties and select the non-dominated points from all the samples. Red represents the end of the workflow.

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

concepts[48,49] for the next iteration. VLE iterations continued until a recommended parameter set was found with a satisfactorily small error in all VLE properties.

The remainder of this section describes the workflow (Figure 2) in detail.

MD Simulations. The GROMACS package[50] was applied to perform MD simulations under the NPT ensemble to compute liquid density at five different temperatures and corresponding vapor pressures. Detailed temperatures and pressures for each refrigerant are documented in Table S1. The number of molecules was 300 for one-carbon refrigerants, R-50 and R-14, and 150 for two-carbon refrigerants, HFC-143a, HFC-134a, and R-170. Molecules were randomly inserted into the simulation box using Packmol[51,52] with an initial density of 1000 kg/m^3 for HFC-143a, 1000 kg/m^3 for HFC-134a, 300 kg/m^3 for R-50, 500 kg/m^3 for R-170, and 1200 kg/m^3 for R-14. A steepest descent algorithm was used in energy minimization. The maximum step size was 0.01 nm, the maximum force was 100.0 kJ/mol/nm, and the maximum number of minimization steps was 50,000. The equilibration phase ran for 0.5 ns, which was followed by a 2.5 ns production run. A leapfrog algorithm[53] for integrating Newton's equations of motion was used. The time step was 1 fs. For the equilibration, the thermostat was the modified Berendsen thermostat[54] with a time constant of 0.1 ps, while the barostat was the Berendsen barostat[55] with a time constant of 0.5 ps. For the production phase, the thermostat was the modified Berendsen thermostat[54] with a time constant of 0.5 ps, while the barostat was the Parrinello−Rahman barostat[56] with a time constant of 1.0 ps. Long-range dispersion corrections were applied for energy and pressure. The fast smooth particle-mesh Ewald method[57] was used for long-range electrostatics. P-LINCS[58] was the constraint used for all bonds with the highest order in the expansion of the constraint coupling matrix equal to 8 and the number of iterations to correct for rotational lengthening in LINCS was set to 4. The short-range cutoff for both Coulombic and van der Waals interactions was 1 nm. Periodic boundary conditions were applied in the x, y, and z directions. All the other settings are listed in the GitHub repository associated with this paper (https://github.com/dowlinglab/HFC-FFO).

MC Simulations. GEMC simulations were employed to study the VLE properties of the five refrigerants using the Cassandra package.[59] The initial number of molecules in the vapor phase and liquid phase was 160 and 640, respectively. The initial vapor and liquid boxes were randomly generated at experimental vapor and liquid density, respectively, using Packmol.[51,52] The liquid box was pre-equilibrated for 5000 sweeps under NPT MC simulations, which was followed by a GEMC equilibration on the whole system for 10,000 sweeps and a production run for 100,000 sweeps using Cassandra. The cutoffs of Coulombic and van der Waals interaction were 1.2 nm for the liquid box and 2.5 nm for the vapor box. Ewald summation was applied for long-range electrostatics with a relative accuracy of 10^{-5}. All bonds were fixed at the nominal bond length. Standard LJ tail corrections were applied to pressure and energy. All the other settings are listed on the GitHub page associated with this work (https://github.com/dowlinglab/HFC-FFO).

Classifier. We used a support vector machine (SVM) classifier to predict the phase of a MD simulation performed with a specific parameter set and state point, represented by the vector x. We label each simulation in the training data set $i \in \mathcal{I}$ either a liquid ($y_i = 1$) if the density is above the threshold or otherwise as a vapor ($y_i = -1$). The SVM then predicts the label for a new postulated simulation with inputs x* as follows:

$$\mathrm{sgn}(\mathbf{w}^T \phi(\mathbf{x}^*) + b) = \mathrm{sgn}\left(\sum_{i \in \mathcal{I}} y_i \alpha_i k(\mathbf{x}_i, \mathbf{x}_j) + b\right) \quad (2)$$

Here, sgn(·) is the sign function. x_i are the input variables (LJ parameters, temperature), and y_i are the labeled outputs for the simulations in the training data set $i \in \mathcal{I}$. Moreover, w and α are the primal and dual variables for the SVM training problem, which is fully described in ref 60. For this classifier, we use a radial basis function (RBF) kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2l^2}\right) \quad (3)$$

Here, $k(x_i, x_j)$ is the kernel function evaluated with respect to data x_i and x_j, and l is a tunable hyperparameter. In the context of eq 2, the function $\phi(\cdot)$ is defined such that $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$, and b is a constant. The SVM classifier was trained using `scikitlearn` with a stopping criterion of 0.001.

GP Model. Single-output GP surrogate models were used to predict the liquid density from MD simulations and saturated liquid density, saturated vapor density, vapor pressure, and enthalpy of vaporization from GEMC simulations. The inputs to the GP model, denoted $\mathbf{x} = [\sigma, \varepsilon, T] \in \mathbb{R}^m$, were LJ parameters and temperatures, and the output, denoted y, was the aforementioned properties. Throughout the workflow in Figure 2, we used an 80%−20% split between training and tested data. A linear mean function, denoted $m(\mathbf{x})$ was applied. For each LD iteration, we compared RBF (eq 3), Matérn$_\nu = \frac{3}{2}$, and Matérn$_\nu = \frac{5}{2}$ kernels. During the VLE iterations, all GP models used a RBF kernel, except for the GP model of vapor density, which used a Matérn$\nu = \frac{5}{2}$ kernel.

The Matérn kernels, shown in eq 4, are defined by the smoothness parameter ν:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l} d_{ij}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l} d_{ij}\right)$$

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 \quad (4)$$

Here, d_{ij} is the Euclidean distance between inputs x_i and x_j, $K_\nu(\cdot)$ is a modified Bessel function, and $\Gamma(\cdot)$ is the gamma function. The hyperparameter l was trained via maximum likelihood estimation (MLE):

$$\ln p(\mathbf{y}|\mathbf{X}, l) = -\frac{1}{2}\mathbf{y}^T(\mathbf{K}(\mathbf{X}, \mathbf{X}|l) + \sigma^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\ln|\mathbf{K}(\mathbf{X}, \mathbf{X}|l) + \sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi \quad (5)$$

Here, $p(y|X, l)$ is the log likelihood of observed data y (i.e., liquid density, vapor density, vapor pressure, or enthalpy of vaporization) given independent data matrix X and hyperparameters l. Let matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ represent the training input data, where each column x_i corresponds to output y_i for observation i in the training data set $\mathcal{I}$. Thus, the vector

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

y ∈ ℝ the output training data for a single thermophysical property. As such, the matrix K(X, X|θ) corresponds to the kernel function $k(\cdot, \cdot)$ evaluated element-wise for all pairwise combinations of the columns of X. Additionally, I is the identity matrix, n is the number of observations, $|\cdot|$ is the matrix determinant, and σ is the standard deviation of the observed data, i.e., random noise from MD or MC simulations.

The output predictions of a GP, by definition, follow a multivariate normal distribution. As such, for a new input x*, there is an analytic formula for the prediction mean and prediction variance:

$$\mu_*(\mathbf{x}^*|\theta) = m(\mathbf{x}^*|\theta) + \mathbf{k}(\mathbf{x}^*, \mathbf{X}|\theta)[\mathbf{K}(\mathbf{X}, \mathbf{X}|\theta) + \sigma^2\mathbf{I}]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{X}|\theta)) \quad (6a)$$

$$\sigma_*^2(\mathbf{x}^*|\theta) = k(\mathbf{x}^*, \mathbf{x}^*|\theta) - \mathbf{k}(\mathbf{x}^*, \mathbf{X}|\theta)[\mathbf{K}(\mathbf{X}, \mathbf{X}|\theta) + \sigma^2\mathbf{I}]^{-1}\mathbf{k}(\mathbf{X}, \mathbf{x}^*|\theta) \quad (6b)$$

These key mathematical properties make GPs excellent emulators of computationally expensive black-box functions, such as molecular simulation. Qualitatively, as the point x* approaches a point x_i in the training data set, the prediction uncertainty σ_*^2 decreases and approaches σ^2 in the limit. Likewise, as the point x* moves away from all points in the training data set X, the prediction uncertainty grows and eventually plateaus. Thus, a GP is a nonparametric model to interpolate between points in the training data set X and y where the kernel choice and hyperparmater θ adjust the smoothness of the mean and scale of the prediction uncertainty.[61] As shown in Figures S2 and S3 in the Supporting Information, the GP predictions can be unphysical, e.g., predict a negative liquid density, especially if the molecular simulation parameters are unrealistic.

LD Iterations. Each LD iteration starts with 1000 isothermal−isobaric (NPT) MD simulations to predict the liquid density of 200 parameter sets at five different temperatures. For the first LD iteration, we use the initial 200 parameter sets generated using a space-filling LHS algorithm. The generated parameter sets were constrained by upper and lower bounds, which were informed from domain knowledge and are documented in Table 1. The upper and lower bounds are determined to cover a broad range of physically reasonable values. Informed by the final nondominated parameter sets of HFC-32 and HFC-125 from ref 24, we slightly extended the upper and lower bounds for each parameter accordingly. The same bounds were initially used for R-14 as other refrigerants; however, very few parameter sets that give satisfying LD predictions were obtained after a few rounds of LD iterations, so we used a broader range of LJ parameters. We consider atoms in the same chemical environment as the same atom type, and no shared atom type is considered between different refrigerants. As shown in Figure 1, there are two atom types for R-50, R-14, and R-170, five atom types for HFC-134a, and four atom types for HFC-143a. Previous work has shown that this atom-typing scheme can lead to overparametrization.[24] Future work will be conducted on a more generalized atom-typing framework for a family of HFCs.

The pressures and temperatures set in the MD simulations were the experimental saturation conditions obtained from the REFPROP package[62] or the NIST Web site[63] (see Table S1).

Table 1. Lower and Upper Parameter Bounds for σ in Å and ε/k_B in Kelvin of R-50, R-14, R-170, HFC-134a, and HFC-143a

Refrigerant	σ	lower	upper	ε/k_B	lower	upper
R-50	σ_C	3.0	4.0	ε_C/k_B	20.0	75.0
	σ_H	1.5	3.0	ε_H/k_B	2.0	10.0
R-14	σ_C	2.0	4.0	ε_C/k_B	10.0	75.0
	σ_F	2.5	3.5	ε_F/k_B	15.0	50.0
R-170	σ_C	3.0	4.0	ε_C/k_B	20.0	75.0
	σ_H	1.5	3.0	ε_H/k_B	2.0	10.0
HFC-134a	σ_{C_1}	3.0	4.0	ε_{C_1}/k_B	20.0	75.0
	σ_{C_2}	3.0	4.0	ε_{C_2}/k_B	20.0	75.0
	σ_{F_1}	2.5	3.5	ε_{F_1}/k_B	15.0	40.0
	σ_{F_2}	2.5	3.5	ε_{F_2}/k_B	15.0	40.0
	σ_H	1.5	3.0	ε_H/k_B	2.0	10.0
HFC-143a	σ_{C_1}	3.0	4.0	ε_{C_1}/k_B	20.0	70.0
	σ_{C_2}	3.0	4.0	ε_{C_2}/k_B	20.0	70.0
	σ_F	2.0	4.0	ε_F/k_B	15.0	40.0
	σ_H	1.5	3.0	ε_H/k_B	2.0	10.0

The resulting simulations were classified into liquid or vapor phase using predefined density-based thresholds that depend on the simulated HFC. The LD thresholds of R-50, R-14, R-170, HFC-134a, and HFC-143a were 200, 1100, 320, 500, and 500 in kg/m^3, respectively. It is important to adjust the liquid density threshold for different molecules to avoid the supercritical region.

A SVM classifier with the RBF kernel was trained to predict whether a given set would yield a liquid or vapor thermodynamic state. The information generated from the previous iteration was used to train this model. Additionally, the GP model with the RBF kernel was trained to predict the simulation density from temperature and LJ parameters. A larger set of 500,000 candidate parameter sets was generated using the LHS method, and their thermodynamic state and densities were predicted with the trained SVM and GP, respectively. A pretrained SVM classifier was used to classify 500,000 candidate sets generated with LHS into liquid samples and vapor samples at the highest temperature. The GP surrogate model was then used to predict the simulation LD results of liquid and vapor samples at each temperature. We selected the best 200 liquid parameter sets across temperatures that yielded an experimental RMSE less than 25 kg/m^3. A space-filling algorithm shown in Algorithm 1 was used to ensure parameter diversity and maximize the distance between the points selected.

Algorithm 1 Space-Filling Downselection of Nondominated Points

```
Given: d > 0, maximum distance between points
Given: [ x_1, ..., x_n ] set of n candidate nondominated parameter sets
Given: N < n target number of candidate parameter sets to advance
Initialize set I = {1, ..., n}
while |I| > N do
    Choose a point from i ∈ I at random
    Compute L_1 norm ℓ_j = ||x_i − x_j||_1,   ∀j ∈ I\{i}
    Remove all points j where j < d, i.e., I ← I\{j : ℓ_j < d}
end while
Return I
```

Given this algorithm, we used a bisection method to adjust the input d in Algorithm 1 such that $|\mathcal{I}| = 200$. When there

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

were fewer than 200 liquid parameter sets with RMSE less than 25 kg/m^3, we selected the rest from the vapor sets with RMSE relative to experimental LD less than 25 kg/m^3 to keep the total number of parameter sets for each iteration at 200. This is a refinement of the procedure used by Befort et al.,[24] where they selected 100 parameter sets from both liquid and vapor samples to balance the inconsistency (e.g., classified as vapor but with low RMSE relative to experimental LD) between the classifier and GP model. However, we only focus on the consistent subset between classifier and GP (i.e., classified as liquid and with low RMSE relative to experimental LD). Without selecting sets from vapor samples, the whole workflow would be more efficient. MD simulations were conducted using the selected set of liquid parameters to start a new iteration of the workflow.

LD iteration stops when the 25 best performing, well-separated parameter sets were selected using an RMSE between simulated and experimental LD of less than 10 kg/m^3. We performed four, three, four, four, and five LD iterations and obtained 800, 600, 800, 800, and 1000 simulated parameter sets for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively.

VLE Iterations. The Cassandra package[59] was applied to perform GEMC simulations to calculate VLE properties of the best performing 25 parameter sets for each refrigerant. The properties of interest included the saturated liquid density, saturated vapor density, vapor pressure, and enthalpy of vaporization, at five different temperatures. Four GP models were trained to predict the saturated liquid density, saturated vapor density, vapor pressure, and enthalpy of vaporization as a function of LJ parameters and temperature. The LD GP model was trained on all LD iterations (except for R-50 which was trained on the last three LD iterations). Parameter sets among all 500,000 parameter sets generated from LHS whose difference between predicted LD RMSE value from GROMACS and Cassandra exceeded 25 kg/m^3 were discarded. VLE properties of the remaining parameter sets were evaluated by four VLE GP surrogate models. The law of rectilinear diameter[64] was applied to calculate the critical temperature and critical density. The RMSE of each VLE property was calculated between GP predictions and experimental data across all five temperatures. The non-dominated parameter sets were obtained by examining the RMSE of these six VLE properties (i.e., saturated liquid density, saturated vapor density, vapor pressure, enthalpy of vaporization, critical temperature, and critical density) using the same approach as Befort et al.[24] More specifically, a parameter set is considered nondominated if there are no other evaluated parameter sets that can improve the RMSE of one VLE property without sacrificing the RMSEs for one or more other VLE properties. Thus, a nondominated parameter set captures an optimal trade-off between objectives, i.e., the nondominated property (finite discrete set) is closely related to Pareto optimally (infinite generalization) in multiobjective optimization.[48,49] A space-filling algorithm was applied to select 25 well-separated parameter sets out of nondominated parameter sets among VLE properties for the next VLE iteration. This was done to ensure diversity among the best parameter sets. After each VLE iteration, we checked if there were parameter sets with errors less than 5% in all VLE properties, including the saturated liquid density, saturated vapor density, vapor pressure, enthalpy of vaporization, critical temperature, and critical density. If no parameter set was found, VLE iteration continued. If parameter sets were found, the set exhibiting the smallest error was recommended. We performed four, three, three, two, and two VLE iterations and obtained 37, 22, 27, 29, and 18 nondominated parameter sets for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Only one parameter set with less than 3%, 2.5%, 3.5%, 2.5%, and 2% error in all VLE properties was recommended for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively.

RESULTS AND DISCUSSION

Machine Learning Model Results. The SVM classifier achieves staggering success in classifying vapor and liquid samples. Figure S1 is the confusion plot, which visualizes the classifier's accuracy. Table 2 shows the details of the accuracy of the classifier on the testing data for each iteration. For all iterations, the accuracy of the classifier is greater than 92%.

Table 2. Accuracy of the SVM Classifier on Testing Data of HFC-143a, HFC-134a, R-50, R-170, and R-14 for Each LD Iteration[a]

Refrigerant	LD-1	LD-2	LD-3	LD-4	LD-5
HFC-143a	96.0%	92.5%	92.8%	94.0%	–
HFC-134a	92.5%	96.5%	96.5%	–	–
R-50	96.0%	98.8%	97.8%	98.0%	–
R-170	96.5%	96.0%	96.2%	96.8%	–
R-14	95.5%	95.5%	93.8%	94.3%	94.8%

[a]LD-n, where n is the iteration number. The classifier was trained on all the available data from current and previous iterations.

Overall, the GPs predicted the results of molecular simulations very accurately. Figure S4 demonstrates that the GP consistently emulates the results of the molecular simulations after training on all discussed VLE properties. Moreover, GPs are particularly relevant to molecular simulations because of their ability to predict over state points. Transferability is an important topic in molecular simulations and the results of Figure S4 demonstrate that using a GP to estimate a molecular simulation inherently accounts for this. Table 3 demonstrates that the overall MAPEs between the GP predictions and molecular simulation results of all iterations of each refrigerant are always less than 10% and frequently less than 3%.

LD Iteration Results. We conducted four, three, four, four, and five LD iterations for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Figure 3 shows the cumulative number of parameter sets plotted versus liquid density MAPE. The performances of LD-1 (blue curves) were reasonably unsatisfying for all refrigerants since unphysical parameter sets were randomly selected by the LHS. Vapor simulations lead to spikes in LD-1 and "jumps" in the curves for later iterations. Vapor density was small relative to liquid density, causing a normalization error in MAPE calculation. Spikes near 100% in LD-1 mean that simulations predicted that the molecule was in the vapor phase at all five temperatures we studied. At the same time, jumps near 20%, 40%, 60%, and 80% in later iterations mean that simulations vaporized at one, two, three, and four studied temperatures, respectively. Over the course of the iterations, the GP surrogate model showed significant improvement, which was reflected by more parameter sets with relatively low MAPE as the number of iterations increased. With more iterations being performed, more

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Overall MAPEs of Liquid Density, Vapor Density, Vapor Pressure, and Enthalpy of Vaporization between GP Predictions and MD or GEMC Results of HFC-143a, HFC-134a, R-50, R-170, and R-14 for all LD or VLE Iterations

Property	Simulation	HFC-143a	HFC-134a	R-50	R-170	R-14
ρ_l	MD	0.89%	0.63%	0.61%	0.84%	0.69%
ρ_l	GEMC	0.3%	0.2%	0.2%	0.2%	0.2%
ρ_v	GEMC	4.6%	9.6%	2.7%	2.9%	4.9%
P_{vap}	GEMC	2.5%	7.6%	1.9%	1.8%	6.5%
ΔH_{vap}	GEMC	0.5%	0.3%	0.2%	0.3%	0.2%

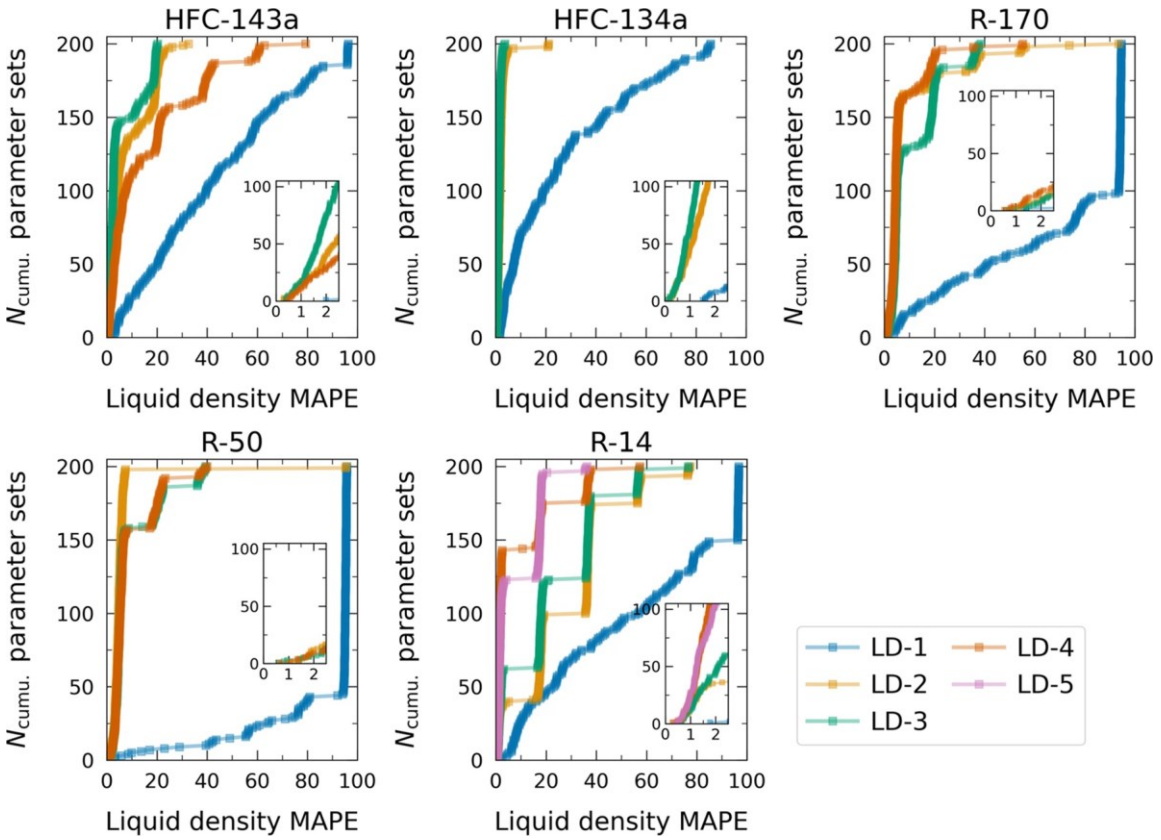


Figure 3. Cumulative number of parameter sets of each LD iteration (LD-n, where n is the iteration number) as a function of liquid density MAPE relative to experimental data for HFC-143a, HFC-134a, R-50, R-170, and R-14. n equals to 4, 3, 4, 4, and 5 for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Blue, yellow, green, red, and purple represent LD-1, LD-2, LD-3, LD-4, and LD-5, respectively. The inset plot is zoomed for MAPE less than 2.5%.

parameter sets provide accurate liquid density. Although the improvement from later iterations ($n \geq 3$) was limited, additional parameter sets with low MAPE were generated. LD iterations stopped when enough (25) parameter sets with RMSE less than 10 kg/m^3 relative to the experiments were obtained.

VLE Iteration Results. We conducted four, three, three, two, and two VLE iterations for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Figure 4 shows the cumulative number of parameter sets as a function of MAPE of VLE properties, including liquid density, vapor density, vapor pressure, enthalpy of vaporization, critical temperature, and critical density. The critical temperature and critical density were not directly included in the optimization workflow and were calculated using the law of the rectilinear diameter.[64] Higher MAPE represents a higher error in estimating certain VLE properties. Significant improvement was observed in VLE-2 relative to VLE-1 in vapor density, vapor pressure, enthalpy of vaporization, and even in critical temperature and critical density. The performance of each VLE iteration is gradually improved based on an overall evaluation of liquid density, vapor density, vapor pressure, enthalpy of vaporization, critical temperature, and critical density. For R-50, fewer parameter sets were found giving lower liquid density MAPE in VLE-2 and 3 compared to VLE-1, since the whole workflow balanced the performance of these six VLE properties using multiobjective optimization concepts.[48,49]

Final Parameter Sets Determination. Finally, we obtained 37, 22, 29, 27, and 18 nondominated parameter sets for HFC-143a, HFC-134a, R-170, R-50, and R-14, respectively, which are plotted in Figure 5. For HFC-143a and HFC-134a, all nondominated parameter sets provide a MAPE less than 20%, and most are less than 10%. The MAPEs of all promising parameter sets are less than 30%, 10%, and 8% for R-50, R-170, and R-14, respectively. Most interesting, the σ_{F_i} values in all nondominated parameter sets for HFC-143a, HFC-134a, and R-14 are all contained in a narrow range. Mathematically, this suggests the physical

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

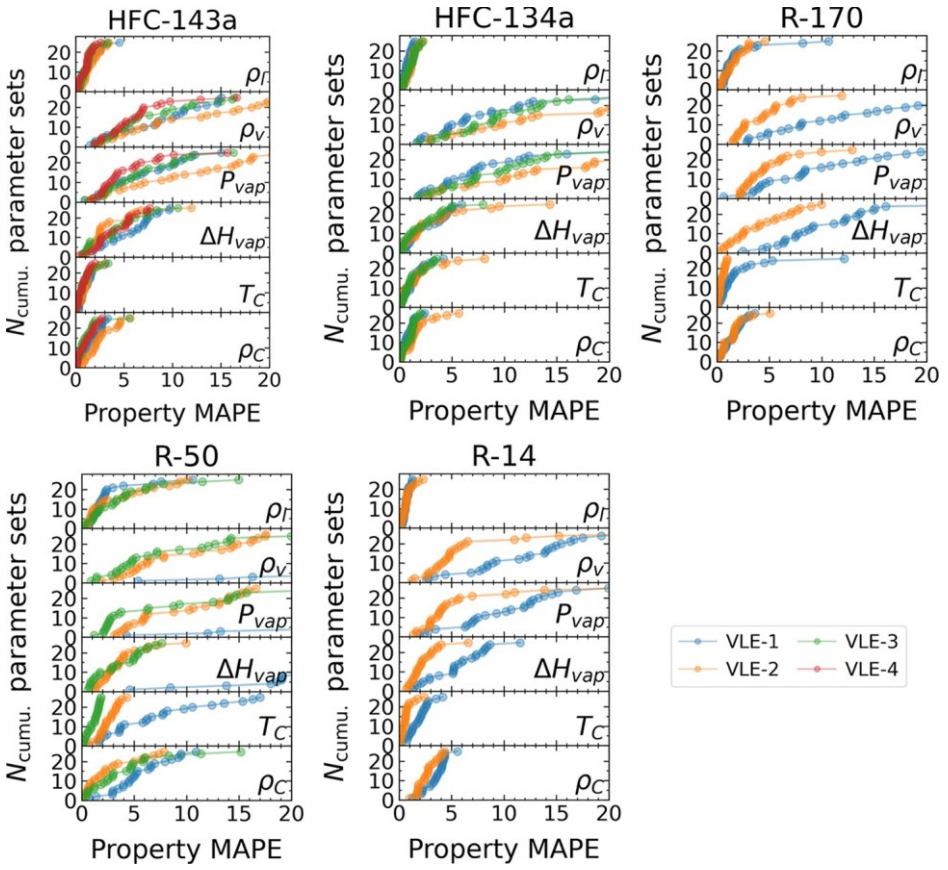


Figure 4. Cumulative number of parameter sets of each VLE iteration (VLE-n, where n is the iteration number) as a function of MAPE of liquid density (ρ_l), vapor density (ρ_v), vapor pressure (P_{vap}), enthalpy of vaporization (ΔH_{vap}), critical temperature (T_C), and critical density (ρ_C) relative to corresponding experimental data for HFC-143a, HFC-134a, R-50, R-170, and R-14. n equals 4, 3, 3, 2, and 2 for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Blue, yellow, green, and red represent VLE-1, VLE-2, VLE-3, and VLE-4, respectively.

property predictions are very sensitive to σ_{F_1}, and this parameter can be inferred with confidence. This is consistent with the observation that fluorine-containing molecules often exhibit unusual physical properties, such that results are highly sensitive to the fluorine parameters. In contrast, the nondominated sets encompass a large range of values for other parameters such as σ_{C_1} and σ_{C_2}. Mathematically, this means there are several distinct values of the parameters that yield good predictions. Physically, this can be interpreted as possible offsetting errors, where the carbon parameters are "effective" parameters that together capture average properties. These results are consistent with our prior work,[24] in which we show via sensitivity and identifiability analyses that the best FF parameters are locally unique. This is not surprising, as nonlinear model calibration problems are nonconvex optimization problems that often exhibit many locally optimal solutions. Our analysis mitigates this concern in part by identifying the nondominated parameter sets that capture trade-offs between multiple physical properties. Furthermore, we hypothesize that creating a generic refrigerant FF will further remove the degeneracy illustrated in Figure 5.

For each system, we recommend a single parameter set that exhibits less than 3% (HFC-143a), 2.5% (HFC-134a), 3.5% (R-50), 2.5% (R-170), and 2% (R-14) error in all VLE properties. These sets are shown in bold blue in Figure 5. The recommended LJ parameters of these five refrigerants are documented in Table 4. Partial charges and intramolecular parameters are obtained from GAFF and ab initio calculations and are documented in Table 4 and Tables S2−S4, respectively.

Our optimized parameter sets for HFC-143a, R-50, R-170, and R-14 show significantly better performance than classical GAFF[12] as can be seen in Figure 5 and Table 5. The MAPEs of GAFF are generally 1−2 orders of magnitude larger than ours. The recommended parameter set of HFC-134a is superior to the OPLS model of Peguin et al.[19] with smaller MAPEs for all VLE properties.

The recommended parameter sets for R-50, R-170, and R-14 are comparable to the united-atom Mie potentials of Potoff et al.[66] who fit three parameters to experimental data: the repulsive exponent n, σ, and ϵ. In this work, we fixed the repulsive exponent to 12 and only tuned σ and ϵ while achieving similar accuracy. For example, our model can more accurately predict vapor density and enthalpy of vaporization of R-50 than the model from Potoff et al.[66] For R-170, the MAPE of enthalpy of vaporization of Potoff is significantly larger than ours, while other properties showed similar accuracy.

The recommended parameter sets of R-50 and R-170 present a similar or better performance relative to the TraPPE[65] force field. For example, the MAPEs of our model for R-170 are 1.8%, 1.6%, 2.1%, and 2.0% for liquid density, vapor density, vapor pressure, and enthalpy of vaporization, respectively, while the MAPEs of TraPPE model are 0.8%, 4.6%, 4.0%, and 8.2%.

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

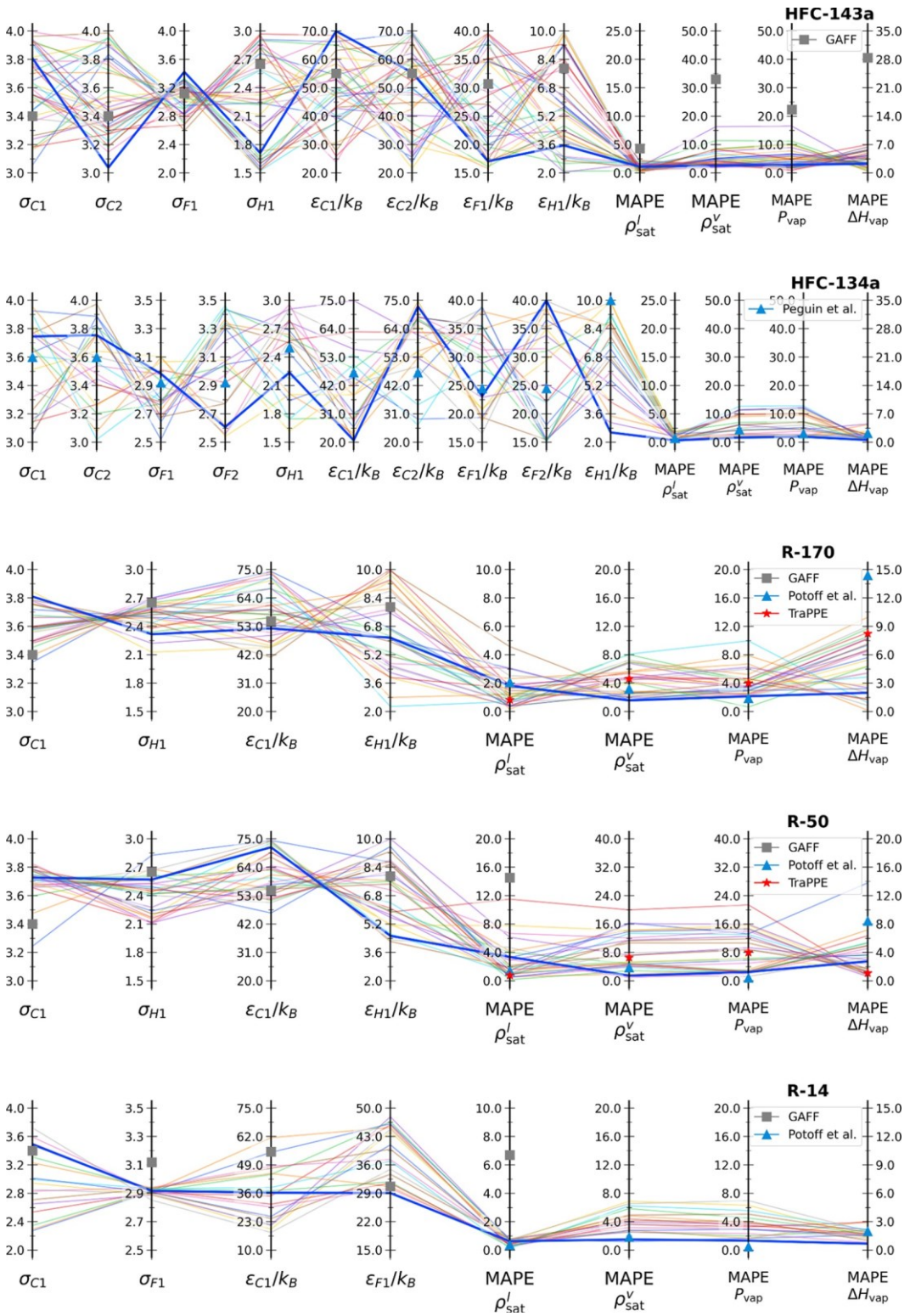


Figure 5. LJ parameter sets with good performance of HFC-143a, HFC-134a, R-50, R-170, and R-14. Each line represents a parameter set. The FF represented by the bold blue line is recommended due to its low MAPE for all properties less than 3%, 2.5%, 3.5%, 2.5%, and 2% for HFC-143a, HFC-134a, R-50, R-170, and R-14, respectively. Others are the nondominated parameter sets. The gray, green, and red squares are the performance of GAFF,[12] TraPPE,[65] and available literature,[19,66] respectively, and their detailed LJ parameters can be found in the Supporting Information. The GAFF MAPEs of R-50, R-170, and R-14 are very inaccurate and not shown here.

The VLE envelope and vapor pressure and enthalpy of vaporization as functions of the temperature of our recommended parameter sets compared with GAFF,[12] TraPPE,[65] other literature values,[19,66] and REFPROP experimental results[62] are shown in Figures 6 and 7, respectively. The corresponding VLE envelopes and vapor pressure and enthalpy of vaporization versus temperature plots of all the nondominated parameter sets are presented in

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Recommended Partial Charges (q) and LJ Parameters (σ and ε) for HFC-134a, HFC-143a, R-170, R-50, and R-14

Refrigerant	Atom	Atom type	q	σ (Å)	ε (K)
HFC-134a	C_1	c3	0.61542	3.745	20.73
	C_2	c3	−0.020709	3.754	72.61
	F_1	f	−0.210427	2.982	23.13
	F_2	f	−0.193556	2.607	39.98
	H	h1	0.115063	2.237	2.55
HFC-143a	C_1	c3	0.78821	3.809	69.97
	C_2	c3	−0.583262	3.037	55.40
	F	f	−0.252614	3.424	17.05
	H	hc	0.184298	1.711	3.55
R-170	C	c3	−0.00612	3.810	52.20
	H	hc	0.00204	2.316	6.16
R-50	C	c3	−0.512608	3.727	71.73
	H	hc	0.128152	2.569	4.53
R-14	C	c3	0.781024	3.490	36.24
	F	f	−0.195256	2.917	29.07

Table 5. MAPEs of Liquid Density (ρ_l), Vapor Density (ρ_v), Vapor Pressure (P_{vap}), and Enthalpy of Vaporization (ΔH_{vap}) of Parameter Sets from This Work (Recommended), GAFF,[12] Literature,[19,66] and TraPPE[65] for HFC-134a, HFC-143a, R-170, R-50, and R-14[a]

Refrigerant	Model	MAPE$_{\rho_l}$	MAPE$_{\rho_v}$	MAPE$_{P_{vap}}$	MAPE$_{\Delta H_{vap}}$
HFC-143a	This work	1.1%	2.6%	2.8%	2.3%
	GAFF[12]	4.3%	33.0%	22.3%	28.4%
HFC-134a	This work	0.3%	1.6%	1.9%	0.5%
	Peguin[19]	2.2%	3.1%	4.4%	0.7%
R-50	This work	3.4%	1.4%	2.5%	2.7%
	GAFF[12]	14.5%	90.8%	59.8%	33.4%
	Potoff[66]	1.4%	3.7%	0.8%	8.4%
	TraPPE[65]	0.8%	6.5%	8.0%	1.1%
R-170	This work	1.8%	1.6%	2.1%	2.0%
	GAFF[12]	22.4%	180.3%	83.0%	50.5%
	Potoff[66]	2.1%	3.2%	1.9%	14.4%
	TraPPE[65]	0.8%	4.6%	4.0%	8.2%
R-14	This work	0.6%	1.5%	1.3%	0.7%
	GAFF[12]	6.7%	72.0%	62.9%	61.5%
	Potoff[66]	0.4%	1.9%	0.5%	2.0%

[a]The lowest MAPE for each property of each refrigerant is shown in bold.

Figures S5 and S6. Detailed values of nondominated LJ parameters and their corresponding VLE properties are documented in the Supporting Information.

Our proposed nondominated parameter sets provide an excellent estimation of the VLE envelope, vapor pressure, and enthalpy of vaporization compared to the experiments. Small interactions at lower temperatures lead to small error spreads for vapor in the VLE envelope, as shown for R-50 in Figure S5. Previous FFs[19,65,66] precisely estimate the VLE envelope.

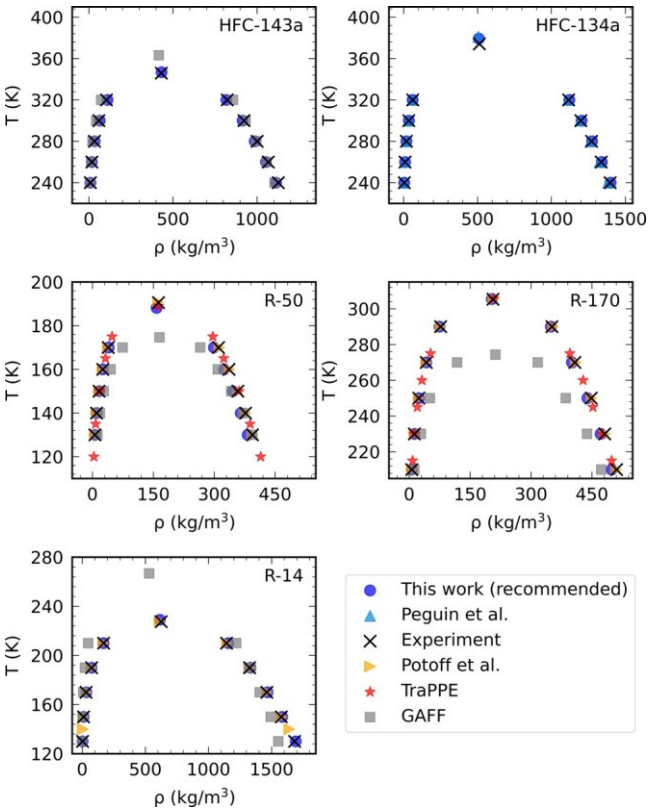


Figure 6. VLE envelopes of the recommended parameter set of HFC-143a, HFC-134a, R-50, R-170, and R-14 (blue circles) with comparison to GAFF[12] (gray squares), TraPPE[65] (red stars), literature[19,66] (blue and yellow triangles), and REFPROP experimental results[62] (black crosses).

Error around the critical point is expected for GAFF. Some literature FFs[19,66] can also predict vapor pressure and enthalpy of vaporization accurately. The TraPPE model estimates vapor pressure quite well, but cannot precisely predict the temperature dependence of enthalpy of vaporization. Prominent deviations between the experiment and GAFF are observed for vapor pressure and enthalpy of vaporization.

Notably, taking R-170 as an example, the wall clock time for each LD iteration is around 2 h, while each VLE iteration is much more computationally expensive and can be finished in 3−4 days. Because each LD iteration requires 1000 simulations (1 core per simulation) and each VLE iteration requires 125 simulations (2 cores per simulation), the total CPU times are 2−3 months and 2−3 years for each LD and VLE iteration, respectively. Thanks to parallel computing, the whole parametrization process can be completed within weeks and save months of computational expenses relative to the traditional hand-tuning process.

CONCLUSIONS

Our machine learning-enabled force field optimization workflow has been refined and generalized to HFC-143a, HFC-134a, R-50, R-170, and R-14. Multiple nondominated parameter sets were found for each refrigerant. A recommended LJ parameter set based on the GAFF functional form is proposed for each refrigerant which surpasses the previously developed force fields in the literature. Based on the five refrigerants studied in this work and HFC-32 and HFC-125 studied before,[24] a generalized HFC force field will be

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

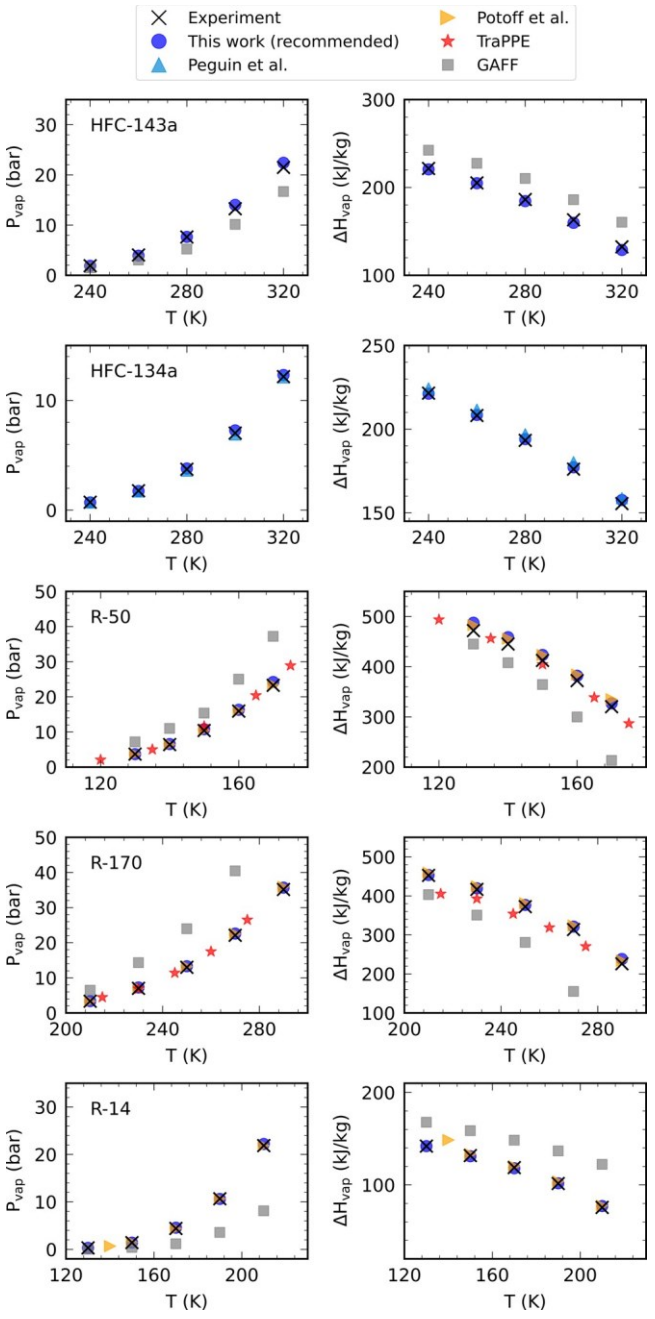


Figure 7. Vapor pressure and enthalpy of vaporization of HFC-143a, HFC-134a, R-50, R-170, and R-14 (blue circles) with comparison to GAFF[12] (gray squares), TraPPE[65] (red stars), literature[19,66] (blue and yellow triangles), and REFPROP experimental results[62] (black crosses).

calibrated in future work. Additionally, we are exploring Bayesian optimization to automate the selection of the single next best simulation (instead of batches of 200 simulations). Further investigations should also be taken to investigate the transferability of this method to property predictions not included as optimization objectives. Relevant properties include diffusivity, viscosity, thermal conductivity, heat capacity, speed of sound, Joule−Thomson coefficient, compressibility, and expansivity. This workflow can also be extended to other complex molecules like hydrofluoroolefins and ionic liquids (ILs). For example, dihedral parameters of ILs that are initially not included in GAFF and hand-tuned before[67] can be optimized using this workflow.

ASSOCIATED CONTENT

Data Availability Statement

The code of this workflow is available to the public at https://github.com/dowlinglab/HFC-FFO. Example files for GROMACS and Cassandra and scripts of workflow.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00338.

Referenced experimental data; detailed values of intramolecular parameters; plots of VLE envelop, vapor pressure, and enthalpy of vaporization (PDF)

Detailed LJ parameters and VLE properties of all the nondominated parameter sets (XLSX)

AUTHOR INFORMATION

Corresponding Author

Edward J. Maginn − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*; orcid.org/0000-0002-6309-1347; Email: ed@nd.edu

Authors

Ning Wang − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*; orcid.org/0000-0002-5515-6198

Montana N. Carlozo − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*; orcid.org/0000-0003-0979-5703

Eliseo Marin-Rimoldi − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*

Bridgette J. Befort − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*; orcid.org/0000-0002-2340-6092

Alexander W. Dowling − *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*; orcid.org/0000-0001-9383-7499

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00338

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are thankful for the financial support from the National Science Foundation via two grants: EFRI DChem: Next-generation Low Global Warming Refrigerants, Award No. 2029354 and Collaborative Research: Development and Application of a Molecular and Process Design Framework for the Separation of Hydrofluorocarbon Mixtures, Award No. CBET-1917474. Computing resources were provided by the Center for Research Computing (CRC) at the University of Notre Dame. We also thank the Shiflett group from The University of Kansas for their collaboration. M.N.C. acknowledges support from the Graduate Assistance in Areas of

National Need fellowship from the Department of Education, Grant No. P200A210048.

REFERENCES

(1) Raabe, G. Molecular Simulation Studies on Refrigerants Past − Present − Future. *Fluid Phase Equilib.* 2019, *485*, 190−198.

(2) Purohit, P.; Höglund-Isaksson, L.; Wagner, F. *Impacts of the Kigali Amendment to Phase-Down Hydrofluorocarbons (HFCs) in Asia*; IIASA, 2018.

(3) Baca, K. R.; Olsen, G. M.; Matamoros Valenciano, L.; Bennett, M. G.; Haggard, D. M.; Befort, B. J.; Garciadiego, A.; Dowling, A. W.; Maginn, E. J.; Shiflett, M. B. Phase Equilibria and Diffusivities of HFC-32 and HFC-125 in Ionic Liquids for the Separation of R-410A. *ACS Sustainable Chem. Eng.* 2022, *10*, 816−830.

(4) Monjur, M. S.; Iftakher, A.; Hasan, M. M. F. Separation Process Synthesis for High-GWP Refrigerant Mixtures: Extractive Distillation Using Ionic Liquids. *Ind. Eng. Chem. Res.* 2022, *61*, 4390−4406.

(5) Morais, A. R. C.; Harders, A. N.; Baca, K. R.; Olsen, G. M.; Befort, B. J.; Dowling, A. W.; Maginn, E. J.; Shiflett, M. B. Phase Equilibria, Diffusivities, and Equation of State Modeling of HFC-32 and HFC-125 in Imidazolium-Based Ionic Liquids for the Separation of R-410A. *Ind. Eng. Chem. Res.* 2020, *59*, 18222−18235.

(6) Finberg, E. A.; May, T. L.; Shiflett, M. B. Multicomponent Refrigerant Separation Using Extractive Distillation with Ionic Liquids. *Ind. Eng. Chem. Res.* 2022, *61*, 9795−9812.

(7) Pardo, F.; Gutiérrez-Hernández, S. V.; Zarca, G.; Urtiaga, A. Toward the Recycling of Low-GWP Hydrofluorocarbon/Hydrofluoroolefin Refrigerant Mixtures Using Composite Ionic Liquid−Polymer Membranes. *ACS Sustainable Chem. Eng.* 2021, *9*, 7012−7021.

(8) Alkhatib, I. I. I.; Albà, C. G.; Darwish, A. S.; Llovell, F.; Vega, L. F. Searching for Sustainable Refrigerants by Bridging Molecular Modeling with Machine Learning. *Ind. Eng. Chem. Res.* 2022, *61*, 7414−7429.

(9) Asensio-Delgado, S.; Pardo, F.; Zarca, G.; Urtiaga, A. Machine Learning For Predicting the Solubility of High-GWP Fluorinated Refrigerants in Ionic Liquids. *J. Mol. Liq.* 2022, *367*, 120472.

(10) Maginn, E. J. From Discovery to Data: What Must Happen for Molecular Simulation to Become a Mainstream Chemical Engineering Tool. *AIChE J.* 2009, *55*, 1304−1310.

(11) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press, 2002.

(12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 2004, *25*, 1157−1174.

(13) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* 1996, *118*, 11225.

(14) Lísal, M.; Vacek, V. Effective Potentials for Liquid Simulation of the Alternative Refrigerants HFC-134a (CF_3CH_2F) and HFC-125 (CF_3CHF_2). *Fluid Phase Equilib.* 1997, *127*, 83−102.

(15) Lísal, M.; Vacek, V. Molecular Dynamics Simulations of Fluorinated Ethanes. *Mol. Phys.* 1996, *87*, 167−187.

(16) Potter, S. C.; Tildesley, D. J.; Burgess, A.; Rogers, S. C. A Transferable Potential Model for the Liquid-Vapour Equilibria of Fluoromethanes. *Mol. Phys.* 1997, *92*, 825−834.

(17) Higashi, S.; Takada, A. Molecular Dynamics Study of Liquid CH_2F_2 (HFC-32). *Mol. Phys.* 1997, *92*, 641−650.

(18) Fermeglia, M.; Ferrone, M.; Pricl, S. Development of an All-Atoms Force Field from Ab Initio Calculations for Alternative Refrigerants. *Fluid Phase Equilib.* 2003, *210*, 105−116.

(19) Peguin, R. P. S.; Kamath, G.; Potoff, J. J.; da Rocha, S. R. P. All-Atom Force Field for the Prediction of Vapor-Liquid Equilibria and Interfacial Properties of HFA134a. *J. Phys. Chem. B* 2009, *113*, 178−187.

(20) Yang, Z.; Gong, M.; Dong, X.; Li, X.; Wu, J. Molecular Modeling and Simulation of Vapor−Liquid Equilibrium of the Refrigerant R152a and its Mixture R152a+R32. *Fluid Phase Equilib.* 2015, *394*, 93−100.

(21) Zhang, N.; Hu, P.; Chen, L.; Zhi, L. Molecular Modeling of Vapor-Liquid Equilibrium Properties of HFC-161 and its Mixture HFC-161+HFO-1234yf. *J. Mol. Liq.* 2020, *306*, 112896.

(22) Raabe, G. Molecular Dynamics Studies on Liquid-Phase Dynamics and Structures of Four Different Fluoropropenes and Their Binary Mixtures with R-32 and CO2. *J. Phys. Chem. B* 2014, *118*, 240−254.

(23) Raabe, G. Molecular Simulation Studies on the Vapor−Liquid Phase Equilibria of Binary Mixtures of R-1234yf and R-1234ze(E) with R-32 and CO_2. *J. Chem. Eng. Data* 2013, *58*, 1867−1873.

(24) Befort, B. J.; DeFever, R. S.; Tow, G. M.; Dowling, A. W.; Maginn, E. J. Machine Learning Directed Optimization of Classical Molecular Modeling Force Fields. *J. Chem. Inf. Model.* 2021, *61*, 4400−4414.

(25) Yu, Y.; Krämer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D. J.; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of Long-Range Dispersion. *J. Chem. Theory Comput.* 2021, *17*, 1562−1580.

(26) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* 2018, *14*, 6076−6092.

(27) Boothroyd, S.; Wang, L.-P.; Mobley, D. L.; Chodera, J. D.; Shirts, M. R. Open Force Field Evaluator: An Automated, Efficient, and Scalable Framework for the Estimation of Physical Properties from Molecular Simulation. *J. Chem. Theory Comput.* 2022, *18*, 3566−3576.

(28) Horton, J. T.; Boothroyd, S.; Wagner, J.; Mitchell, J. A.; Gokey, T.; Dotson, D. L.; Behara, P. K.; Ramaswamy, V. K.; Mackey, M.; Chodera, J. D.; Anwar, J.; Mobley, D. L.; Cole, D. J. Open Force Field BespokeFit: Automating Bespoke Torsion Parametrization at Scale. *J. Chem. Inf. Model.* 2022, *62*, 5622−5633.

(29) Madin, O. C.; Shirts, M. R. Using physical property surrogate models to perform multi-fidelity global optimization of force field parameters. *Digital Discovery* 2023, na DOI: 10.1039/D2DD00138A.

(30) Razi, M.; Narayan, A.; Kirby, R. M.; Bedrov, D. Force-Field Coefficient Optimization of Coarse-Grained Molecular Dynamics Models with a Small Computational Budget. *Comput. Mater. Sci.* 2020, *176*, 109518.

(31) Tran, A.; Tranchida, J.; Wildey, T.; Thompson, A. P. Multi-Fidelity Machine-Learning with Uncertainty Quantification and Bayesian Optimization for Materials Design: Application to Ternary Random Alloys. *J. Chem. Phys.* 2020, *153*, 074705.

(32) Bisbo, M. K.; Hammer, B. Efficient Global Structure Optimization with a Machine-Learned Surrogate Model. *Phys. Rev. Lett.* 2020, *124*, 086102.

(33) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D.; Roux, B. Optimized Lennard-Jones parameters for druglike small molecules. *J. Chem. Theory Comput.* 2018, *14*, 3121−3131.

(34) Paulechka, E.; Kroenlein, K.; Kazakov, A.; Frenkel, M. A Systematic Approach for Development of an OPLS-Like Force Field and its Application to Hydrofluorocarbons. *J. Phys. Chem. B* 2012, *116*, 14389−14397.

(35) MacKerell, A. D.; Karplus, M. Importance of Attractive van der Waals Contribution in Empirical Energy Function Models for the Heat of Vaporization of Polar Liquids. *J. Phys. Chem.* 1991, *95*, 10559−10560.

(36) Liang, G.; Bays, J. P.; Bowen, J. P. Ab Initio Calculations and Molecular Mechanics (MM3) Force Field Development for Sulfonamide and its Alkyl Derivatives. *J. Mol. Struct.: THEOCHEM* 1997, *401*, 165−179.

(37) Raabe, G. Molecular Simulation Studies in Hydrofluoroolefine (HFO) Working Fluids and Their Blends. *Sci. Technol. Built Environ* 2016, *22*, 1077−1089.

(38) Allinger, N. L.; Zhou, X.; Bergsma, J. Molecular Mechanics Parameters. *J. Mol. Struct.: THEOCHEM* 1994, *312*, 69−83.
(39) Norrby, P.-O.; Brandt, P. Deriving Force Field Parameters for Coordination Complexes. *J. Coord. Chem.* 2001, *212*, 79−109.
(40) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective Force Fields for Condensed Phase Systems from Ab Initio Molecular Dynamics Simulation: A New Method for Force-Matching. *J. Chem. Phys.* 2004, *120*, 10896−10913.
(41) Höller, J.; Bickert, P.; Schwartz, P.; Von Kurnatowski, M.; Kerber, J.; Kunzle, N.; Lorenz, H.-M.; Asprion, N.; Blagov, S.; Bortz, M. Parameter Estimation Strategies in Thermodynamics. *ChemEngineering* 2019, *3*, 56.
(42) Zahariev, F.; De Silva, N.; Gordon, M. S.; Windus, T. L.; Perez Garcia, M. ParFit: A Python-Based Object-Oriented Program for Fitting Molecular Mechanics Parameters to ab Initio Data. *J. Chem. Inf. Model* 2017, *57*, 391−396.
(43) Ivanov, M. V.; Talipov, M. R.; Timerghazin, Q. K. Genetic Algorithm Optimization of Point Charges in Force Field Development: Challenges and Insights. *J. Phys. Chem. A* 2015, *119*, 1422−1434.
(44) Jaramillo-Botero, A.; Naserifar, S.; Goddard, W. A. General Multiobjective Force Field Optimization Framework, with Application to Reactive Force Fields for Silicon Carbide. *J. Chem. Theory Comput.* 2014, *10*, 1426−1439.
(45) Norrby, P.-O.; Liljefors, T. Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data. *J. Comput. Chem.* 1998, *19*, 1146−1166.
(46) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* 1993, *97*, 10269−10280.
(47) Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M. A.; Savoie, B. M. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force Fields. *J. Chem. Inf. Model.* 2021, *61*, 5013−5027.
(48) Miettinen, K. *Nonlinear Multiobjective Optimization*; Springer: New York, 1998.
(49) Dowling, A. W.; Ruiz-Mercado, G.; Zavala, V. M. A Framework For Multi-Stakeholder Decision-Making and Conflict Resolution. *Comput. Chem. Eng.* 2016, *90*, 136−150.
(50) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 2015, *1*, 19−25.
(51) Martínez, J. M.; Martínez, L. Packing Optimization for Automated Generation of Complex System's Initial Configurations for Molecular Dynamics and Docking. *J. Comput. Chem.* 2003, *24*, 819−825.
(52) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* 2009, *30*, 2157−2164.
(53) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press, 1988.
(54) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* 2007, *126*, 014101.
(55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 1984, *81*, 3684−3690.
(56) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* 1981, *52*, 7182−7190.
(57) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 1995, *103*, 8577−8593.
(58) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* 2008, *4*, 116−122.
(59) Shah, J. K.; Marin-Rimoldi, E.; Mullen, R. G.; Keene, B. P.; Khan, S.; Paluch, A. S.; Rai, N.; Romanielo, L. L.; Rosch, T. W.; Yoo, B.; Maginn, E. J. Cassandra: An Open Source Monte Carlo Package for Molecular Simulation. *J. Comput. Chem.* 2017, *38*, 1727−1739.
(60) Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2011, *2*, 1−27.
(61) Gramacy, R. B. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*; CRC Press, 2020.
(62) Lemmon, E. W.; Bell, I.; Huber, M. L.; McLinden, M. O. *NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP*, Version 10.0; National Institute of Standards and Technology, 2018. https://www.nist.gov/srd/refprop.
(63) Lemmon, E. W.; Bell, I. H.; Huber, M. L.; McLinden, M. O. In *Thermophysical Properties of Fluid Systems*; Linstrom, P., Mallard, W., Eds.; National Institute of Standards and Technology, 2023.
(64) Zollweg, J. A.; Mulholland, G. W. On the Law of the Rectilinear Diameter. *J. Chem. Phys.* 1972, *57*, 1021−1025.
(65) Eggimann, B. L.; Sunnarborg, A. J.; Stern, H. D.; Bliss, A. P.; Siepmann, J. I. An online parameter and property database for the TraPPE force field. *Mol. Simul.* 2014, *40*, 101−105.
(66) Potoff, J. J.; Bernard-Brunel, D. A. Mie Potentials for Phase Equilibria Calculations: Application to Alkanes and Perfluoroalkanes. *J. Phys. Chem. B* 2009, *113*, 14725−14731.
(67) Wang, N.; Zhang, Y.; Maginn, E. J. Molecular Dynamics Study of the Ionic Liquid 1-n-Hexyl-3-methylimidazolium Tris-(pentafluoroethyl)trifluorophosphate ($[C_6C_1im][FAP]$): Force Field Development and the Effect of $[FAP]^-$ Isomer Content on Properties. *J. Ionic Liq.* 2022, *2*, 100040.

https://doi.org/10.1021/acs.jctc.3c00338
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX