Delft University of Technology

Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency by Combining Predictive Thermodynamics and Machine Learning

Fleitmann, Lorenz; Ackermann, Philipp; Schilling, Johannes; Kleinekorte, Johanna; Rittig, Jan G.; vom Lehn, Florian; Schweidtmann, Artur M.; Pitsch, Heinz; Leonhard, Kai

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Article

Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency by Combining Predictive Thermodynamics and Machine Learning

Lorenz Fleitmann,[∇] Philipp Ackermann,[∇] Johannes Schilling, Johanna Kleinekorte, Jan G. Rittig, Florian vom Lehn, Artur M. Schweidtmann, Heinz Pitsch, Kai Leonhard, Alexander Mitsos, André Bardow,* and Manuel Dahmen*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Co-design of alternative fuels and future spark-ignition (SI) engines allows very high engine efficiencies to be achieved. To tailor the fuel's molecular structure to the needs of SI engines with very high compression ratios, computer-aided molecular design (CAMD) of renewable fuels has received considerable attention over the past decade. To date, CAMD for fuels is typically performed by computationally screening the physicochemical properties of single molecules against property targets. However, achievable SI engine efficiency is the result of the combined effect of various fuel properties, and molecules should not be discarded because of individual unfavorable properties that can be compensated for. Therefore, we present an optimization-based fuel design method directly targeting SI engine efficiency as the objective function. Specifically, we employ an empirical model to assess the achievable relative engine efficiency increase compared to conventional RON95 gasoline for each candidate fuel as a function of fuel properties. For this purpose, we integrate the automated prediction of various fuel properties into the fuel design method: Thermodynamic properties are calculated by COSMO-RS; combustion properties, indicators for environment, health and safety, and synthesizability are predicted using machine learning models. The method is applied to design pure-component fuels and binary ethanol-containing fuel blends. The optimal pure-component fuel *tert*-butyl formate is predicted to yield a relative efficiency increase of approximately 8% and the optimal fuel blend with ethanol and 3,4-dimethyl-3-propan-2-yl-1-pentene of 19%.
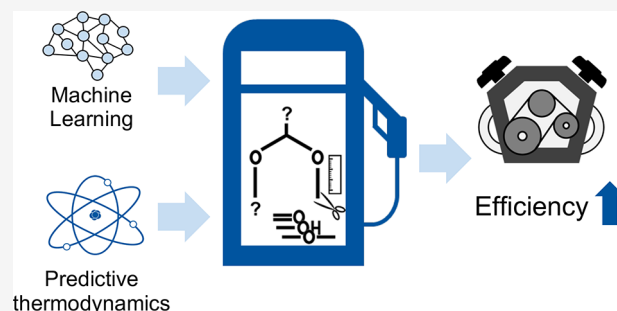
1. INTRODUCTION

The molecular structure of a fuel is a crucial degree of freedom for sustainable mobility.[1] This presents an opportunity to co-optimize internal combustion engines and fuel. Engine efficiencies can be increased significantly by increasing the compression ration or turbo charging,[2] as shown in laboratory experiments.[3−5] In advanced spark-ignition (SI) engine concepts, the achievable engine efficiency strongly depends on the autoignition tendency of the fuel—and thus the fuel's molecular structure or composition. Accordingly, alternative fuels have been found to vastly outperform RON95 gasoline in single-cylinder research engines.[3−8] The molecular structure of alternative, renewable fuels can be tailored in-silico by computer-aided molecular design (CAMD) methods, representing a special case of product design.[9,10] CAMD methods combine molecule databases or molecular structure generation with predictive models to assess the suitability of a candidate molecule for a given application based on physicochemical and thermodynamic properties.[11,12] To date, fuels are usually designed for individual physicochemical property targets as

surrogate measures for engine performance rather than the expected engine efficiency itself.[13]

Some studies on fuel design rely on database screenings using experimental data rather than models for property prediction: McCormick et al.[14] screened a database of approximately 500 potential biomass-based blendstocks and blends to identify feasible gasoline blends. To assess the candidates, experimental data were collected from various databases for physicochemical properties, environment, health and safety indicators, and corrosivity. Similarly, Fioroni et al.[15] screened a database for potential diesel blendstocks based on thermodynamic properties and cetane numbers. Using the database created by Fioroni et al.,[15] Huo et al.[16] and Huq et al.[17] evaluated chemocatalytic
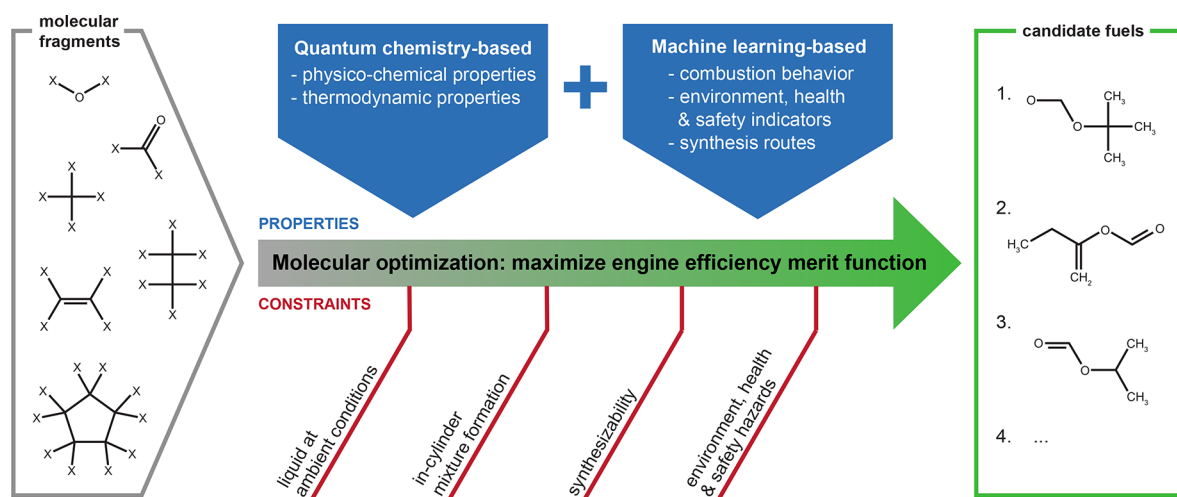
Figure 1. Fuel design via molecular optimization maximizing engine efficiency. Predictive quantum chemistry-based and machine learning-based models are used to calculate the properties for engine efficiency evaluation and to verify constraints on technical requirements, chemical synthesizability, and environment, health and safety hazards.

conversion pathways from potential biobased platform chemicals to hydrocarbons targeting physicochemical and combustion properties. Recently, Kuzhagaliyeva et al.[18] published a data-driven framework to design gasoline blends with tailored properties from a database of fuel molecules.

To expand the molecular design space beyond molecules contained in databases, generate-and-test approaches have been developed. The idea of generate-and-test CAMD is to create candidate structures first and assess their fuel properties subsequently by predictive models. Hechinger[19] employed the structure generator MOLGEN[20] and dedicated Quantitative Structure–Property Relationship (QSPR) models to predict physicochemical properties. The combustion performance of the candidate molecules was not assessed by a model-based approach but manually a posteriori because of limited training data.[21] Dahmen and Marquardt[22] later extended the generate-and-test approach by a group contribution method for the derived cetane number[23] to include combustion behavior. The authors tailored the structure generation to model catalytic refunctionalization of platform chemicals derived from lignocellulosic biomass. This approach yielded a list of candidate fuels that meet a range of fuel properties associated with high engine efficiency and high synthesizability.[22] Recently, Rittig et al.[24] employed generative graph machine learning models to design molecules with maximum research octane number and octane sensitivity.

Fuel design can also be formulated as a mathematical program as proposed by Gani,[25] which allows the use of deterministic optimization techniques to tailor the molecular structure of a fuel. The optimization-based approach was first applied to determine the composition of biofuel blends of preselected blend components respecting fuel standards[26,27] and was later extended to arbitrary components using a decomposed optimization strategy.[28] The integrated design of the molecular structures and their optimal composition in a blend was finally achieved by formulating and solving the blend design problem as a mixed integer nonlinear program based on functional groups as molecular building blocks.[29−31]

Optimization-based fuel design has attracted particular attention in combination with the selection of optimal conversion routes. Marvin et al.[32] used a rule-based reaction network generator to generate possible gasoline fuel compo-

nents and their production pathways. Based on the reaction network, gasoline blends were optimized with respect to production process performance constrained by fuel blend properties. In contrast, Ng et al.[33] designed biobased fuel blends by first optimizing the properties of a blend and then solving a superstructure optimization problem for an integrated biorefinery. Dahmen and Marquardt[34] combined blend design with mass-based screening of processing pathways using experimental yields to obtain renewable fuel blends maximizing resource efficiency. The method was extended by early stage process design using process network flux analysis, allowing the minimization of production cost and global warming impact of the designed fuel blend.[35] Subsequent engine testing of selected blends[34,36] has confirmed the superior engine performance compared to fossil gasoline.[4,5]

The studies mentioned above identified promising molecules and blends based on a list of target properties. However, such a fuel design does not consider the combined effect that the individual properties exert on engine performance. To date, no method has been proposed that explicitly designs fuels for maximum engine efficiency.

To consider engine efficiency as an explicit design objective, an engine model is required that predicts engine efficiency based on the fuel's physicochemical properties. Recently, two models for spark-ignition (SI) engines were presented in the literature: The first model is a zero-dimensional engine model,[37] developed to calculate a fuel's maximum engine efficiency considering knock limitation. The model has already been coupled with reaction network analysis to find the optimal upgrading of lignin pyrolysis oil.[38] Moreover, the model has been applied to a detailed performance evaluation of 50 preselected fuel candidates.[39] The second model is the engine efficiency merit function,[2,40] i.e., a correlation that predicts the relative engine efficiency increase compared to a base fuel, e.g., RON 95 gasoline, based on fuel properties, such as research octane number, octane sensitivity, and heat of vaporization. vom Lehn et al.[41] and Li et al.[42] have used the merit function to rank candidate fuels within database screenings. However, the screening studies are limited to existing database molecules and do not discover novel molecular structures.

In this work, we present an optimization-based method to design fuels for advanced highly boosted spark-ignition engines
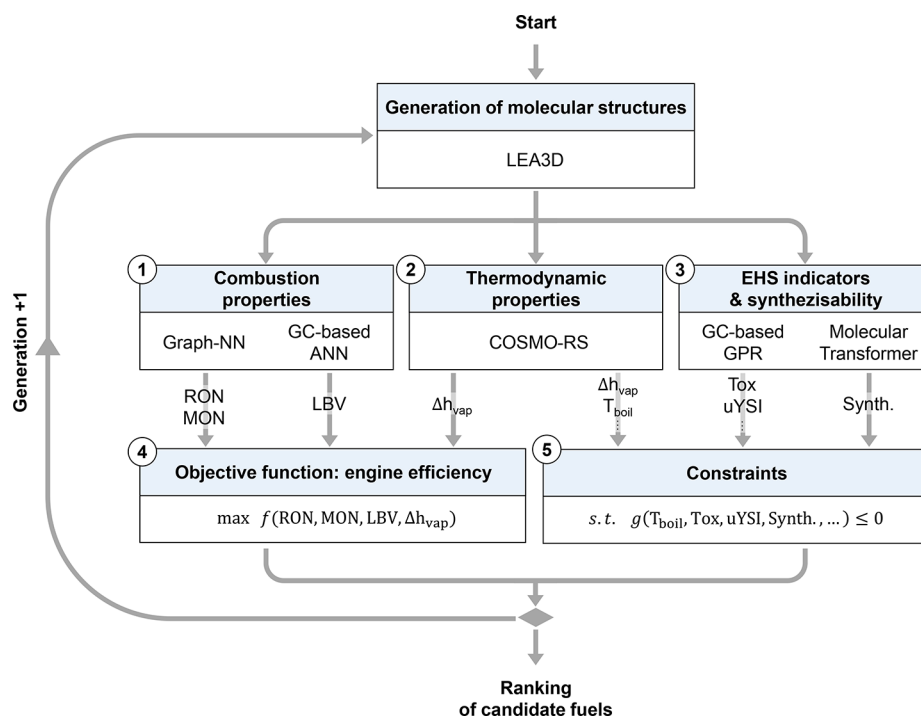
Figure 2. Fuel design method for maximum engine efficiency considering constraints on environment, health, and safety hazards, as well as synthesizability. For molecular optimization of candidate fuels, property prediction and performance assessment are integrated into the genetic algorithm LEA3D in five steps (1−5): prediction of combustion (1) and thermodynamic properties (2), and environment, health, and safety indicators including synthesizability (3), evaluation of objective function f (4) and constraints g (5).

with dedicated high compression ratios using the recently developed engine efficiency merit function[2,40] as an objective function (Figure 1). We evaluate engine efficiency as well as technical and practical feasibility of each candidate fuel by combining quantum chemistry-based predictive thermodynamics and machine learning models. For this purpose, we predict thermodynamic equilibrium and combustion properties as well as environment, health, and safety indicators and chemical synthesizability. In particular, we use dedicated models for the main determinants of SI engine fuel efficiency: research and motor octane numbers, laminar burning velocity, and enthalpy of vaporization. The capabilities of the method are demonstrated in the design of pure-component fuels and components of binary blends with ethanol.

The paper is structured as follows: In Section 2, the details of the fuel design method are explained. We outline the prediction models and methods and describe the fuel design constraints. In Section 3, the fuel design method is applied to the design of pure-component fuels (Section 3.1) and components of binary blends with ethanol (Section 3.2). In Section 4, we summarize our work and conclude with a brief outlook on future research.

2. FUEL DESIGN METHOD

Our fuel design method uses molecular optimization to maximize the predicted achievable engine efficiency increase of a fuel combusted in dedicated SI engines. Various constraints are imposed to design both efficient and safe fuels. Specifically, we predict thermodynamic and combustion properties as well as environmental, health, and safety (EHS) indicators and synthesizability.

For thermodynamic properties, established models from the literature use group contribution (GC)[43] or quantum chemistry-based methods,[44] e.g., COSMO-RS.[45] GC methods have also been applied to predict EHS indicators of chemicals.[46] Recently, advanced machine learning-based methods, e.g., deep neural networks and

Bayesian regression, have progressed rapidly in the field of molecular property prediction.[47] Therefore, our fuel design method is based on a hybrid approach for property prediction: Thermodynamic properties are predicted using quantum chemistry-based COSMO-RS, and combustion and EHS properties as well as synthesizability are predicted using machine learning-based models.

Based on the predicted properties, each candidate fuel is evaluated with the engine efficiency merit function. Both the property prediction and the evaluation of the objective function and constraints are integrated into a molecular optimization framework that is based on COSMO-CAMD.[48] COSMO-CAMD employs the genetic algorithm LEA3D,[49] which optimizes molecular structures using predefined 3D-molecular fragments through evolutionary optimization, and COSMO-RS for property prediction.[45]

The fuel design method involves five steps for candidate evaluation in each generation of the genetic optimization (Figure 2):

1. Prediction of combustion properties
2. Prediction of thermodynamic properties
3. Prediction of EHS indicators and chemical synthesizability
4. Objective function evaluation
5. Constraints evaluation

Based on the fitness values (i.e., objective function values) of the current generation of molecules, a next generation is created through the genetic operations crossover and mutation. The method proceeds to systematically explore the molecular design space until a predefined maximum number of generations is met.

In the following subsections, we briefly explain each step in the fuel design method. Details on the used soft- and hardware are included in the Supporting Information. The Supporting Information also contains the molecular fragments, which are specified as building blocks for the genetic algorithm. In this work, we include fragments to design oxygenated and nonoxygenated hydrocarbons.

2.1. Combustion Properties. Combustion properties that substantially influence SI engine efficiency are the research and the motor octane numbers (RON and MON), octane sensitivity (OS), as well as the laminar burning velocity (LBV).[2] We predict RON and

MON with our graph neural network (GNN) by Schweidtmann et al.[50] that was trained by simultaneous, so-called multitask learning on training data of RON, MON, and derived cetane number (DCN) of oxygenated and nonoxygenated hydrocarbons. By multitask learning, a higher prediction accuracy is achieved compared to the accuracy of single-task learning of the individual properties. The GNN directly uses the molecular graph and thus eliminates the need for manual feature selection. OS is calculated by subtracting MON from RON.[51]

The LBV is predicted with our group contribution (GC) based artificial neural network (ANN) that uses the occurrences of predefined structural groups in a molecule as input features.[52] The structural groups were originally proposed by Joback and Reid.[53] The LBV not only depends on the molecular structure but also on combustion parameters, i.e., temperature, pressure, and the fuel-air equivalence ratio, which are therefore additional inputs to the ANN. As suggested by Farrell et al.,[40] the LBV is evaluated for a stoichiometric mixture at ambient pressure and 358 K. It should be noted that these conditions differ from typical engine conditions. However, as Szybist et al.[2] point out, LBV measurements at engine-relevant conditions are associated with high uncertainties.

2.2. Thermodynamic Properties. Thermodynamic properties are predicted with COSMO-RS on the TZVP-MF level.[45] The TZVP-MF level offers a good balance between computational cost and accuracy for CAMD applications.[48] First, full geometry optimization by the quantum chemical density functional theory (DFT) is performed for each pure component based on its 3D-molecular structure, and the screening charge density profiles (σ-surfaces) are calculated. The BP86 functional and a TZVP basis set are employed on semiempirically generated conformers.[45] The time-consuming DFT calculations only need to be performed once for a molecule since the results are stored in a local database for reuse, e.g., in the next generation. Based on the interactions between the screening charges and statistical thermodynamics, COSMO-RS computes many thermodynamic properties of pure components and mixtures with low computational effort.

We use COSMO-RS to predict boiling/bubble points and enthalpies of vaporization of pure components and mixtures. Furthermore, melting points of pure components are available via a random forest-based QSPR model trained on structural molecular information and the σ-moment descriptors from COSMO-RS by Loschen and Klamt.[54] For fuel blends, COSMO-RS additionally calculates liquid–liquid-equilibria to estimate immiscibility and phase segregation.

2.3. Environment, Health, and Safety Indicators and Chemical Synthesizability. In addition to technical performance and technical feasibility, the candidate fuel needs to be assessed with regard to practical constraints as an optimally designed fuel should allow for safe handling and minimum hazards to environment and health.[5] Therefore, we include the prediction of environment, health, and safety (EHS) indicators to design nonhazardous fuels. Furthermore, some computer-designed molecules are chemically feasible but challenging or even impossible to synthesize in practice.[55] To exclude such compounds, we include a preliminary assessment of chemical synthesizability based on machine learning.

2.3.1. EHS Properties. Alshehri et al.[56] recently presented models for predicting various EHS indicators of pure components using group contribution (GC) based Gaussian process regression (GPR). A similar approach was presented by Li et al.[57] for the prediction of sooting tendencies. We follow the approaches by Alshehri et al.[56] and Li et al.[57] and consider the following EHS indicators through GC-based GPR prediction as constraints for the fuel design:

- Autoignition temperature (AiT)[58]
- Bioconcentration factor (BCF)[59]
- Aqueous toxicity as a lethal concentration for fathead minnow fish (LC$_{50}$(FM))[60]
- Oral toxicity as lethal dose for rats (LD$_{50}$)[61]
- Permissible exposure limit using the OSHA time-weighted average (PEL$_{OSHA-TWA}$)[62]
- Chemical tendency to form soot expressed through the unified yield sooting index (uYSI)[63]

For integration in the fuel design method, the models by Alshehri et al.[56] for AiT, BCF, LC$_{50}$(FM), LD$_{50}$, and PEL$_{OSHA-TWA}$ are retrained using UNIFAC groups as descriptors and the training and test data from Alshehri et al.[56] The uYSI model is developed using the data from McEnally et al.[64] Note that the uYSI does not predict engine-out soot emissions but rather the chemical tendency of a fuel to form soot. A more practical measure for engine-out soot emissions would be the Particulate Matter Index (PMI),[65] where the number of double bond equivalents as a proxy for the chemical tendency to form soot is divided by the vapor pressure as a measure for in-cylinder mixture formation quality. However, the different oxygenate functionalities of alternative fuels are differently effective in reducing soot formation.[66] Accordingly, comparing different soot indices, Leach et al.[67] found that the correlation between the number of double-bound equivalents and uYSI is stronger for hydrocarbon fuels than for oxygenated fuels. Unfortunately, no model is available for predicting nano soot number density, which is becoming more important in regulations.[68]

For the setup and accuracy assessment of the models, the training data for each model is split into a set for training and testing.[69] The test set contains approximately 10% of the training data and is not used within the training to assess the accuracy of the model on unseen data. The data are split so that the statistical distribution of the features in the test and training sets are similar.[69] A test set with a statistical distribution similar to the training set represents the model domain well and thus reflects model performance across the whole domain rather than just in a particular region. For this purpose, 10,000 random splits are performed, and the split with the lowest Kullback–Leibler divergence[70] is chosen, indicating the most similar and uniform statistical distribution between the training and test sets.

The models are set up by fragmenting the molecules contained in the training data into UNIFAC groups using the automated fragmentation tool by Müller.[71] The kernels for the GPR of each model are selected by employing the automated kernel-search algorithm developed by Duvenaud et al.,[72] Duvenaud,[73] and Lloyd et al.[74]

The models achieve an accuracy comparable to the models in the literature[46,56] (see Table 1). The accuracy of the predictions on the test

Table 1. Data Set Sizes N_{Train} and N_{Test} and Prediction Accuracies on the Test Sets of the EHS Indicators Using Group Contribution-Based GPR Models for the Categories Autoignition Temperature (AiT), Bioconcentration Factor (BCF), Aqueous Toxicity of Fathead Minnow Fish (LC$_{50}$(FM)), Oral Rat Toxicity (LD$_{50}$), Permissible Exposure Limit (PEL$_{OSHA-TWA}$), and Unified Yield Sooting Index (uYSI)[a]

EHS indicator	N_{Train}	N_{Test}	R^2_{test}	RMSE$_{test}$	nRMSE$_{test}$ (%)
AiT	487	54	0.78	58 K	6.4
log(BCF)	366	41	0.78	0.69	11
$-$log(LC$_{50}$(FM))	490	54	0.66	0.75	8.6
$-$log(LD$_{50}$)	2157	240	0.58	0.40	9.1
$-$log(PEL$_{OSHA-TWA}$)	346	38	0.60	1.1	13
uYSI	397	44	0.99	39	2.9
average performance	-	-	0.73	-	8.5

[a]The accuracy is measured by the coefficient of determination (R^2), the root-mean-square error (RMSE), and the root-mean-square error normalized by the range of values (nRMSE).

sets measured by the coefficient of determination (R^2) equals on average $R^2 = 0.73$. The corresponding root-mean-square error normalized by the range of values (nRMSE) is nRMSE = 8.5%. Parity plots of predicted and target values of the test sets visualizing prediction accuracy can be found in the Supporting Information.

2.3.2. Chemical Synthesizability. By assembling molecular fragments, the LEA3D algorithm generates molecules that always satisfy chemical feasibility, i.e., the octet rule. However, a chemically feasible molecule is not necessarily similar to known molecules and may

therefore be technologically challenging to obtain, i.e., hardly synthesizable or only synthesizable with considerable effort via numerous synthesis steps.

The chemical synthesizability of candidate molecules can be assessed through retrosynthesis algorithms,[55] which generate synthesis routes for a given product and thus allow an investigation into whether and how a chemical can be readily synthesized. Various retrosynthesis models have been developed in the past.[75] For the fuel design method, we employ a graph exploration algorithm for retrosynthesis developed by Schwaller et al.[76] The algorithm is based on the molecular transformer, a multihead attention-based neural network model for forward synthesis prediction with high accuracy.[77]

In this work, we consider a fuel synthesizable if a maximum of three subsequent reactions are required from commercially available reactants to the desired fuel. Otherwise, the synthesis route is deemed too costly to be viable. Moreover, the confidence of the retrosynthesis algorithm in the synthesis route has to be greater than 50%.

Our definition of chemical synthesizability discards candidate fuels for which efficient synthesis routes are likely unknown. However, we note that our assessment of synthesizability does not ensure that the molecules considered synthesizable can be produced in large quantities, at low cost, or from renewable resources, which should also be a design target for a novel fuel but cannot yet be predicted.

2.4. Engine Efficiency Merit Function as an Objective Function. Various fuel properties can have a positive impact on achievable SI engine efficiency, including RON, OS, Δh_{vap}, and LBV.[2,40] Alternative fuels can exhibit favorable values in one or more of these properties. To evaluate the potential of an alternative fuel candidate for use in advanced highly boosted SI engines, the impact of these fuel properties on efficiency has been empirically quantified through experimental sensitivity analyses under stoichiometric boosted combustion conditions, resulting in the so-called *engine efficiency merit function*.[2,40] The merit function linearly correlates the fuel properties with the achievable improvement in maximum brake thermal engine efficiency (η) compared to a reference fuel, e.g., RON95 gasoline.

Using the merit function, we calculate the expected relative engine efficiency increase of a candidate fuel based on RON, MON, Δh_{vap}, and LBV compared to RON95 gasoline:[2,40]

$$
\begin{aligned}
\frac{merit}{100\%} &= \frac{\eta - \eta_{ref}}{\eta_{ref}} \\
&= \underbrace{\frac{RON - RON_{ref}}{1.6}}_{\text{octane number}} - K \cdot \underbrace{\frac{RON - MON - (RON_{ref} - MON_{ref})}{1.6}}_{\text{octane sensitivity}} \\
&+ 0.0085 \cdot \underbrace{\frac{\frac{\Delta h_{vap}}{AFR + 1} - \frac{\Delta h_{vap,ref}}{AFR_{ref} + 1}}{1.6}}_{\text{effective octane rating}} + \underbrace{\frac{\frac{\Delta h_{vap}}{AFR + 1} - \frac{\Delta h_{vap,ref}}{AFR_{ref} + 1}}{15.2}}_{\text{charge cooling}} \\
&+ \underbrace{\frac{LBV - LBV_{ref}}{5.4}}_{\text{laminar burning velocity}}
\end{aligned}
\tag{1}
$$

Each term in the merit function reflects an empirically found linear influence on maximum engine efficiency:[2,40] the influence of octane number, octane sensitivity, effective octane rating, charge cooling, and laminar burning velocity. The reference values for RON, MON, and Δh_{vap} are taken from Leitner et al.[1] (Table 2). Lacking a value for LBV from Leitner et al.,[1] we use the LBV of a commercial RON95 measured by Dirrenberger et

al.[78] as a reference value (Table 2). The stoichiometric air-to-fuel ratio of each candidate fuel is denoted by AFR, and the parameter K is a normalized value describing the engine's operating conditions relative to those of the RON and MON tests, i.e., $K_{RON} = 0$ and $K_{MON} = 1$, respectively.[51]

Since the K value depends on engine operating conditions, selecting a single, representative value for our screening is difficult. For modern downsized, turbocharged SI engines, the K parameter is usually negative.[79−81] Kassai et al.[80] determined values in a single-cylinder research engine with a moderate compression ratio of 10.5 between −0.1 and −1.9, with K being the lowest at high intake pressure and low engine speed. Since we aim at fuels for engines with higher compression ratios (e.g., 16.4 in our recent work[5]), we pragmatically chose $K = -1.5$. In the results section, we also analyze how the choice of the K value affects the predicted efficiency gains.

2.5. Property Constraints. The candidate fuels need to meet several thermodynamic, environmental, and practical requirements that are formulated as design constraints (Table 3). The constraints on thermodynamic properties are taken from a previous design study by Dahmen and Marquardt.[22] The normal boiling and melting points of the candidate fuels are constrained to ensure that the fuel is liquid at ambient conditions. Additional constraints on the maximum boiling point and the maximum enthalpy of vaporization ensure sufficient volatility and, thus, proper in-cylinder mixture formation under cold conditions. If the boiling point or the enthalpy of vaporization of a candidate fuel is too high, the candidate fuel may not completely evaporate under cold conditions but dissolve in the engine oil potentially causing engine failure due to oil dilution.[82−84] Note that today's fossil fuel standards use the Reid vapor pressure and characteristic points on the distillation curve to address cold start issues and neglect the enthalpy of vaporization.[85,86] However, studies on pure alcohol fuels such as ethanol, 1-butanol, and 2-butanol have linked high enthalpies of vaporization to higher pollutant formation[87,88] and excessive oil dilution.[84] Further research is needed to better define appropriate upper limits on the boiling point and enthalpy of vaporization in the case of pure-component alternative fuels.

The candidate fuels' EHS indicators are constrained to ensure that the candidate fuels are less hazardous than RON95 gasoline with respect to AiT, $LC_{50}(FM)$, LD_{50}, and $PEL_{OSHA-TWA}$. For BCF, candidate fuels must not be bioaccumulative, i.e., $\log(BCF) < 3.3$ according to Arnot and Gobas.[59] The sooting tendency expressed through the uYSI is not restricted by regulations or policy but should be as small as possible for clean and efficient combustion.[2] A strict upper bound is difficult to define. In this work, the uYSI value of *n*-hexane is considered acceptable and therefore chosen as an upper bound.

Since the EHS indicators are predicted with Gaussian Process Regression (GPR), predicted values are provided with uncertainty quantification. This prediction uncertainty is considered in chance constraints to minimize the number of incorrectly discarded fuel candidates. Candidate fuels are only discarded if a property's 95% confidence interval violates a constraint, i.e., if a constraint's lower bound lb > $\Omega_{max} = \Omega + 1.96 \cdot \sigma_{\Omega}$, or if a constraint's upper bound ub < $\Omega_{min} = \Omega - 1.96 \cdot \sigma_{\Omega}$, where Ω is the considered property and σ_{Ω} is the property's prediction uncertainty.

2.6. Evaluation of Mixture Properties. For the design of binary blends with ethanol (cf. Section 3.2), mixture properties have to be predicted. Mixture bubble points and mixture enthalpies of vaporization are calculated considering nonideal thermodynamic behavior using COSMO-RS. The prediction of nonideal behavior is a particular strength of COSMO-RS. Pragmatically, we constrain the bubble point temperature of the mixture to the same value (120 °C) as the normal boiling point used in the pure-component design. Note that mixtures of components with strongly different evaporation characteristics may exhibit preferential evaporation, with in-cylinder mixture inhomogeneity, potentially causing wall wetting and oil dilution.[90−92] Consideration of preferential evaporation effects is, however, beyond the scope of this work. With the given data-driven models, the nonideal mixture behavior for the combustion properties and EHS indicators

Table 2. Reference Values for RON95 Gasoline Used in the Engine Efficiency Merit Function

property	reference value	reference
RON_{ref}	96	Leitner et al.[1]
MON_{ref}	85	Leitner et al.[1]
$\Delta h_{vap,ref}$	350 kJ kg^{-1}	Leitner et al.[1]
AFR_{ref}	14	Leitner et al.[1]
LBV_{ref}	48 cm s^{-1}	Dirrenberger et al.[78]

Table 3. Property Constraints for the Fuel Design Used in This Work[a]

property	constraint	reason	reference
T_{melt}	$T_{melt} \leq -20$ °C	liquid fuel	Dahmen and Marquardt[22]
$T_{boil/bubble}$	$T_{boil/bubble} \geq 60$ °C	liquid fuel	Dahmen and Marquardt[22]
	$T_{boil/bubble} \leq 120$ °C	in-cylinder mixture formation	Dahmen and Marquardt[22]
Δh_{vap}	$\Delta h_{vap} \leq 60 \mathrm{kJ\ kg_{air}^{-1}}$	in-cylinder mixture formation	Dahmen and Marquardt[22]
AiT	AiT ≥ 553.15 °C	RON95 gasoline value	BP Europa[89]
BCF	$\log(\text{BCF}) \leq 3.3$	REACH administrative	Arnot and Gobas[59]
$\mathrm{LC_{50}(FM)}$	$-\log(\mathrm{LC_{50}(FM)}) \leq 4$	RON95 gasoline value	BP Europa[89]
$\mathrm{LD_{50}}$	$-\log(\mathrm{LD_{50}}) \leq 1.3$	RON95 gasoline value	BP Europa[89]
$\mathrm{PEL_{OSHA-TWA}}$	$-\log(\mathrm{PEL_{OSHA-TWA}}) \leq 1.96$	RON95 gasoline value	BP Europa[89]
uYSI	uYSI ≤ 30	n-hexane value	Das et al.[63]

[a]T_{melt}: melting point temperature; $T_{boil/bubble}$: boiling/bubble point temperature; Δh_{vap}: enthalpy of vaporization; AiT: autoignition temperature; BCF: bioconcentration factor; $\mathrm{LC_{50}(FM)}$: aqueous toxicity of fathead minnow fish; $\mathrm{LD_{50}}$: oral rat toxicity; $\mathrm{PEL_{OSHA-TWA}}$: permissible exposure limit; uYSI: unified yield sooting index.
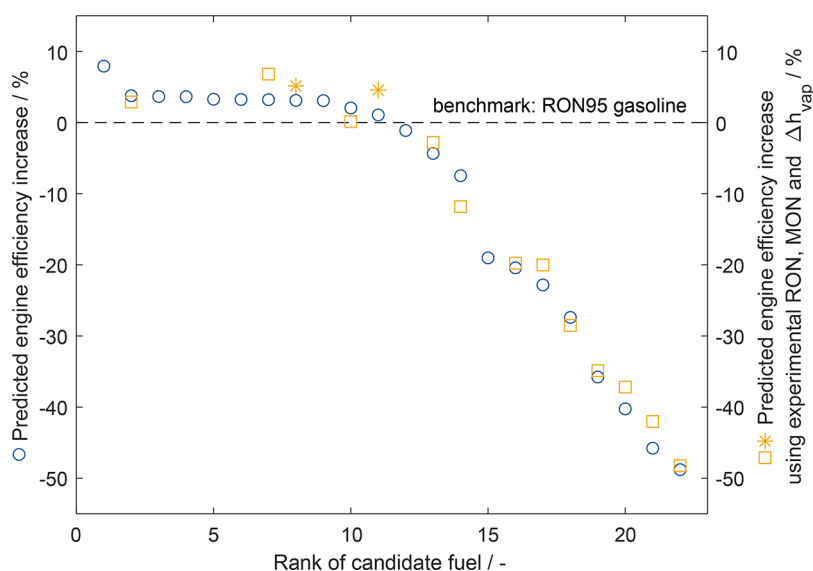

Figure 3. Predicted relative engine efficiency increase for the pure-component design. The blue circles represent the model predictions for the 22 candidate fuels. The orange squares and asterisks are calculated using experimental values for RON, MON, and Δh_{vap}. The experimental values of the molecules marked with asterisks were not used for training of the GNN.

cannot be predicted. In the absence of more accurate, nonlinear models, we use a linear molar mixing rule for these mixture properties:

$$\Omega_{blend} = \sum_{i=1}^{n} x_i \Omega_i \tag{2}$$

In this equation, Ω_i stands for the predicted pure-component properties, and x_i is the mole fraction of component i in the blend. The mixture property is denoted by Ω_{blend}.

For combustion properties, the linear molar mixing rule approximates nonideal behavior more accurately than, e.g., a linear liquid volume-based mixing rule, in particular for blends with ethanol.[41] For the EHS indicators, the linear molar mixing rule is in line with previous blend design studies,[28,29] following the concept of dose addition for toxicity ($\mathrm{LC_{50}(FM)}$, $\mathrm{LD_{50}}$)[93] and exposure hazards ($\mathrm{PEL_{OSHA-TWA}}$).[94]

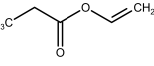## 3. DESIGN OF PURE FUELS AND FUEL BLENDS FOR SPARK-IGNITION ENGINES
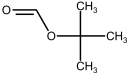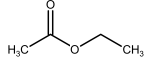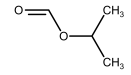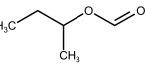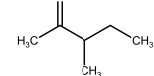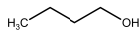
We apply the fuel design method to design (1) pure-component fuels and (2) blend components for binary fuel blends with ethanol.

3.1. Pure-Component Fuel Design. The fuel design method is started twice with 50 generations and 40 candidate molecules per generation to accommodate for the stochastic

nature of the approach. In total, the method investigates 1,033 unique molecules in approximately 3 days and 9 h in parallel on 24 computer cores (see Supporting Information for details on the hardware). Of these 1,033 unique molecules, 22 are candidate fuel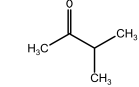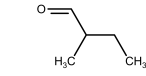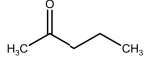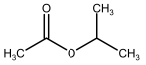s that fulfill all constraints (Figure 3). Eleven candidate fuels outperform the benchmark RON95 gasoline in predicted engine efficiency (Table 4). As the optimal fuel, we identify *tert*-butyl formate with a predicted increase in engine efficiency of approximately 7.9%, followed by ethyl acetate (3.8%), isopropyl formate (3.7%), and vinyl propionate (3.6%). The remaining 6 candidate fuels increasing engine efficiency achieve only minor improvements in predicted engine efficiency between 1% and 3.3%.

From the 22 identified candidate fuels, 14 candidate fuels have been considered in the database screening by vom Lehn et al.[41] that relied on experimental property data. We use this experimental data to recalculate the relative engine efficiency increase and compare the results with the predicted relative engine efficiency increase. The mean absolute error (MAE) of the predicted engine efficiency increase is only 2.2%, indicating an accurate assessment by the fuel design method (cf. Figure 3). However, 12 of these 14 candidate fuels were also included in the training data set of the GNN that contributes RON and

Table 4. Details on the Pure-Component Fuel Design for the 11 Candidate Fuels Predicted to Exceed the Engine Efficiency of RON95 Gasoline

Molecule	pure-component properties					Engine efficiency increase
	T_{boil} / K	RON / -	MON / -	Δh_{vap} / kJ kg$_{air, \lambda=1}^{-1}$	LBV / cm s^{-1}	
tert-butyl formate	371.6	115.5	108.6	37.5	38.6	7.92 %
ethyl acetate	372.0	116.7	116.7	49.1	41.2	3.77 %
isopropyl formate	357.7	113.1	110.9	47.3	42.3	3.65 %
vinyl propionate	379.9	111.3	108.3	42.7	46.9	3.63 %
sec-butyl formate	378.9	108.6	103.1	39.1	43.0	3.27 %
3-methyl-2-pentanone	390.5	104.6	96.4	30.7	47.8	3.24 %
1-butanol	387.7	97.8	88.5	54.9	50.1	3.21 %
3-methyl-2-butanone	368.9	104.7	96.9	34.9	46.5	3.11 %
2-methylbutyraldehyde	351.3	103.3	95.2	33.3	50.2	3.08 %
2-pentanone	372.2	106.7	102.2	35.5	49.5	2.04 %
isopropyl acetate	391.8	115.1	115.5	40.2	38.6	1.06 %

MON values to the engine efficiency assessment,[50] leading to a high prediction accuracy of these candidates. Nevertheless, the MAE for 3-methyl-2-butanone and isopropyl acetate, which were not included in the training data set, is comparable with MAE = 2.8% to the MAE of the training data (MAE$_{train}$ = 2.1%), indicating generalizability beyond the training data of the GNN.

Using the experimental data, 1-butanol (rank 7, *merit*$_{exp}$ = 6.8%), 3-methyl-2-butanone (rank 8, *merit*$_{exp}$ = 5.1%), and isopropyl acetate (rank 11, *merit*$_{exp}$ = 4.6%) are highly promising candidate fuels achieving higher engine efficiency according to

the engine efficiency merit function than predicted during the design. Thus, the fuel design method suggests promising candidate fuels. Still, the final ranking requires subsequent experimental verification.

In contrast to the present study, vom Lehn et al.[41] identified methanol, methyl formate, and ethanol as the highest-ranking candidate fuels. However, under the constraints of the present study, methanol and ethanol exceed the maximum permissible heat of vaporization with predicted values of 214 kJ kg$_{air}^{-1}$ (experimental[95] 183 kJ kg$_{air}^{-1}$) and 106 kJ kg$_{air}^{-1}$ (experimental[95]
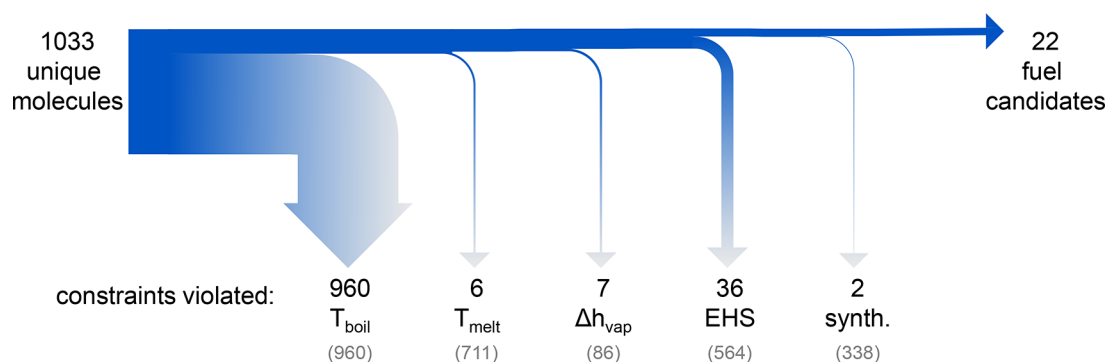
Figure 4. Influence of property constraints on the number of candidate fuels in the pure-component fuel design. The gray numbers in brackets are the total number of candidate fuels violating the corresponding constraint if not already discarded by previously applied constraints.

93 kJ kg_{air}^{-1}), respectively. Methyl formate violates the lower bound on the boiling point with a predicted boiling temperature of 38 °C (experimental[95] 32 °C). Formates with longer alkyl chains and higher boiling points have been identified as promising pure-component fuels in the present study as well as other esters. In particular, ethyl acetate and isopropyl acetate are suggested as pure-component fuels by the present study, vom Lehn et al.,[41] and Dahmen and Marquardt.[22] Moreover, 3-methyl-2-pentanone and 2-pentanone have been proposed by Dahmen and Marquardt[22] as blend candidates with moderate knock resistance, which is confirmed by the predicted moderate pure-component relative engine efficiency increase found in the present study (3.2% and 2.0%, respectively). 1-Butanol and its isomers have also been studied in the literature as 1-butanol has similar knock resistance to RON95 and is known to fulfill the property constraints for an SI fuel.[22] The fuel design in this work is thus confirmed by findings from the literature but also yields additional candidate fuels that have not yet been investigated.

3.1.1. Influence of Property Constraints on Optimal Fuel Candidates. In the design of pure-component fuels for SI engines, the tight property constraints drastically limits the number of candidate molecules, as also discovered by Dahmen and Marquardt.[22] In the present study, 1011 of 1033 candidates are already discarded before the engine efficiency evaluation due to constraint violation (Figure 4). The most selective constraint is the constraint on the boiling point, which is also enforced first: 960 of 1011 excluded candidate fuels violate the constraint, with 929 of 960 candidates exceeding the maximum boiling point. The subsequently applied constraints on melting point, enthalpy of vaporization, EHS indicators, and synthesizability discard additional 6, 7, 36, and 2 molecules, respectively. The majority of candidate molecules not only violate the boiling point constraint but also the melting point temperature, EHS, or synthesizability constraints (gray numbers in Figure 4).

We relax the constraints on boiling point and enthalpy of vaporization to investigate whether technical advances in engine design could substantially increase the number of feasible candidate fuels. A relaxation of the constraints on T_{boil} and Δh_{vap} by 10 K and 10 kJ kg_{air}^{-1} compared to the original values (cf. Table 3) yields three additional fuel candidates: propen-2-ol (relative engine efficiency increase 12%, Δh_{vap} = 69 kJ kg_{air}^{-1}), 2-propanol (relative engine efficiency increase 10%, Δh_{vap} = 69 kJ kg_{air}^{-1}), and 2-hydroxy-2-methylpropanal (relative engine efficiency increase 8.5%, T_{boil} = 120.2 °C). Further relaxation of the constraints by 10 K and 10 kJ kg_{air}^{-1} to 140 °C and 80 kJ kg_{air}^{-1}, respectively, additionally yields 1-propanol (relative engine efficiency increase 12%, %, Δh_{vap} = 72 kJ kg_{air}^{-1}).

Moreover, various highly branched alkenes and alkanes are designed under the relaxed property constraints, e.g., 3,3,4-trimethyl-1-pentene, 4,4,5-trimethyl-1-hexene, or 2,3,3,4-tetra-methylpentane. The designed highly branched alkenes and alkanes are predicted to exhibit high RONs (104−110) and high octane sensitivities (12−16), leading to predicted relative engine efficiency increases between 8.5% and 11%. However, high uYSI values (74−84) are predicted for these molecules, which are only considered fuel candidates by the algorithm as their 95% confidence intervals reach below the threshold value of uYSI = 30. Since highly branched alkenes and alkanes are known to cause soot formation, the actual values of the corresponding molecules are likely in the area of uYSI = 70−80, as also suggested by the online uYSI estimator by St. John et al.[96]

In conclusion, pure-component fuel design for high-efficiency SI engines is extremely challenging as it yields only a limited number of candidates. Moreover, relaxations in property constraints only result in a few additional candidate fuels.

3.2. Design of Binary Blends with Ethanol. Unfavorable properties of individual molecules can often be balanced by blending.[34−36] Therefore, in fuel design, blending constitutes an additional degree of freedom that can be used to broaden the number of candidate fuels. In this section, we apply the fuel design method to binary blends with ethanol.

Ethanol is an established blend component for commercial fuels.[1] It is known for its excellent knock resistance (high RON) and its high enthalpy of vaporization that provides a charge cooling effect. Both properties increase engine efficiency,[97] leading to a predicted relative engine efficiency increase by the merit function of 25%. However, as a pure-component fuel, ethanol is troublesome as its high enthalpy of vaporization (cf. Section 3.1) can lead to oil dilution under cold conditions.[82−84] Therefore, ethanol is a suitable base component for a blend due to its favorable combustion properties but must be balanced with a tailor-made secondary component.

For the purpose of blend design, we optimize both the molecular structure and the molar fraction of the candidate blend component in a binary blend with ethanol. We run the fuel design method twice for 50 generations with 40 candidate molecules per generation. In total, the method investigates 1,310 unique blends in 9 days and 20 h in parallel on 24 computer cores (see Supporting Information for details on the hardware). From these unique blends, 226 fulfill the property constraints and are thus candidate fuel blends (see Supporting Information for details). 184 candidate fuel blends lead to a positive merit function value indicating a relative engine efficiency increase compared to RON95 gasoline (Figure 5). Compared to the
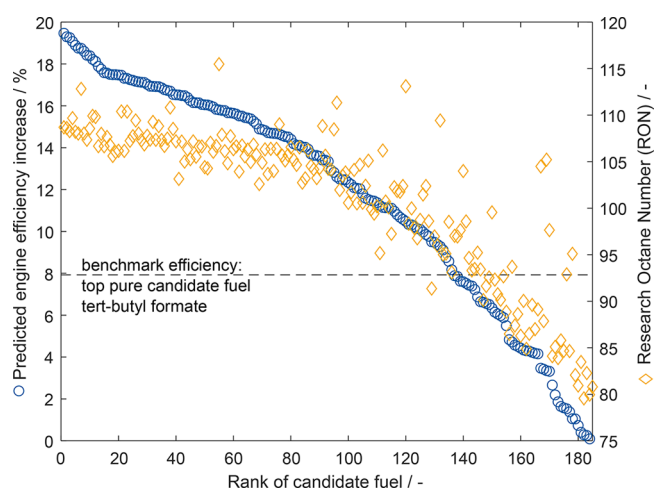
Figure 5. Predicted relative engine efficiency increase for the 184 binary blends with ethanol. The blue circles represent the predictions from the predictive models. The orange diamonds are the predicted blend RONs.
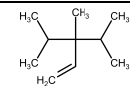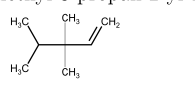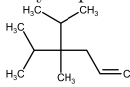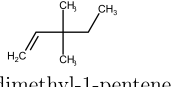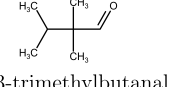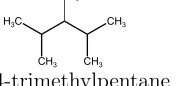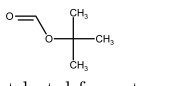
optimal pure-component fuel *tert*-butyl formate, 136 fuel blends yield a higher engine efficiency. Therefore, as Figures 4 and 5 demonstrate, the design space for fuel blends is much larger than the design space for pure-component fuels and offers many more possibilities for increasing engine efficiency.

The optimal fuel blend with ethanol is obtained with 3,4-dimethyl-3-propan-2-yl-1-pentene, a highly branched alkene. The relative engine efficiency increase of this optimal blend is predicted to be 19.5% (Table 5). The majority of the blend is composed of ethanol (83 mol %) and only 17 mol % of 3,4-dimethyl-3-propan-2-yl-1-pentene. The blends with the second and third highest engine efficiency increases are formed with 22 mol % 3,3,4-trimethyl-1-pentene, and with 16 mol % 4,5-dimethyl-4-propan-2-yl-1-hexene, again highly branched alkenes predicted to relatively increase engine efficiency by approximately 19.3%.

The top candidates combine a high RON (104−107) and a high octane sensitivity (15−16), as this combination is known to enable high engine efficiency in modern highly boosted engines.[80] We observe that alkenes are generally high-ranking blend components: 18 of the top 50 candidate blends are formed with an alkene and 1 with an alkadiene. The positive effect of the vinyl groups on engine efficiency is known from the literature.[41]

The top 3 blend components are not suitable as pure-component fuels, e.g., due to constraint violations on boiling point, toxicity, or soot formation. Since the top 3 candidates are large alkenes (C8−C11), they have high boiling points of 129 °C, 91 °C, and 145 °C, and high sooting tendencies as evident by predicted uYSI values of between 75 and 107. Moreover, the components are highly toxic to aqueous organisms as indicated by a predicted $-\log(LC_{50}(FM)) = 4.0-4.8$ and potentially have a low exposure limit (predicted $-\log(PEL_{OSHA-TWA}) = 3.3-3.5$).

Table 5. Blend Design Results: The Three Highest-Ranking Candidate Blend Components, the Highest-Ranking Commercially Available Blend Components, and the Optimal Pure-Component Fuel *tert*-Butyl Formate[a]

Molecular structure	Blend properties		Relative engine efficiency increase		Molar fraction in blend x_{blend}	Rank in design	
	RON_{blend}	$T_{bubble,blend}$	pure comp.	blend		eng. eff.	RON_{blend}
3,4-dimethyl-3-propan-2-yl-1-pentene	108.7	352.7 K	10.2 %	19.5 %	0.175	1	17
3,3,4-trimethyl-1-pentene	108.6	346.2 K	10.9 %	19.3 %	0.222	2	18
4,5-dimethyl-4-propan-2-yl-1-hexene	108.3	353.8 K	8.28 %	19.3 %	0.159	3	22
3,3-dimethyl-1-pentene	108.0	337.5 K	9.18 %	18.6 %	0.261	8	25
2,2,3-trimethylbutanal	109.2	355.4 K	8.38 %	18.1 %	0.275	12	9
2,3,4-trimethylpentane	107.7	344.2 K	6.42 %	17.2 %	0.214	25	31
tert-butyl formate	112.0	354.3 K	7.92 %	15.8 %	0.453	55	1

[a]The predicted pure-component relative engine efficiency increase is calculated to show the effect of blend design. Note that except for *tert*-butyl formate, the components do not satisfy all property constraints as pure components.

In a blend with ethanol, ethanol is predicted to compensate for these properties, while the low enthalpies of vaporization of the top blend components ($18-21$ kJ kg_{air}^{-1}) help to mitigate ethanol's high enthalpy of vaporization.

The top 3 blend components are synthesizable as indicated by retrosynthesis through 1 or 2 reactions from commercially available components. However, they are currently not commercially available themselves as determined by database searches. The highest-ranking commercially available component is 3,3-dimethyl-1-pentene on rank 8. An optimized blend of 26 mol % 3,3-dimethyl-1-pentene and 74 mol % ethanol increases engine efficiency by approximately 18.6% compared to RON95. Similar to the top 3 blend components, 3,3-dimethyl-1-pentene has a high RON of 105 and an octane sensitivity of 15. The predicted EHS indicators suggest similar toxicity ($-\log(LC_{50}(FM)) = 3.8, -\log(LD_{50}) = 1.6$) and permissible exposure limit ($-\log(PEL_{OSHA\text{-}TWA}) = 3.8$), but lower soot formation (uYSI = 57).

A commercially available alternative to 3,3-dimethyl-1-pentene is 2,2,3-trimethylbutanal on rank 12. A blend of 28 mol % 2,2,3-trimethylbutanal and 72 mol % ethanol is predicted to increase engine efficiency by 18%. Compared to 3,3-dimethyl-1-pentene, 2,2,3-trimethylbutanal has the advantage of not containing a vinyl group but instead contains an aldehyde group. However, molecules with vinyl or aldehyde groups can both age the fuel blend and reduce its stability due to polymerization[98] and high reactivity.[99] The highest-ranking commercially available candidate blend component without a vinyl or aldehyde group is 2,3,4-trimethylpentane on rank 25. A blend of 21 mol % 2,3,4-trimethylpentane and 79 mol % ethanol is predicted to increase engine efficiency by 17%. In comparison to the top 3 and the high-ranking commercially available blend components, the predicted EHS indicators of 2,3,4-trimethylpentane suggest safer handling and use with a higher permissible exposure limit ($-\log(PEL_{OSHA\text{-}TWA}) = 1.9$) and lower toxicity ($-\log(LC_{50}(FM)) = 3.6, -\log(LD_{50}) = 1.4$). Therefore, 2,3,4-trimethylpentane could be the most promising commercially available blend component in this design, highlighting the additional requirements for the practical selection of a promising fuel besides engine efficiency.

3.2.1. Comparison between Ranking by RON and by Engine Efficiency Increase. We challenge the blend design maximizing the merit function of engine efficiency increase (blue circles, Figure 5) with a blend design maximizing an individual blend property, here, the RON of the blend (orange diamonds, Figure 5). In general, the correlation between the predicted engine efficiency increase of the blend and the RON of the blend is strong, as evident by a Pearson's correlation coefficient of $\rho = 0.96$. The strong correlation is not surprising since the RON is a key property determining engine efficiency with the largest impact on the engine efficiency increase. Consequently, a blend design maximizing RON would be sufficient to identify many high-ranking candidate blends and could discard low-ranking candidate blends.

However, the correlation becomes weak among the high-ranking candidates in engine efficiency increase; e.g., for the highest-ranking 50 candidate blends, the correlation coefficient equals only $\rho = 0.49$. Only 2 of the top 10 blend components in the RON maximization are among the top 10 blend components in the engine efficiency maximization. Three of the top 10 blend components in the RON maximization are not even among the top 50 blend components in the engine efficiency maximization.

The blend component maximizing RON is the best pure-component fuel *tert*-butyl formate. The pure-component RON of *tert*-butyl formate (RON = 116) is higher than the RON of ethanol (RON = 109). Since *tert*-butyl formate already meets the property constraints as a pure component, ethanol neither increases the blending RON nor contributes to fulfilling constraints. Therefore, the "blend" maximizing RON does not contain any ethanol and is equal to pure *tert*-butyl formate. Apart from *tert*-butyl formate, 8 more "blends" would not contain ethanol or reduce the ethanol fraction if optimized for RON.

Thus, focusing the optimization on a single characteristic property, e.g., the blending RON, is insufficient for a ranking which accurately reflects engine efficiency. The most promising candidate blend components yield not only a high RON of the blend but also balance octane sensitivity, heat of vaporization, and laminar burning velocity. Therefore, CAMD and fuel design should use an application-based objective function that combines the effects of the individual properties.

3.2.2. Comparison between Pure-Component and Blend Results. Of the 226 candidate blend components, 26 fulfill the pure-component constraints. Ten of these candidates yield positive engine efficiency increases as pure-component fuels and thus are predicted to increase engine efficiency compared to RON95 gasoline. Six of these candidates were not discovered in the pure-component design. Particularly high-ranking are 2,2-dimethylbutyraldehyde, isobutyraldehyde, and isovaleraldehyde with predicted engine efficiency increases of 6.3%, 3.9%, and 3.6%, respectively. The additionally identified pure-component candidates indicate that our pure-component molecular design study is not exhaustive. For a more extensive list of candidates, the method needs to be rerun, e.g., for more generations, or the general solution strategy needs to be improved.

To compare the predicted relative engine efficiency increase of the blends to that of the pure components, we calculate the pure-component relative engine efficiency increase for all blend components ignoring property constraint violations (Figure 6). As evident by a correlation coefficient of $\rho = 0.69$, the pure-
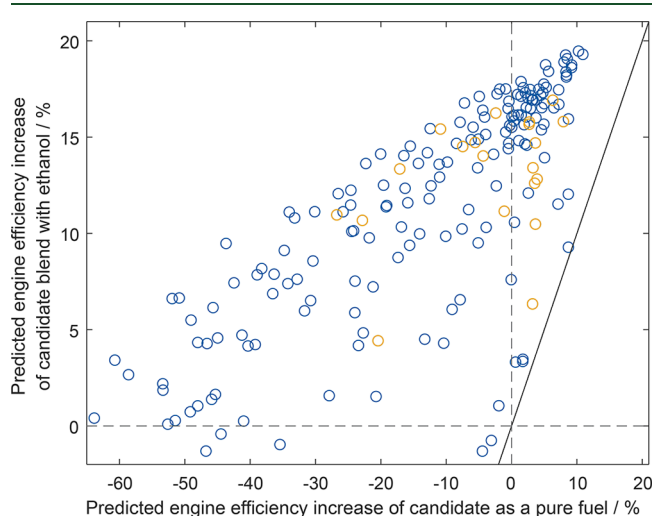
Figure 6. Parity plot comparing pure-component and blend merit function value of the candidate fuels. Each circle stands for one blend component. The orange circles represent blend components that fulfill the property constraints of the pure-component design as well. The black line indicates equal efficiency increase as a pure substance and as a blend component. The dashed gray lines represent the benchmark engine efficiency of RON95.
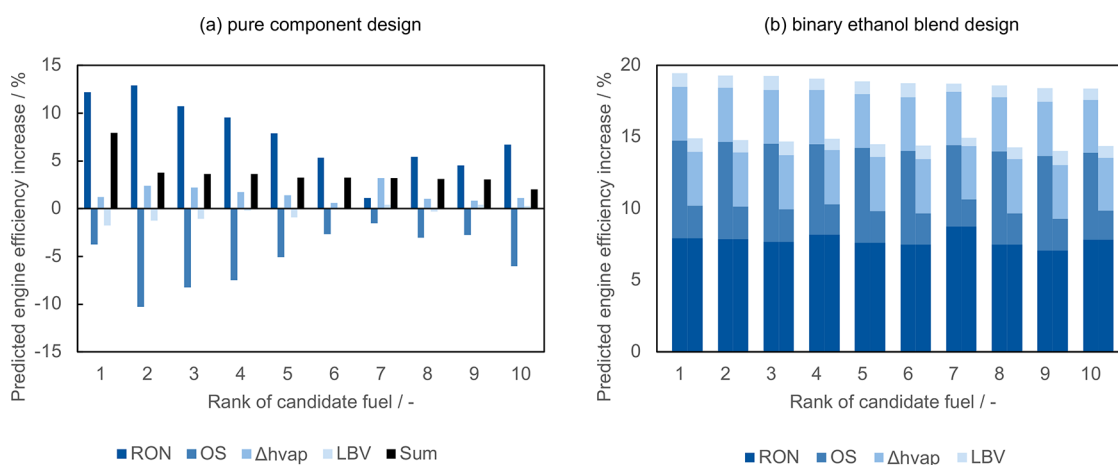
Figure 7. Bar plots showing the contributions of fuel properties to expected relative engine efficiency increase for the top 10 pure-component fuels (left) and top 10 binary blends with ethanol (right). In the right plot, the contributions are plotted for two K values, -1.5 (left bar) and -0.5 (right bar), highlighting the influence of the K value on the predicted engine efficiency increase.

component engine efficiency increase is a good indicator of the engine efficiency of the blend. High-ranking pure-component fuels usually lead to high-ranking blends; e.g., 18 of the 25 components with the highest pure efficiency increases are among the top 50 blend components. However, not all high-ranking blend components necessarily have a high pure-component efficiency increase: 19 of the top 50 blend components are not among the top 50 pure components. Thus, the effects of blending go beyond simple weighting of the objective function.

Generally, all candidate fuel blends are predicted to achieve a higher engine efficiency than the pure blend components if optimization targets maximum engine efficiency increase. In particular, even the optimal pure-component fuel *tert*-butyl formate benefits from blending with ethanol, raising the achievable relative engine efficiency increase to 16% for a blend with 55 mol % ethanol. The optimal balance of molecular properties also enables high-ranking blends for molecules with low pure-component efficiency increases. For example, 4-ethyl-4,5-dimethyl-1-hexene ranks 77th in the pure-component ranking and improves to rank 17 in a blend with 82 mol % ethanol through an increase in predicted engine efficiency from -0.9% as a pure component to 18% in the fuel blend.

The increased engine efficiency of the blends compared to the pure-component fuels is primarily attributed to the favorable combustion properties of ethanol. Ethanol has the highest predicted relative engine efficiency increase of the pure components (25%). Therefore, only 18 of the 184 blend components with a positive efficiency increase are blended with ethanol at more than 50 mol %. The majority of blend components act as enablers for ethanol since pure ethanol violates the constraint on the heat of vaporization. For 179 of the 184 blends, the blend's heat of vaporization matches the constraint limit of 60 kJ $\text{kg}_{\text{air},\lambda\,=\,1}^{-1}$. Conversely, 144 of the 184 candidate blend components violate the maximum boiling temperature as pure components but are feasible in the blend through a reduction of the bubble point by ethanol. Thus, blending allows us to meet the strict property constraints and significantly enlarges the molecular design space. To fulfill fuel property constraints for SI engines, blend design is therefore key.

3.2.3. Contribution of Fuel Properties to Expected Engine Efficiency Increase. We examine the contributions of the fuel properties to the predicted relative engine efficiency increase

(Figure 7). In the pure-component design, all top 10 candidates show positive contributions from RON and Δh_{vap}, while OS always has a negative contribution (Figure 7, left). For most candidates, the RON is the strongest contributor. The first five candidates are esters, where both low OS and low LBV show negative impacts on efficiency. For the seventh candidate, 1-butanol, Δh_{vap} has the biggest impact on efficiency.

For the binary ethanol blends, the picture looks markedly different (cf. Figure 7, right): Here, all four properties (RON, OS, Δh_{vap}, and LBV) positively impact engine efficiency. Ethanol having a large share in all designed blends is well-known for its high RON, OS, Δh_{vap}, and LBV.

As the relevance of OS for engine efficiency strongly depends on the K value, we also plot the results for an alternative choice of $K = -0.5$. Due to the linear contribution of the K value, the predicted efficiency increase due to OS is simply divided by three, lowering the total relative efficiency increase of the top blend from 19.5% ($K = -1.5$) to 14.9% ($K = -0.5$). Note that the choice for the K value also changes the ranking. Considering the uncertainty of the K factor, the ranking should thus not be considered as absolute. Instead, it seems appropriate to utilize expert knowledge, e.g., on the potential for large-scale production, to further prioritize among the vast number of promising blend candidates.

3.2.4. Comparison between Engine Efficiency Increase and Direct CO_2 Emissions Caused As an Alternative Objective. An optimal renewable fuel should not only lead to high engine efficiency but ultimately must enable sustainable mobility with low environmental impact, e.g., with low CO_2 emissions. Therefore, for each candidate blend, we calculate the direct CO_2 emissions generated by driving 100 km, based on the maximum engine efficiency. We compare the CO_2 emissions of the blends to the CO_2 emissions of RON95 gasoline, assuming a fuel consumption of RON95 of 7 L/100 km. For the calculation, we assume constant engine operation at the optimal operating point achieving maximum engine efficiency. Although this assumption limits the significance for practical implementation because no full driving cycle is considered, we can investigate whether the maximization of engine efficiency also minimizes direct CO_2 emissions, which are additionally influenced by the heating value and the amount of carbon of each fuel.

To calculate the CO_2 emissions of the blends ($\dot{m}_{\text{CO}_2}$), we assume that the thermal engine efficiency increase of a fuel

reduces the energy demand for the same engine power P_{engine} as with RON95:

$$\dot{m}_{CO_2} = \frac{M_{CO_2}}{M_{blend}} \cdot N_{C,blend} \cdot \dot{m}_{blend}$$

$$= \frac{M_{CO_2}}{M_{blend}} \cdot N_{C,blend} \cdot \frac{\dot{m}_{RON95} \cdot LHV_{RON95}}{\frac{merit}{100\%} + 1} \cdot \frac{1}{LHV_{blend}}$$

(3)

In this equation, the mass-based fuel consumption of the candidate blend and RON95 are denoted by $\dot{m}$. The molar mass of the candidate blend and CO_2 are denoted by M, and $N_{C,blend}$ stands for the average number of carbon atoms per molecule in the blend. The lower heating value of a fuel blend (LHV_{blend}) is calculated from linear mixing of the pure-component heating values (eq 2). The pure components' lower heating values are calculated from the stoichiometric combustion reaction using standard enthalpies of formation. The gas-phase standard enthalpy of formation for each candidate fuel is predicted from a GC-based GPR similar to the EHS hazards (see Section 2.3, $N_{Train} = 697$, $N_{Test} = 78$, Accuracy: $R^2_{test} = 1.00$, $RMSE_{test} = 16$ kJ mol^{-1}, nRMSE$_{test} = 1.5\%$). More details and property data used for RON95 can be found in the Supporting Information.

In total, 177 candidate blends with an increased engine efficiency compared to RON95 (Figure 8) reduce the direct
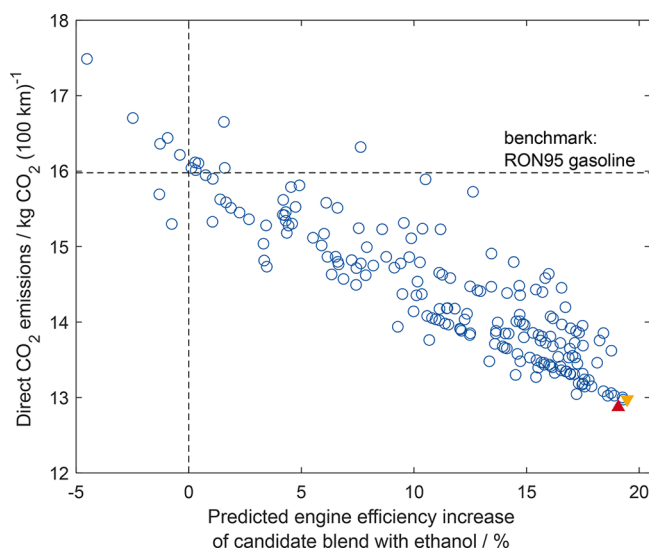

Figure 8. Direct CO_2 emissions from fuel combustion compared to predicted relative engine efficiency increase for each candidate blend. The red upward-pointing triangle is the candidate blend with the lowest direct CO_2 emissions. The orange downward-pointing triangle is the candidate blend with the highest predicted relative engine efficiency increase. The dashed gray lines represent the benchmark engine efficiency and emissions of RON95.

CO_2 emissions compared to RON95. The blend with the lowest CO_2 emissions (red upward-pointing triangle in Figure 8) contains 19 mol % 2,3,3,4-tetramethyl-pentane and 81 mol % ethanol and reduces the exhaust gas CO_2 emissions compared to RON95 by 19% to 12.9 kg CO_2/100 km. 2,3,3,4-tetramethyl-pentane also achieves one of the highest relative efficiency increases with 19%, ranking fourth in engine efficiency increase. The blend with the highest efficiency increase (orange downward-pointing triangle in Figure 8) also reduces direct CO_2 emissions by approximately 19% to 13.0 kg CO_2/100 km

and ranks third in CO_2 emissions reduction. Generally, high engine efficiency is closely aligned with lower direct CO_2 emissions ($\rho = -0.88$). However, the energy and carbon content of the fuel also significantly affect CO_2 emissions. Thus, a fuel design maximizing engine efficiency is not sufficient to guarantee minimum direct CO_2 emissions.

Besides the direct exhaust emissions of a vehicle, emissions during the production of the blend are highly relevant for environmental assessment.[100] Therefore, a meaningful fuel design with an environmental objective must include a "well-to-wheel" assessment.

4. CONCLUSION

We present a method for the molecular design of fuels and fuel blends for future, dedicated spark-ignition engines with very high compression ratios that uses an empirical model of engine efficiency as the objective function. The molecular optimization integrates the prediction of various properties to evaluate the feasibility and expected engine efficiency of each candidate fuel. Thermodynamic properties are calculated using COSMO-RS. Combustion properties are predicted by a graph neural network and a group contribution-based artificial neural network from the literature. Constraints on environmental, health, and safety indicators are assessed using Gaussian process regression.

The method is applied to design candidate fuels with high spark-ignition engine efficiency in two case studies: (1) the design of pure-component fuels and (2) the design of binary blends with ethanol. The application highlights the challenge of finding suitable pure-component fuels that meet all property constraints. In our design, only 11 of the 1033 investigated molecules fulfill the property constraints and increase predicted engine efficiency compared to RON95 gasoline. As an optimal pure-component fuel, we identify *tert*-butyl formate with a predicted relative engine efficiency increase of 7.9%.

We extend the molecular design space by designing a two-component fuel blend with ethanol. We identify 184 blend components that exceed the engine efficiency of RON95. 136 of the 184 candidate blends also exceed the predicted engine efficiency of the best pure-component fuel *tert*-butyl formate, highlighting the significant potential of blend design. As optimal blend components, we identify highly branched alkenes increasing engine efficiency by up to 19%. In most cases, the designed blend component represents the minor constituent of the blend but satisfies the property constraints to enable the inclusion of a substantial amount of ethanol.

The fuel design method has successfully designed candidate fuels that are known to increase engine efficiency from experiments in the literature. However, for the less well-known candidate fuels proposed by the method, the predicted properties need to be confirmed. Furthermore, experimental engine tests are required to confirm the predicted efficiency increases.

In future work, the accuracy of the fuel design method should be improved, since the uncertainties in the predictions are still large for some properties. To increase the reliability and the significance of the predictions, the prediction uncertainties have already been considered for constraint evaluation and could be extended to design under uncertainty. Moreover, accuracy can be improved by integrating additional models accounting for nonideal mixture behavior. The mixture property models except for COSMO-RS are currently linear molar mixing rules, e.g., for toxicity and RON. However, toxicity and RON are known to frequently exhibit nonideal mixing behavior.[41,101] Moreover, for

the design of renewable fuels, e.g., produced from biomass and carbon dioxide, the assessment of synthesizability should be adapted to constrain the molecular design space to appropriate processing pathways. To lower the market-entry barriers of alternative fuels, blending with fossil gasoline to comply with existing fuel standards could be considered already in the design phase.

The integration of more accurate prediction of nonideal mixture behavior needs to consider the overall computational complexity of the method. In the presented fuel design method, the computationally expensive DFT calculations required by COSMO-RS limit the number of molecule evaluations and thus the exploration of the molecular design space. To investigate more molecules, a future method could employ a hierarchical approach balancing exploration and accuracy, or employ computationally efficient machine learning models also for thermodynamics.[102−104]

Future work could also extend our design method to a higher number of blend components or even optimize the number of blend components as a design degree of freedom, since this work showed the opportunities of blend design for spark-ignition engine fuels.

Finally, to improve the engine efficiency assessment, a more detailed engine model is desirable that considers a typical driving cycle instead of a correlation of potential engine efficiency. A detailed engine model would also eliminate the difficult task of choosing a single, representative K factor. Ultimately, the model for fuel assessment should consider not only the combustion of the fuel but the emissions of the entire life cycle of the fuel aiming at "well-to-wheel" optimization.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.energyfuels.2c03296.

> Details on software and hardware; prediction accuracy of environment, health, and safety indicators; fragment database of the genetic algorithm; influence of property constraints on the number of feasible candidate blends; calculation and property data for the estimation of direct CO_2 emissions (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

André Bardow − *Energy & Process Systems Engineering, ETH Zurich, 8092 Zurich, Switzerland; Institute of Energy and Climate Research (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany;* ⓘ orcid.org/0000-0002-3831-0691; Email: abardow@ethz.ch

Manuel Dahmen − *Institute of Energy and Climate Research (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany;* ⓘ orcid.org/0000-0003-2757-5253; Email: m.dahmen@fz-juelich.de

Authors

Lorenz Fleitmann − *Energy & Process Systems Engineering, ETH Zurich, 8092 Zurich, Switzerland; Institute for Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany;* ⓘ orcid.org/0000-0002-6350-5116

Philipp Ackermann − *Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany*

Johannes Schilling − *Energy & Process Systems Engineering, ETH Zurich, 8092 Zurich, Switzerland*

Johanna Kleinekorte − *Institute for Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany;* ⓘ orcid.org/0000-0002-1895-1539

Jan G. Rittig − *Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany*

Florian vom Lehn − *Institute for Combustion Technology (ITV), RWTH Aachen University, 52056 Aachen, Germany;* ⓘ orcid.org/0000-0002-7215-4530

Artur M. Schweidtmann − *Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany; Department of Chemical Engineering, Delft University of Technology, Delft 2629 HZ, The Netherlands*

Heinz Pitsch − *Institute for Combustion Technology (ITV), RWTH Aachen University, 52056 Aachen, Germany*

Kai Leonhard − *Institute for Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany;* ⓘ orcid.org/0000-0001-6231-6957

Alexander Mitsos − *JARA-ENERGY, 52056 Aachen, Germany; Process Systems Engineering (AVT.SVT), RWTH Aachen University, 52074 Aachen, Germany; Institute of Energy and Climate Research (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.energyfuels.2c03296

Author Contributions

[∇]L.F. and P.A. contributed equally to this work.

Author Contributions

L.F. developed the fuel design method and integrated the individual property prediction methods. P.A. contributed the conceptualization of the fuel design based on the engine efficiency as an application-based objective function. L.F. investigated and validated the design results with respect to the CAMD methodology. P.A. investigated and validated the design results with respect to the fuel design application. L.F. and P.A. jointly wrote the first draft of the paper. L.F.: Conceptualization, Methodology, Software, Validation, Investigation, Visualization, Writing - original draft, Writing - review and editing. P.A.: Conceptualization, Methodology, Validation, Investigation, Writing - original draft, Writing - review and editing. J.S.: Methodology, Writing - review and editing. J.K.: Software, Writing - review and editing. J.G.R.: Software, Writing - review and editing. F.v.L.: Software, Writing - review and editing. A.M.S.: Conceptualization, Writing - review and editing. H.P.: Supervision, Funding acquisition, Writing - review and editing. K.L.: Conceptualization, Supervision, Writing - review and editing. A.M.: Conceptualization, Supervision, Funding acquisition, Writing - review and editing. A.B.: Conceptualization, Supervision, Funding acquisition, Writing - review and editing. M.D.: Conceptualization, Supervision, Writing - review and editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ NOMENCLATURE

Abbreviations

AFR = stoichiometric mass air-to-fuel ratio

AiT = autoignition temperature

ANN = artificial neural network

BCF = bioconcentration factor

CAMD = computer-aided molecular design

DCN = derived cetane number

DFT = density functional theory

EHS = environment, health, and safety

GC = group contribution

GNN = graph neural network

GPR = Gaussian process regression

LBV = laminar burning velocity

LC_{50} = aqueous toxicity as lethal concentration for fish

LD_{50} = oral toxicity as lethal dose for rats

LHV = lower heating value

MAE = mean absolute error

MON = motor octane number

(n)RMSE = (normalized) root-mean-square error

OS = octane sensitivity

PEL = permissible exposure limit

QSPR = quantitative structure−property relationship

RON = research octane number

SI = spark ignition

Tox = toxicity

uYSI = unified yield sooting index

Greek Symbols

Δh_{vap} = enthalpy of vaporization

η = maximum brake thermal engine efficiency

Ω = general variable for properties

Latin Symbols

K = parameter describing an engine's operating conditions

$\dot{m}$ = mass flow

$N_{Train/Test}$ = number of data points in training/test set

R^2 = coefficient of determination

$T_{boil/bubble}$ = boiling/bubble point temperature

x_i = mole fraction

■ REFERENCES

(1) Leitner, W.; Klankermayer, J.; Pischinger, S.; Pitsch, H.; Kohse-Höinghaus, K. Advanced Biofuels and Beyond: Chemistry Solutions for Propulsion and Production. *Angewandte Chemie (International ed.)* **2017**, *56*, 5412−5452.

(2) Szybist, J. P.; Busch, S.; McCormick, R. L.; Pihl, J. A.; Splitter, D. A.; Ratcliff, M. A.; Kolodziej, C. P.; Storey, J. M.; Moses-DeBusk, M.; Vuilleumier, D.; Sjöberg, M.; Sluder, C. S.; Rockstroh, T.; Miles, P. What fuel properties enable higher thermal efficiency in spark-ignited engines? *Prog. Energy Combust. Sci.* **2021**, *82*, 100876.

(3) Verhelst, S.; Turner, J. W. g.; Sileghem, L.; Vancoillie, J. Methanol as a fuel for internal combustion engines. *Prog. Energy Combust. Sci.* **2019**, *70*, 43−88.

(4) Burkardt, P.; Ottenwälder, T.; König, A.; Viell, J.; Mitsos, A.; Wouters, C.; Marquardt, W.; Pischinger, S.; Dahmen, M. Toward co-optimization of renewable fuel blend production and combustion in ultra-high efficiency SI engines. *International Journal of Engine Research* **2021**, *24*, 146808742110409.

(5) Ackermann, P.; et al. Designed to Be Green, Economic, and Efficient: A Ketone-Ester-Alcohol-Alkane Blend for Future Spark-Ignition Engines. *ChemSusChem* **2021**, *14*, 5254−5264.

(6) Wang, C.; Xu, H.; Daniel, R.; Ghafourian, A.; Herreros, J. M.; Shuai, S.; Ma, X. Combustion characteristics and emissions of 2-methylfuran compared to 2,5-dimethylfuran, gasoline and ethanol in a DISI engine. *Fuel* **2013**, *103*, 200−211.

(7) Hoppe, F.; Heuser, B.; Thewes, M.; Kremer, F.; Pischinger, S.; Dahmen, M.; Hechinger, M.; Marquardt, W. Tailor-made fuels for future engine concepts. *International Journal of Engine Research* **2016**, *17*, 16−27.

(8) Nithyanandan, K.; Zhang, J.; Li, Y.; Wu, H.; Lee, T. H.; Lin, Y.; Lee, C.-f. F. Improved SI engine efficiency using Acetone−Butanol−Ethanol (ABE). *Fuel* **2016**, *174*, 333−343.

(9) Gani, R.; Zhang, L. Editorial overview: Frontiers of Chemical Engineering: Chemical Product Design. *Current Opinion in Chemical Engineering* **2020**, *27*, A1−A3.

(10) Zhang, L.; Mao, H.; Liu, Q.; Gani, R. Chemical product design − recent advances and perspectives. *Current Opinion in Chemical Engineering* **2020**, *27*, 22−34.

(11) Ng, L. Y.; Chong, F. K.; Chemmangattuvalappil, N. G. Challenges and opportunities in computer-aided molecular design. *Comput. Chem. Eng.* **2015**, *81*, 115−129.

(12) Zhang, L.; Babi, D. K.; Gani, R. New Vistas in Chemical Product and Process Design. *Annu. Rev. Chem. Biomol. Eng.* **2016**, *7*, 557−582.

(13) König, A.; Marquardt, W.; Mitsos, A.; Viell, J.; Dahmen, M. Integrated design of renewable fuels and their production processes: recent advances and challenges. *Current Opinion in Chemical Engineering* **2020**, *27*, 45−50.

(14) McCormick, R. L.; Fioroni, G.; Fouts, L.; Christensen, E.; Yanowitz, J.; Polikarpov, E.; Albrecht, K.; Gaspar, D. J.; Gladden, J.; George, A. Selection Criteria and Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Advanced Spark-Ignition Engines. *SAE International Journal of Fuels and Lubricants* **2017**, *10*, 442−460.

(15) Fioroni, G.; Fouts, L.; Luecke, J.; Vardon, D.; Huq, N.; Christensen, E.; Huo, X.; Alleman, T.; McCormick, R.; Kass, M.; Polikarpov, E.; Kukkadapu, G.; Whitesides, R. A. Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Mixing Controlled Compression Ignition Combustion. *SAE International Journal of Advances and Current Practices in Mobility* **2019**, *1*, 1117−1138.

(16) Huo, X.; et al. Tailoring diesel bioblendstock from integrated catalytic upgrading of carboxylic acids: a "fuel property first" approach. *Green Chem.* **2019**, *21*, 5813−5827.

(17) Huq, N. A.; et al. Performance-advantaged ether diesel bioblendstock production by a priori design. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 26421−26430.

(18) Kuzhagaliyeva, N.; Horváth, S.; Williams, J.; Nicolle, A.; Sarathy, S. M. Artificial intelligence-driven design of fuel mixtures. *Communications Chemistry* **2022**, *5*, 1−10.

(19) Hechinger, M. *Model-based identification of promising biofuel candidates for spark-ignited engines.* Dissertation, RWTH Aachen; Berichte aus der Aachener Verfahrenstechnik - Prozesstechnik, AVT, RWTH Aachen University; VDI-Verl.: Düsseldorf, 2014; Vol. *940*.

(20) Gugisch, R.; Kerber, A.; Kohnert, A.; Laue, R.; Meringer, M.; Rücker, C.; Wassermann, A. *Advances in Mathematical Chemistry and Applications*; Elsevier, 2015; pp 113−138.

(21) Hechinger, M.; Dahmen, M.; Villeda, J. J. V.; Marquardt, W. In *11th International Symposium on Process Systems Engineering*; Karimi, I. A., Srinivasan, R., Eds.; Computer Aided Chemical Engineering; Elsevier, 2012; Vol. *31*; pp 1341−1345.

(22) Dahmen, M.; Marquardt, W. Model-Based Design of Tailor-Made Biofuels. *Energy Fuels* **2016**, *30*, 1109−1134.

(23) Dahmen, M.; Marquardt, W. A novel group contribution method for the prediction of the derived cetane number of oxygenated hydrocarbons. *Energy Fuels* **2015**, *29*, 5781−5801.

(24) Rittig, J. G.; Ritzert, M.; Schweidtmann, A. M.; Winkler, S.; Weber, J. M.; Morsch, P.; Heufer, K. A.; Grohe, M.; Mitsos, A.;

Dahmen, M. Graph machine learning for design of high−octane fuels. *AIChE J.* **2022**, e17971.

(25) Gani, R. Chemical product design: challenges and opportunities. *Comput. Chem. Eng.* **2004**, *28*, 2441−2457.

(26) Kashinath, S. A. A.; Manan, Z. A.; Hashim, H.; Alwi, S. R. W. Design of green diesel from biofuels using computer aided technique. *Comput. Chem. Eng.* **2012**, *41*, 88−92.

(27) Hashim, H.; Narayanasamy, M.; Yunus, N. A.; Shiun, L. J.; Muis, Z. A.; Ho, W. S. A cleaner and greener fuel: Biofuel blend formulation and emission assessment. *Journal of Cleaner Production* **2017**, *146*, 208−217.

(28) Yunus, N. A.; Gernaey, K. V.; Woodley, J. M.; Gani, R. A systematic methodology for design of tailor-made blended products. *Comput. Chem. Eng.* **2014**, *66*, 201−213.

(29) Zhang, L.; Kalakul, S.; Liu, L.; Elbashir, N. O.; Du, J.; Gani, R. A Computer-Aided Methodology for Mixture-Blend Design. Applications to Tailor-Made Design of Surrogate Fuels. *Ind. Eng. Chem. Res.* **2018**, *57*, 7008−7020.

(30) Kalakul, S.; Zhang, L.; Fang, Z.; Choudhury, H. A.; Intikhab, S.; Elbashir, N.; Eden, M. R.; Gani, R. Computer aided chemical product design − ProCAPD and tailor-made blended products. *Comput. Chem. Eng.* **2018**, *116*, 37−55.

(31) Liu, Q.; Zhang, L.; Liu, L.; Du, J.; Tula, A. K.; Eden, M.; Gani, R. OptCAMD: An optimization-based framework and tool for molecular and mixture product design. *Comput. Chem. Eng.* **2019**, *124*, 285−301.

(32) Marvin, W. A.; Rangarajan, S.; Daoutidis, P. Automated Generation and Optimal Selection of Biofuel-Gasoline Blends and Their Synthesis Routes. *Energy Fuels* **2013**, *27*, 3585−3594.

(33) Ng, L. Y.; Andiappan, V.; Chemmangattuvalappil, N. G.; Ng, D. K. A systematic methodology for optimal mixture design in an integrated biorefinery. *Comput. Chem. Eng.* **2015**, *81*, 288−309.

(34) Dahmen, M.; Marquardt, W. Model-based formulation of biofuel blends by simultaneous product and pathway design. *Energy Fuels* **2017**, *31*, 4096−4121.

(35) König, A.; Neidhardt, L.; Viell, J.; Mitsos, A.; Dahmen, M. Integrated design of processes and products: Optimal renewable fuels. *Comput. Chem. Eng.* **2020**, *134*, 106712.

(36) König, A.; Siska, M.; Schweidtmann, A. M.; Rittig, J. G.; Viell, J.; Mitsos, A.; Dahmen, M. Designing production-optimal alternative fuels for conventional, flexible-fuel, and ultra-high efficiency engines. *Chem. Eng. Sci.* **2021**, *237*, 116562.

(37) Gschwend, D.; Soltic, P.; Edinger, P.; Wokaun, A.; Vogel, F. Performance evaluation of gasoline alternatives using a thermodynamic spark-ignition engine model. *Sustainable Energy Fuels* **2017**, *1*, 1991−2005.

(38) Gschwend, D.; Müller, S.; Wokaun, A.; Vogel, F. Optimum Fuel for Spark Ignition Engines from Lignin Pyrolysis Oil. *Energy Fuels* **2018**, *32*, 9388−9398.

(39) Gschwend, D.; Soltic, P.; Wokaun, A.; Vogel, F. Review and Performance Evaluation of Fifty Alternative Liquid Fuels for Spark-Ignition Engines. *Energy Fuels* **2019**, *33*, 2186−2196.

(40) Farrell, J. T.; Zigler, B. T.; Ratcliff, M. A.; Miles, P.; Kolodziej, C.; Sjoberg, M.; Sluder, S.; Szybist, J.; Wagner, S.; Splitter, D.; Pihl, J.; Toops, T.; Debusk, M.; Storey, J.; Vuilleumier, D.Co-Optimization of Fuels & Engines: Efficiency Merit Function for Spark-Ignition Engines; Revisions and Improvements Based on FY16-17 Research. DOI: 10.2172/1463450.

(41) vom Lehn, F.; Cai, L.; Tripathi, R.; Broda, R.; Pitsch, H. A property database of fuel compounds with emphasis on spark-ignition engine applications. *Applications in Energy and Combustion Science* **2021**, *5*, 100018.

(42) Li, R.; Herreros, J. M.; Tsolakis, A.; Yang, W. Integrated machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics for high throughput fuel screening toward internal combustion engine. *Fuel* **2022**, *307*, 121908.

(43) Gani, R. Group contribution-based property estimation methods: advances and perspectives. *Current Opinion in Chemical Engineering* **2019**, *23*, 184−196.

(44) Gertig, C.; Leonhard, K.; Bardow, A. Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects. *Current Opinion in Chemical Engineering* **2020**, *27*, 89−97.

(45) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101−122.

(46) Hukkerikar, A. S.; Kalakul, S.; Sarup, B.; Young, D. M.; Sin, G.; Gani, R. Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution + (GC +) Property Models and Uncertainty Analysis. *J. Chem. Inf. Model.* **2012**, *52*, 2823−2839.

(47) Walters, W. P.; Barzilay, R. Applications of Deep Learning in Molecule Generation and Molecular Property Prediction. *Acc. Chem. Res.* **2021**, *54*, 263−270.

(48) Scheffczyk, J.; Fleitmann, L.; Schwarz, A.; Lampe, M.; Bardow, A.; Leonhard, K. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chem. Eng. Sci.* **2017**, *159*, 84−92.

(49) Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: a computer-aided ligand design for structure-based drug design. *J. Med. Chem.* **2005**, *48*, 2457−2468.

(50) Schweidtmann, A. M.; Rittig, J. G.; König, A.; Grohe, M.; Mitsos, A.; Dahmen, M. Graph Neural Networks for Prediction of Fuel Ignition Quality. *Energy Fuels* **2020**, *34*, 11395−11407.

(51) Kalghatgi, G. T.*Fuel Anti-Knock Quality - Part I. Engine Studies*. SAE Technical Paper 2001-01-3584; SAE, 2001.

(52) vom Lehn, F.; Cai, L.; Copa Cáceres, B.; Pitsch, H. Exploring the fuel structure dependence of laminar burning velocity: A machine learning based group contribution approach. *Combust. Flame* **2021**, *232*, 111525.

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

(54) Loschen, C.; Klamt, A. In *Computational Pharmaceutical Solid State Chemistry*; Abramov, Y. A., Ed.; Wiley, 2016; pp 211−233.

(55) Gao, W.; Coley, C. W. The Synthesizability of Molecules Proposed by Generative Models. *J. Chem. Inf. Model.* **2020**, *60*, 5714−5723.

(56) Alshehri, A. S.; Tula, A. K.; You, F.; Gani, R. Next generation pure component property estimation models: With and without machine learning techniques. *AIChE J.* **2021**, e17469.

(57) Li, R.; Herreros, J. M.; Tsolakis, A.; Yang, W. Machine learning and deep learning enabled fuel sooting tendency prediction from molecular structure. *Journal of Molecular Graphics & Modelling* **2022**, *111*, 108083.

(58) ASTM International,*ASTM Standard Test Method E659-78*; The American Society for Testing and Materials: West Conshohocken, PA, United States, 2000.

(59) Arnot, J. A.; Gobas, F. A. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environmental Reviews* **2006**, *14*, 257−297.

(60) Ankley, G. T.; Villeneuve, D. L. The fathead minnow in aquatic toxicology: Past, present and future. *Aquatic Toxicology* **2006**, *78*, 91−102.

(61) Walum, E. Acute oral toxicity. *Environ. Health Perspect.* **1998**, *106*, 497−503.

(62) Spear, R. C.; Selvin, S. OSHA's permissible exposure limits: regulatory compliance versus health risk. *Risk Anal.* **1989**, *9*, 579−586.

(63) Das, D. D.; St. John, P. C.; McEnally, C. S.; Kim, S.; Pfefferle, L. D. Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale. *Combust. Flame* **2018**, *190*, 349−364.

(64) McEnally, C. S.; Das, D. D.; Pfefferle, L. D.*Yield Sooting Index Database Vol. 2: Sooting Tendencies of a Wide Range of Fuel Compounds on a Unified Scale*; Harvard Dataverse, V1, 2017.

(65) Aikawa, K.; Sakurai, T.; Jetter, J. J. Development of a Predictive Model for Gasoline Vehicle Particulate Matter Emissions. *SAE International Journal of Fuels and Lubricants* **2010**, *3*, 610−622.

(66) Westbrook, C. K.; Pitz, W. J.; Curran, H. J. Chemical Kinetic Modeling Study of the Effects of Oxygenated Hydrocarbons on Soot Emissions from Diesel Engines. *J. Phys. Chem. A* **2006**, *110*, 6912−6922.

(67) Leach, F.; Chapman, E.; Jetter, J. J.; Rubino, L.; Christensen, E. D.; St. John, P. C.; Fioroni, G. M.; McCormick, R. L. A Review and Perspective on Particulate Matter Indices Linking Fuel Composition to Particulate Emissions from Gasoline Engines. *SAE International Journal of Fuels and Lubricants* **2022**, *15*, 3−28.

(68) Samaras, Z. C.; Andersson, J.; Bergmann, A.; Hausberger, S.; Toumasatos, Z.; Keskinen, J.; Haisch, C.; Kontses, A.; Ntziachristos, L. D.; Landl, L.; Mamakos, A.; Bainschab, M. Measuring Automotive Exhaust Particles Down to 10 nm. *SAE International Journal of Advances and Current Practices in Mobility* **2021**, *3*, 539−550.

(69) Goodfellow, I.; Bengio, Y.; Courville, A.*Deep Learning*; MIT Press, 2016.

(70) Kullback, S.; Leibler, R. A. On information and sufficiency. *Annals of Mathematical Statistics* **1951**, *22*, 79−86.

(71) Müller, S. Flexible heuristic algorithm for automatic molecule fragmentation: application to the UNIFAC group contribution model. *Journal of Cheminformatics* **2019**, *11*, 57.

(72) Duvenaud, D.; Lloyd, J.; Grosse, R.; Tenenbaum, J.; Zoubin, G. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. *Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA* **2013**, *28*, 1166−1174.

(73) Duvenaud, D.*Automatic model construction with Gaussian processes*. Ph.D. thesis, Apollo - University of Cambridge Repository, 2014.

(74) Lloyd, J.; Duvenaud, D.; Grosse, R.; Tenenbaum, J.; Ghahramani, Z. Automatic Construction and Natural-Language Description of Nonparametric Regression Models. *Proceedings of the AAAI Conference on Artificial Intelligence* **2014**, *28*, 1242−1250.

(75) Sun, Y.; Sahinidis, N. V. Computer-aided retrosynthetic design: fundamentals, tools, and outlook. *Current Opinion in Chemical Engineering* **2022**, *35*, 100721.

(76) Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using transformer-based models and a hypergraph exploration strategy. *Chemical Science* **2020**, *11*, 3316−3325.

(77) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Science* **2019**, *5*, 1572−1583.

(78) Dirrenberger, P.; Glaude, P. A.; Bounaceur, R.; Le Gall, H.; Da Cruz, A. P.; Konnov, A. A.; Battin-Leclerc, F. Laminar burning velocity of gasolines with addition of ethanol. *Fuel* **2014**, *115*, 162−169.

(79) Remmert, S.; Campbell, S.; Cracknell, R.; Schuetze, A.; Lewis, A.; Giles, K.; Akehurst, S.; Turner, J.; Popplewell, A.; Patel, R. Octane Appetite: The Relevance of a Lower Limit to the MON Specification in a Downsized, Highly Boosted DISI Engine. *SAE International Journal of Fuels and Lubricants* **2014**, *7*, 743−755.

(80) Kassai, M.; Aksu, C.; Shiraishi, T.; Cracknell, R.; Shibuya, M.*Mechanism Analysis on the Effect of Fuel Properties on Knocking Performance at Boosted Conditions*. SAE Technical Paper 2019-01-0035; SAE, 2019.

(81) Singh, E., Mohammed, A., Gorbatenko, I., Sarathy, M.On the Relevance of Octane Sensitivity in Heavily Downsized Spark-Ignited Engines. 15th International Conference on Engines & Vehicles, 12−16 September 2021, Capri, Napoli, Italy, CEVOLVER, 2021.

(82) Larsen, U.; Johansen, T.; Schramm, J.*Ethanol as a Future Fuel for Road Transportation: Main report*. https://www.osti.gov/etdeweb/biblio/21580860.

(83) Thewes, M.; Muether, M.; Pischinger, S.; Budde, M.; Brunn, A.; Sehr, A.; Adomeit, P.; Klankermayer, J. Analysis of the Impact of 2-Methylfuran on Mixture Formation and Combustion in a Direct-Injection Spark-Ignition Engine. *Energy Fuels* **2011**, *25*, 5549−5561.

(84) Hoppe, F.; Burke, U.; Thewes, M.; Heufer, A.; Kremer, F.; Pischinger, S. Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines. *Fuel* **2016**, *167*, 106−117.

(85) *Standard Specification for Automotive Spark-Ignition Engine Fuel*; American Society for Testing and Materials, 2021.

(86) DIN German Institute for Standardization,Automotive fuels - Unleaded petrol - Requirements and test methods; German version EN 228:2012+A1:2017.

(87) Chen, L.; Stone, R. Measurement of Enthalpies of Vaporization of Isooctane and Ethanol Blends and Their Effects on PM Emissions from a GDI Engine. *Energy Fuels* **2011**, *25*, 1254−1259.

(88) Thewes, M.; Müther, M.; Brassat, A.; Pischinger, S.; Sehr, A. Analysis of the Effect of Bio-Fuels on the Combustion in a Downsized DI SI Engine. *SAE International Journal of Fuels and Lubricants* **2012**, *5*, 274−288.

(89) BP Europa, BP Benzin Bleifrei 95: Safety Data Sheet No. SCH2106. https://www.bp.com/content/dam/bp/country-sites/de_ch/switzerland/home/produkt-und-services/sicherheitsdatenblaetter/bp_benzin_bleifrei_95_eng.pdf.

(90) Itani, L. M.; Bruneaux, G.; Di Lella, A.; Schulz, C. Two-tracer LIF imaging of preferential evaporation of multi-component gasoline fuel sprays under engine conditions. *Proceedings of the Combustion Institute* **2015**, *35*, 2915−2922.

(91) Bardi, M.; Di Lella, A.; Bruneaux, G. A novel approach for quantitative measurements of preferential evaporation of fuel by means of two-tracer laser induced fluorescence. *Fuel* **2019**, *239*, 521−533.

(92) Kranz, P.; Kaiser, S. A. LIF-based imaging of preferential evaporation of a multi-component gasoline surrogate in a direct-injection engine. *Proceedings of the Combustion Institute* **2019**, *37*, 1365−1372.

(93) Altenburger, R.; Nendza, M.; Schüürmann, G. Mixture toxicity and its modeling by quantitative structure-activity relationships. *Environ. Toxicol. Chem.* **2003**, *22*, 1900−1915.

(94) Craig, D. K.; Baskett, R. L.; Davis, J. S.; Dukes, L.; Hansen, D. J.; Petrocchi, A. J.; Powell, T. J.; Sutherland, P. J.; Tuccinardi, T. E. Recommended default methodology for analysis of airborne exposures to mixtures of chemicals in emergencies. *Appl. Occup. Environ. Hyg.* **1999**, *14*, 609−617.

(95) Majer, V.; Svoboda, V.*Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Chemical data series; Blackwell Scientific Publications: Oxford, 1985; Vol. *32*.

(96) St. John, P. C.; Bartlett, M.; Kim, S.Group-contribution predictions of Yield Sooting Index (YSI) - YSI Estimator. 2022; https://ysi.ml.nrel.gov/.

(97) Shirazi, S. A.; Abdollahipoor, B.; Windom, B.; Reardon, K. F.; Foust, T. D. Effects of blending C3-C4 alcohols on motor gasoline properties and performance of spark ignition engines: A review. *Fuel Process. Technol.* **2020**, *197*, 106194.

(98) Pereira, R.; Pasa, V. Effect of mono-olefins and diolefins on the stability of automotive gasoline. *Fuel* **2006**, *85*, 1860−1865.

(99) Baehr, C.; Smith, G. J.; Sleeman, D.; Zevaco, T. A.; Raffelt, K.; Dahmen, N. Aldehydes and ketones in pyrolysis oil: analytical determination and their role in the aging process. *RSC Adv.* **2022**, *12*, 7374−7382.

(100) Deutz, S.; Bongartz, D.; Heuser, B.; Kätelhön, A.; Schulze Langenhorst, L.; Omari, A.; Walters, M.; Klankermayer, J.; Leitner, W.; Mitsos, A.; Pischinger, S.; Bardow, A. Cleaner production of cleaner fuels: wind-to-wheel − environmental assessment of CO_2-based oxymethylene ether as a drop-in fuel. *Energy Environ. Sci.* **2018**, *11*, 331−343.

(101) Smith, K. E. C.; Schmidt, S. N.; Dom, N.; Blust, R.; Holmstrup, M.; Mayer, P. Baseline toxic mixtures of non-toxic chemicals: "solubility addition" increases exposure for solid hydrophobic chemicals. *Environ. Sci. Technol.* **2013**, *47*, 2026−2033.

(102) Jirasek, F.; Alves, R. A. S.; Damay, J.; Vandermeulen, R. A.; Bamler, R.; Bortz, M.; Mandt, S.; Kloft, M.; Hasse, H. Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion. *J. Phys. Chem. Lett.* **2020**, *11*, 981−985.

(103) Sanchez Medina, E. I.; Linke, S.; Stoll, M.; Sundmacher, K. Graph neural networks for the prediction of infinite dilution activity coefficients. *Digital Discovery* **2022**, *1*, 216−225.

(104) Winter, B.; Winter, C.; Schilling, J.; Bardow, A. A smile is all you need: predicting limiting activity coefficients from SMILES with natural language processing. *Digital Discovery* **2022**, *1*, 859−869.

📖 Recommended by ACS

Comparative Study on Hydraulic and Electrical Transport Properties of Carbonate Rocks Based on Rock Typing

Zhao Zhang, Jie Li, *et al.*

FEBRUARY 06, 2023
ENERGY & FUELS

READ 🗗

Role of Different Catalysts on a Direct Growth Carbon Nanotube for Supercapacitor Electrodes

Melkiyur Isacfranklin, Dhayalan Velauthapillai, *et al.*

FEBRUARY 14, 2023
ENERGY & FUELS

READ 🗗

A Single-Atom Pd Catalyst Anchored on a Porous Organic Polymer for Highly Efficient Telomerization of 1,3-Butadiene with Methanol

Zhaozhan Wang, Yong Yang, *et al.*

FEBRUARY 07, 2023
INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH

READ 🗗

One on one with Melissa Gish

Matthew A. Wiebe, special to C&EN.

APRIL 11, 2022
C&EN GLOBAL ENTERPRISE

READ 🗗

Get More Suggestions >