# Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation

Downloaded from: https://research.chalmers.se, 2026-04-06 08:01 UTC

Citation for the original published paper (version of record):

Vargas-Hernández, R., Jorner, K., Pollice, R. et al (2023). Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation. Journal of Chemical Physics, 158(10). http://dx.doi.org/10.1063/5.0137103

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

The Journal of Chemical Physics

# Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation

Cite as: J. Chem. Phys. **158**, 104801 (2023); https://doi.org/10.1063/5.0137103
Submitted: 30 November 2022 • Accepted: 17 February 2023 • Accepted Manuscript Online: 17 February 2023 • Published Online: 08 March 2023

**Rodrigo A. Vargas-Hernández, Kjell Jorner, Robert Pollice, et al.**

COLLECTIONS

Paper published as part of the special topic on Modern Semiempirical Electronic Structure Methods

View Online
Export Citation
CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

Phase diagrams—Why they matter and how to predict them
The Journal of Chemical Physics **158**, 030902 (2023); https://doi.org/10.1063/5.0131028

Learning pair potentials using differentiable simulations
The Journal of Chemical Physics **158**, 044113 (2023); https://doi.org/10.1063/5.0126475

A mapping approach to surface hopping
The Journal of Chemical Physics **158**, 104111 (2023); https://doi.org/10.1063/5.0139734

AIP Publishing

© 2023 Author(s).

The Journal of Chemical Physics

ARTICLE

scitation.org/journal/jcp

# Inverse molecular design and parameter optimization with Hückel theory using automatic differentiation

**Cite as: J. Chem. Phys. 158, 104801 (2023); doi: 10.1063/5.0137103**
**Submitted: 30 November 2022 • Accepted: 17 February 2023 •**
**Published Online: 8 March 2023**

View Online Export Citation CrossMark

**Rodrigo A. Vargas-Hernández,**[1,2,a),b)] **Kjell Jorner,**[1,3,4] **Robert Pollice,**[1,3,c)] **and Alán Aspuru-Guzik**[1,2,3,5,6,7]

AFFILIATIONS

[1] Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada
[2] Vector Institute for Artificial Intelligence, 661 University Ave. Suite 710, Toronto, Ontario M5G 1M1, Canada
[3] Department of Computer Science, University of Toronto, 40 St. George St., Toronto, Ontario M5S 2E4, Canada
[4] Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Kemigården 4, SE-41258 Gothenburg, Sweden
[5] Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College St., Toronto, Ontario M5S 3E5, Canada
[6] Department of Materials Science and Engineering, University of Toronto, 184 College St., Toronto, Ontario M5S 3E4, Canada
[7] Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 661 University Ave., Toronto, Ontario M5G 1M1, Canada

**Note:** This paper is part of the JCP Special Topic on Modern Semiempirical Electronic Structure Methods.
[a)] **Current address:** Department of Chemistry and Chemical Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M1 Canada.
[b)] **Author to whom correspondence should be addressed:** vargashr@mcmaster.ca
[c)] **Current address:** Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, Groningen 9747 AG, The Netherlands.

## ABSTRACT

Semiempirical quantum chemistry has recently seen a renaissance with applications in high-throughput virtual screening and machine learning. The simplest semiempirical model still in widespread use in chemistry is Hückel's $\pi$-electron molecular orbital theory. In this work, we implemented a Hückel program using differentiable programming with the `JAX` framework based on limited modifications of a pre-existing NumPy version. The auto-differentiable Hückel code enabled efficient gradient-based optimization of model parameters tuned for excitation energies and molecular polarizabilities, respectively, based on as few as 100 data points from density functional theory simulations. In particular, the facile computation of the polarizability, a second-order derivative, via auto-differentiation shows the potential of differentiable programming to bypass the need for numeric differentiation or derivation of analytical expressions. Finally, we employ gradient-based optimization of atom identity for inverse design of organic electronic materials with targeted orbital energy gaps and polarizabilities. Optimized structures are obtained after as little as 15 iterations using standard gradient-based optimization algorithms.

*Published under an exclusive license by AIP Publishing.* https://doi.org/10.1063/5.0137103

## I. INTRODUCTION

Mathematical models that are both predictive and provide insight are a cornerstone of the physical sciences. However, accurate models for complicated processes often have no analytical solution and require large computational resources to solve numerically. At the same time, they also tend to be hard to interpret, as highlighted by Mulliken's famous quote "the more accurate the calculations became, the more the concepts tended to vanish into thin air."[1] Approximate models with problem-specific parameters are, therefore, used in practice, but finding optimal values for these parameters can be non-trivial. Parameter optimization normally requires considerable amounts of reference data and is done either manually or with algorithms that do not take advantage of first or higher order derivatives as the corresponding analytical expressions are often unavailable.

Published under an exclusive license by AIP Publishing

In chemistry, the Schrödinger equation is an archetype of such a mathematical model that describes the interactions between nuclei and electrons in both atoms and molecules. However, (near) exact solutions are too computationally expensive for most molecules of interest. Quantum chemistry is an entire research field dedicated to finding computationally efficient solutions to the Schrödinger equation by introducing prudent approximations or reformulations.[2] One approach that was extremely successful in the early days of quantum chemistry is the use of so-called *semiempirical* (SE) approximations.[3] The central idea is the use of problem-specific parameters to simplify the mathematical form of the Schrödinger equation. One of the earliest SE models was Hückel's method to treat the $\pi$-electrons in organic molecules.[4–7] Traditionally, the parameters in the Hückel method were derived manually by human scientists with the aim to reproduce properties for well-known reference molecules,[8,9] or they were derived from more accurate calculations.[10] Over the years, the Hückel method has been used for pedagogical purposes and for obtaining physical insight into problems in organic chemistry[11] and photochemistry and photophysics.[12,13] However, it can also be used as a fast method for the prediction of molecular properties[14] and for inverse design of molecules with desired target properties.[15,16]

The recent upsurge in machine learning (ML), and specifically deep neural networks, created a need for robust and efficient algorithms to co-optimize a very large number of model parameters for various architectures. This problem is now solved by automatic differentiation (AD), a technique to evaluate the derivatives of mathematical expressions via the chain rule.[17] Importantly, AD removes the need to determine analytic expressions for derivatives and makes complicated mathematical models amenable to gradient-based optimization, allowing them to be applied in the same way as general supervised machine learning models. Regular machine learning approaches like deep neural networks are meant to be very general mathematical models with a large number of parameters. Through learning, they can adapt to essentially any problem provided sufficient training data are available. In contrast, physics-based mathematical models have expressions that are specific to a certain type of problem to be solved and feature a much smaller number of parameters. Implementing physical models, such as quantum chemistry, within AD frameworks enables the use of default learning algorithms for parameter optimization with a potentially much smaller training data requirement. Along these lines, there are auto-differentiable versions of Hartree–Fock,[18,19] density functional theory (DFT),[19–23] excited state mean-field theory,[24] and other applications in physical sciences.[21,25–35] Over all, AD has been used to accelerate the calculation of gradient physical methods and to blend with ML algorithms. AD has also been fundamental for constructing more accurate semiempirical methods when combined with ML algorithms.[36–40]

In this work, we developed an auto-differentiable implementation of the Hückel method by minimal adaptation of an initially developed NumPy[41] version into the `JAX`[42] AD framework. We use this model to demonstrate the ease and efficiency of parameter fitting based on computational reference datasets for both excitation energies and molecular polarizabilities, a property calculated via a second-order derivative. Additionally, we demonstrate that our AD model allows for gradient-based inverse design by regarding the atomic composition of a molecular system as an adjustable parameter to find molecules with targeted properties.[16] The corresponding code is made publicly available, allowing it to be applied to a large variety of chemical problems. As the Hückel calculations are extremely fast, our workflow allows for facile development of property-specific models that can be readily used in molecular generative models that require on the order of $10^5$–$10^6$ property evaluations.

This paper is structured as follows: we first present a short introduction on automatic differentiation and the Hückel model (Secs. II A and II B). Following that, we execute inverse design of molecules as a fully differentiable procedure (Sec. III A) and perform optimization of the Hückel model parameters using modern gradient-based methods (Sec. III B).

## II. METHODS

### A. Automatic differentiation

Gradients and high-order derivatives are at the core of physical simulations. For physical models, common approaches to evaluate derivatives of any order are closed-form solutions, symbolic differentiation, and numerical differentiation, i.e., finite-differences.[17,43] For any function represented as a computer program, AD[17] is an alternative way to compute gradients and higher-order derivatives. AD makes use of the chain rule for differentiation to create a program that computes the gradients during evaluation. There are two main modes in AD, *forward* and *reverse* modes. For scalar functions, reverse mode is more efficient as differentiation requires a single evaluation of the function to fully compute the Jacobian. An example of reverse mode differentiation is the backpropagation algorithm that is used for training neural networks. For more details about AD, we refer the reader to Ref. 17.

The optimization of ML models is mostly done with methods that require the gradient of the *loss* or *error function* ($\mathcal{L}$) with respect to the model parameters ($\boldsymbol{\theta}$), $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})$. All contemporary ML libraries, e.g., `Tensorflow`,[44] `PyTorch`,[45] and `JAX`,[42] are built on top of an AD engine, which computes $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})$ for any ML model. Given the robustness of AD libraries, differentiating physical models[43] could be done similarly to modern ML algorithms.

### B. Hückel model

The Hückel model, a well-known semiempirical quantum chemistry model,[4–7] was first proposed to describe the interactions of $\pi$-electrons in conjugated unsaturated hydrocarbons. In the Hückel model, this interaction is restricted to electrons centered at nearest-neighbor atoms. Generally, the Hückel model is considered a tight-binding type Hamiltonian [Eq. (1)], where the on-site and hopping parameters are commonly denoted in the literature as $\alpha_\ell$ and $\beta_{\ell,k}$, respectively. The matrix elements of the Hückel Hamiltonian are given by

$$H_{\ell,k} = \begin{cases} \alpha_\ell, & \ell = k, \\ \beta_{\ell,k}, & \ell \text{ and } k \text{ are adjacent}, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where the $\alpha_\ell$ parameters roughly represent the energy of an electron in a p orbital of the corresponding element, and the $\beta_{\ell,k}$ parameters describe the energy of an electron in the bond $\ell - k$. Extensions of

Published under an exclusive license by AIP Publishing

the Hückel Hamiltonian are possible and can, for instance, incorporate distance-dependence via $\beta_{\ell,k} = \beta^0_{\ell,k}\, g(\mathbf{R}_{\ell,k})$ (cf. Sec. III B). For more details, we refer the reader to standard quantum chemistry textbooks.[46,47]

Notably, any molecular property computed with the Hückel method depends directly on the $\alpha_\ell$ and $\beta_{\ell,k}$ parameters. Therefore, by tuning their values, one can either construct a more accurate Hückel model for a given molecule and property (cf. Sec. III B) or search for atomic compositions that optimize target properties given a preset connectivity (cf. Sec. III A). In Secs. III A and III B, we demonstrate how AD can be used to facilitate both these types of problems.

## III. RESULTS AND DISCUSSION

### A. Inverse molecular design

Inverse molecular design can be carried out via gradient-based optimization methods, as shown in Ref. 16. The Hückel model can be extended to search for the molecular structure with a desired property. Both the diagonal and off-diagonal elements of the Hückel Hamiltonian matrix can be described by a weighted average of atom types at each site,

$$H_{\ell,k} = \begin{cases} \sum_i^M b_\ell^i \alpha_\ell^i, & \ell = k, \\ \sum_i^M \sum_j^M b_\ell^i b_k^j \beta_{\ell,k}^{ij}, & \ell \text{ and } k \text{ are adjacent}, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

where $b_\ell^i$ is the weight of the atom of type $i$ for site $\ell$.

For a meaningful description, the weights of each site must be normalized, i.e., $\sum_i^M b_\ell^i = 1$. $M$ is the total number of atom types considered in the search. As a proof of concept, we consider eight different molecular frameworks,[16] which are displayed in Fig. 1. The $x$-symbol indicates the sites with variable atom types to be optimized. We only considered carbon $(C)$, nitrogen $(N)$, and phosphorus $(P)$, i.e., $M = 3$, as these atom types each contribute one electron, assuming that the remaining valences of carbon will be satisfied with a bond to an implicit hydrogen atom, and can be incorporated interchangeably at all sites with two neighbors in the $\pi$-framework (Fig. 1). Therefore, we defined the following vector of atom type weight parameters: $\mathbf{b}_\ell = [b_\ell^C, b_\ell^N, b_\ell^P]$. For clarity, $\mathbf{b}$ jointly describes the $\mathbf{b}_\ell$ parameters for all search sites in a molecule, i.e., $\mathbf{b} = \{\mathbf{b}_\ell\}^N$. For all the results presented, the values of the $\alpha_\ell$ and $\beta^0_{\ell,k}$ parameters were previously optimized with respect to the desired property (cf. Sec. III B).

For the set of eight molecules considered (cf. Fig. 1), we search for the type of atom $\mathbf{b}$ at each site that gives the lowest HOMO–LUMO gap [Eq. (3)], denoted as $\varepsilon_{HL}$, and the maximum polarizability denoted as $\langle\tilde{\alpha}\rangle$ [Eq. (4)]. $\varepsilon_{HL}$ is defined as

$$\varepsilon_{HL} = \epsilon_{LUMO} - \epsilon_{HOMO}, \tag{3}$$

where $\epsilon_{HOMO}$ and $\epsilon_{LUMO}$ are the eigenvalues of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), respectively. The polarizability function is defined as

$$\langle\tilde{\alpha}\rangle = \frac{1}{3}(\tilde{\alpha}_{xx} + \tilde{\alpha}_{yy} + \tilde{\alpha}_{zz}), \tag{4}$$

where the $\tilde{\alpha}_{ij}$ elements are the polarizability components defined as[16,48–51]

$$\tilde{\alpha}_{ij} = -\frac{\partial^2 E}{\partial F_i \partial F_j}. \tag{5}$$

The $F_i$ terms are the components of the electric field, $\vec{F} = [F_x, F_y, F_z]$, and $E$ is the electronic energy of the system. The elements of the polarizability tensor are usually computed using an approximate finite-differences (FD) approach called the finite field method, which

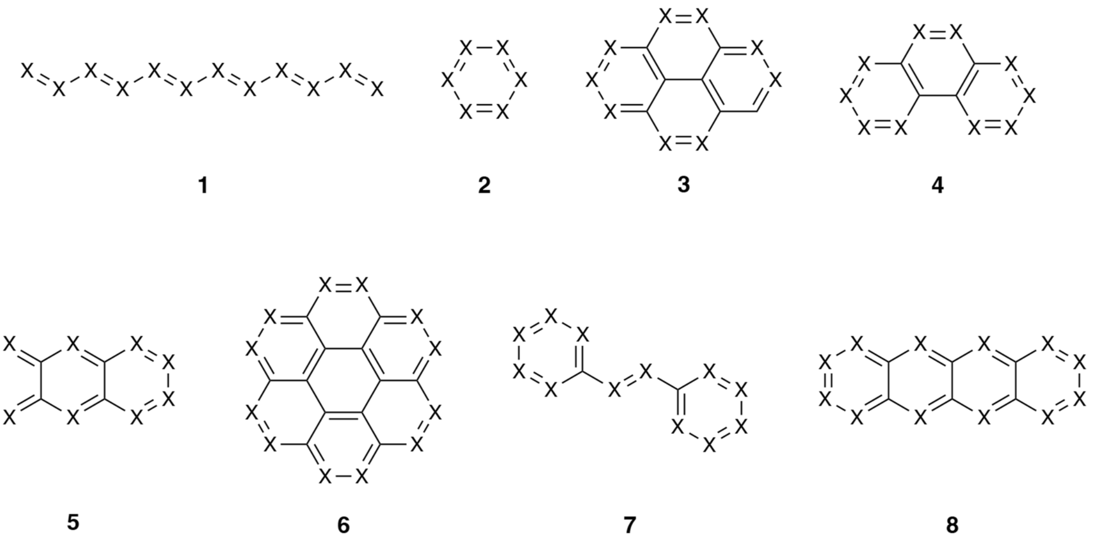

**FIG. 1.** The eight molecular frameworks considered for inverse molecular design in this work.[16] The $x$-symbol represents atomic sites whose identity is optimized.

was introduced in quantum chemistry by Cohen and Roothaan.[48] The response of the energy of the system to the electric field is truncated to second order, and numerical differentiation is used to solve for the polarizability,[52]

$$\tilde{\alpha}_{ii} = \frac{2E(0) - [E(-F_i) + E(+F_i)]}{F_i^2}, \tag{6}$$

where the electronic energy is evaluated several times, typically three times for each diagonal element, and four times for each cross term. Equation (6) is homologous to approximating a second-order derivative via central differences.[52] Notably, if the parameters $\mathbf{b}$ are to be optimized using a gradient-based method, the Jacobians $\nabla_{\mathbf{b}}\varepsilon_{HL}$ and $\nabla_{\mathbf{b}}\langle\tilde{\alpha}\rangle$ will also be constructed using an FD approach. However, this will increase the number of energy evaluations needed, especially for $\langle\tilde{\alpha}\rangle$ as the elements of $\nabla_{\mathbf{b}}\langle\tilde{\alpha}\rangle$ are third-order derivatives: $\frac{\partial}{\partial \mathbf{b}^{\ell}}\frac{\partial^2 \tilde{\alpha}_{ii}}{\partial F_i^2}$. Thus, for a single element of $\nabla_{\mathbf{b}}\langle\tilde{\alpha}\rangle$, using FD will require 18 energy calculations, $\mathcal{O}(18 \times \|\mathbf{b}\|)$, where $\|\mathbf{b}\|$ is the total number of parameters in $\mathbf{b}$. For $\varepsilon_{HL}$, using FD, we only require $\mathcal{O}(2 \times \|\mathbf{b}\|)$, as $\nabla_{\mathbf{b}}\varepsilon_{HL}$ is a first-order derivative.

In contrast, using modern AD frameworks, we can efficiently compute, with a single energy calculation (i.e., one forward pass), the Jacobian of $\varepsilon_{HL}$ with respect to $\mathbf{b}$. For $\langle\tilde{\alpha}\rangle$, the number of total energy evaluations depends on the dimension of the external field to construct the diagonal elements of the Hessian [Eq. (5)], which we also computed using AD. The Jacobian of $\langle\tilde{\alpha}\rangle$ with respect to $\mathbf{b}$, a third-order derivative, can be constructed from only three energy evaluations using AD,[17] a drastic reduction from the 18 required for FD.

After implementation of the Hückel model using the `JAX` ecosystem,[42] we could fully differentiate both observables, $\varepsilon_{HL}$ and $\langle\tilde{\alpha}\rangle$. Importantly, using `JAX` allowed us to convert our existing Python-based Hückel code very easily by replacing calls to `NumPy` with almost equivalent calls to the `JAX.Numpy` package. For optimization, we used the Broyden–Fletcher–Goldfarb–Shanno (`BFGS`) algorithm via the `JaxOpt` library.[53] Instead of using a constrained optimization scheme to satisfy the site-normalization restriction for $\mathbf{b}$, we used the softmax function [Eq. (7)],

$$b^i = \frac{\exp(\underline{b}^i)}{\sum_i^M \exp(\underline{b}^i)}, \tag{7}$$

where $\underline{b}$ represents the unnormalized $b$ parameters.

Given the flexibility of the `JAX` ecosystem, we were able to test other gradient optimization algorithms, such as `Adam`[54] and canonical gradient-descent, but found the `BFGS` to be most efficient as it required, on average, 15 or less total iterations to reach convergence (cf. Fig. 2). The `Adam` and gradient-descent algorithms, with an exponential learning rate decay, each needed more than 30 iterations to minimize $\varepsilon_{HL}$ or $\langle\tilde{\alpha}\rangle$. Notably, we initialized the values for all $\underline{b}^i$ parameters by sampling a uniform distribution, $\underline{b}^i \sim \mathcal{U}(-1, 1)$. Instead of using literature Hückel parameters, we used our optimized parameters for each target observable where 5000 training molecules were used. More details are described in Sec. III B. As can be seen in Figs. 3 and 4, we found that our random initialization of $\underline{b}$ allows us to sample a wide range of molecules with a broad range of values for both objectives, $\varepsilon_{HL}^{\text{initial}}$ and $\langle\tilde{\alpha}\rangle_{\text{initial}}$.

Because of the statistical description of the molecules by the $\mathbf{b}$ parameters, the optimal parameters ($\mathbf{b}^*$) found by optimization are not one-hot[55] vectors that correspond to only one atom type per site in the molecule but rather a linear combination of multiple atom types. We define the observable value for this unphysical molecule as $y_{\text{virtual}}$ and that for the real molecule as $y_{\text{feasible}}$ (i.e., the most probable atom is picked for each site to define the real molecule). An example of this is displayed for framework **3** in Fig. 5, where we show the change in $\mathbf{b}$ throughout the optimization for both objectives ($\varepsilon_{HL}$ and $\langle\tilde{\alpha}\rangle$). As we can observe, the change in $\varepsilon_{HL}$ from the initial random molecule to the optimal one is close to 1 eV. The optimizations in Fig. 5 were done using `Adam` only to properly illustrate the change in $\mathbf{b}$ as the change between iterations is smoother and more readily discernible. Importantly, this shows that the generative model can shift the distribution of properties toward the target property with

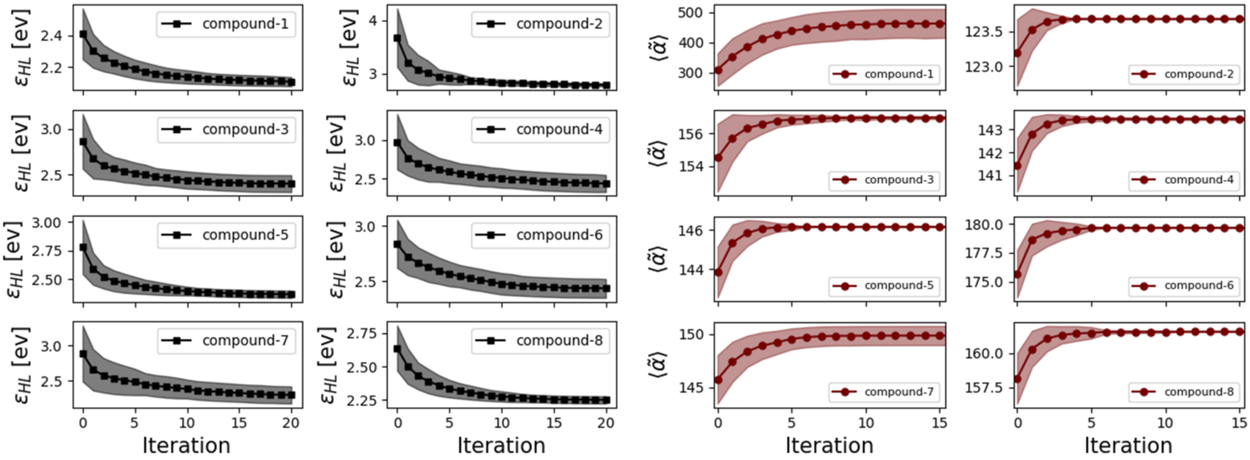

**FIG. 2.** Average learning curve for $\varepsilon_{HL}$ (left panel) and $\langle\tilde{\alpha}\rangle$ (right panel) for 250 random initial molecules based on the eight different molecular frameworks. We use the `BFGS` algorithm to optimize both observables. More details about the random initialization are provided in the text.

Published under an exclusive license by AIP Publishing

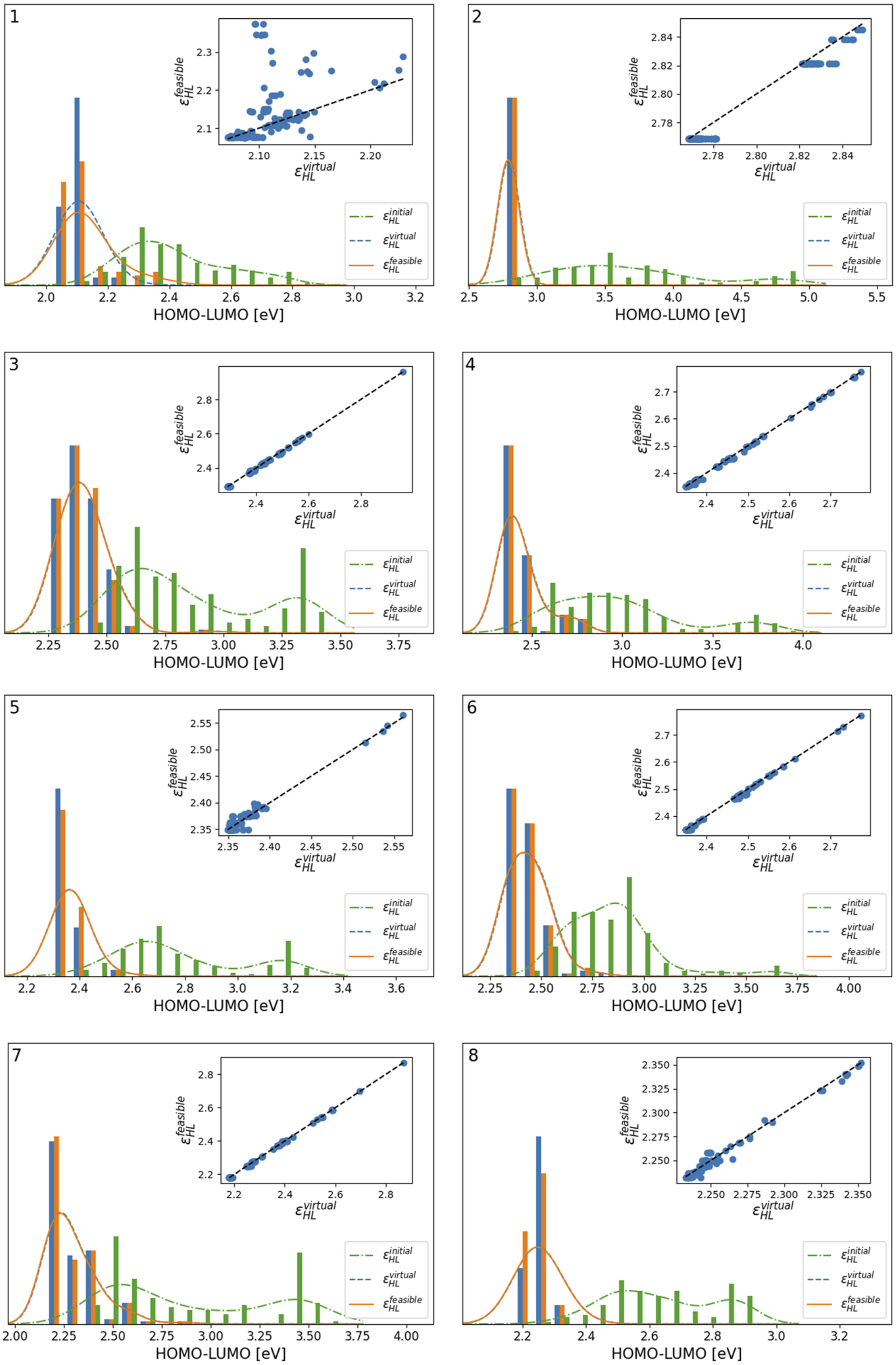

**FIG. 3.** Histograms of the optimized HOMO–LUMO gap [Eq. (3)] for 250 random initial molecules ($\varepsilon_{HL}^{\text{initial}}$). The inset panels compare the similarity between $\varepsilon_{HL}$ computed with the value of **b** at the end of the optimization protocol ($\varepsilon_{HL}^{\text{virtual}}$) and the values of $\varepsilon_{HL}$ selecting the most probable atoms given $\mathbf{b}^*$ ($\varepsilon_{HL}^{\text{feasible}}$). Curves represent the derived histograms using kernel density estimation, $\varepsilon_{HL}^{\text{feasible}}$ (solid), $\varepsilon_{HL}^{\text{virtual}}$ (dashed), and $\varepsilon_{HL}^{\text{initial}}$ (dotted-dashed). All molecules were optimized using the `BFGS` algorithm. Molecular frameworks are displayed in Fig. 1.

Published under an exclusive license by AIP Publishing

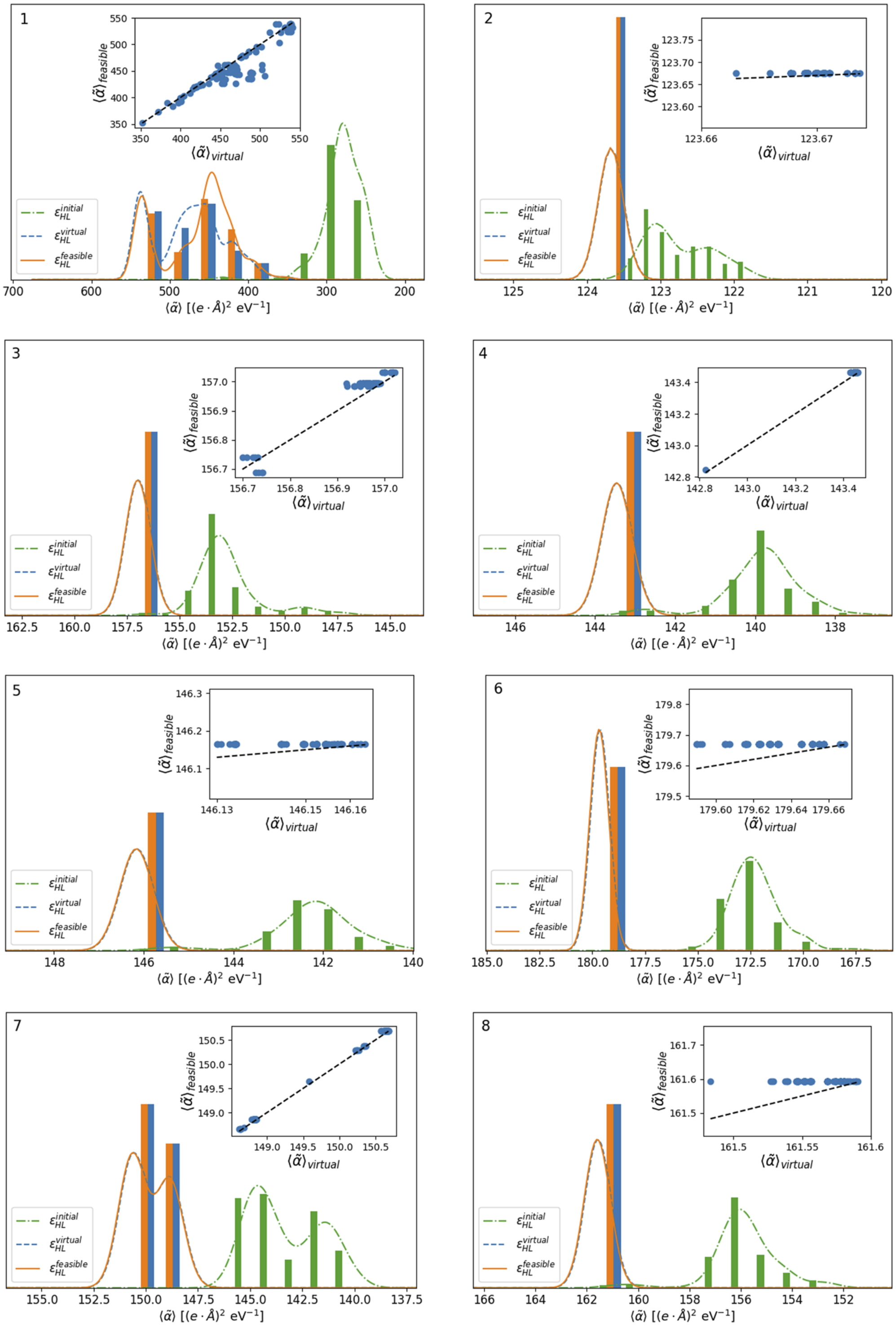

**FIG. 4.** Histograms of the optimized $\langle\tilde{\alpha}\rangle$ [Eq. (4)] for 250 random initial molecules ($\langle\tilde{\alpha}\rangle^{\text{initial}}$). The inset panels compare the similarity between $\langle\tilde{\alpha}\rangle$ computed with the value of **b** at the end of the optimization protocol ($\langle\tilde{\alpha}\rangle^{\text{virtual}}$) and the values of $\langle\tilde{\alpha}\rangle$ selecting the most probable atoms given $\mathbf{b}^*$ ($\langle\tilde{\alpha}\rangle^{\text{feasible}}$). Curves represent the derived histograms using kernel density estimation, $\langle\tilde{\alpha}\rangle^{\text{feasible}}$ (solid), $\langle\tilde{\alpha}\rangle^{\text{virtual}}$ (dashed), and $\langle\tilde{\alpha}\rangle^{\text{initial}}$ (dotted-dashed). All molecules were optimized using the `BFGS` algorithm. Molecular frameworks are displayed in Fig. 1.

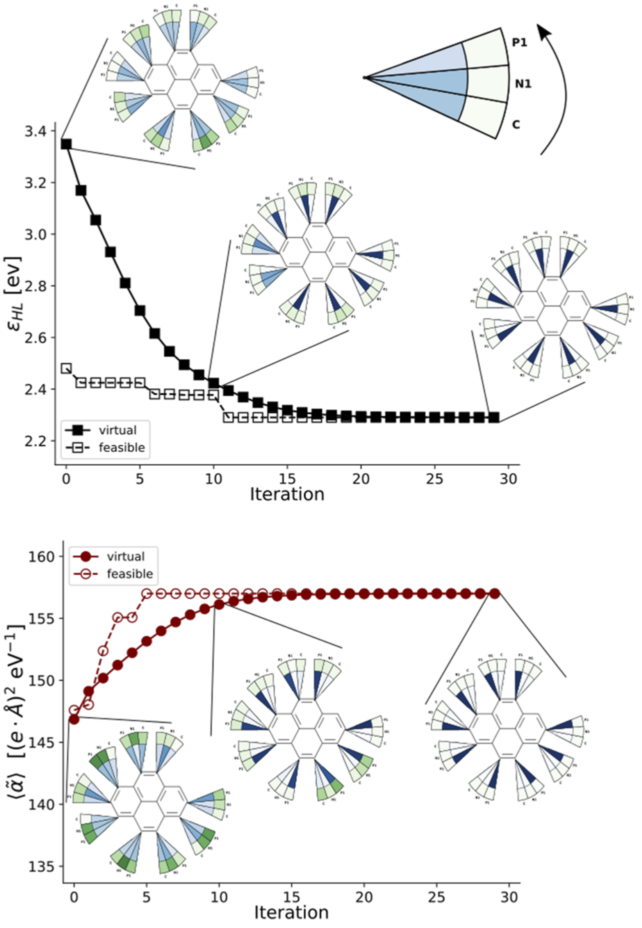

**FIG. 5.** Change in the parameters **b** for $\varepsilon_{HL}$ (left panel) and $\langle \tilde{\alpha} \rangle$ (right panel) during the optimization of a single random initial virtual molecule based on framework **3**. For the initial, one intermediate and the final **b**, we plot the values of **b** for molecular framework **3**. For both panels, the filled markers represent the values of $\varepsilon_{HL}$ and $\langle \tilde{\alpha} \rangle$ computed with **b** at each iteration (virtual), and the empty markers represent the observable values computed only with the most probable atoms given **b** at each iteration (feasible). For each search site, we only considered three different atom types, namely, C, N, and P; $\mathbf{b}_\ell = [b_\ell^{\mathrm{C}}, b_\ell^{\mathrm{N}}, b_\ell^{\mathrm{P}}]$. The optimization of $\varepsilon_{HL}$ and $\langle \tilde{\alpha} \rangle$ with respect to **b** was carried out with `Adam` using a learning rate of 0.2. More details are provided in the main text.

little dependence on the random initialization of **b** (Figs. 3 and 4). We also observe that, for the majority of the optimized molecules, $y_{virtual}$ and $y_{feasible}$ are linearly correlated, indicating that the optimization converged essentially to feasible molecules. The property distributions of $y_{\mathrm{virtual}}$ and $y_{\mathrm{feasible}}$ are reasonably close, even in the few cases when the correlations are poor.

Figure 6 displays the molecules with the lowest $\varepsilon_{HL}$ and maximum $\langle \tilde{\alpha} \rangle$ from the ensemble of different optimizations. First, we notice that there is a higher amount of phosphorus atoms in the molecules when $\langle \tilde{\alpha} \rangle$ was the target property. This is not unexpected as molecular polarizabilities, while not simply a sum of the atomic polarizabilities, are strongly influenced by the atomic polarizabilities of the constituent atoms.[56,57] As phosphorus is a third-row element in the same group with nitrogen and atomic polarizability increases significantly when increasing the row number, its atomic polarizability, both in free atoms[58] and in molecules,[59] is significantly larger than both nitrogen and carbon. Therefore, incorporating a large number of phosphorus atoms is expected to be a viable strategy to maximize the molecular polarizability in all of the molecular frameworks considered. For the molecules with the lowest $\varepsilon_{HL}$, we see extensive incorporation of both nitrogen and phosphorus atoms. This can be understood in terms of the effect of heteroatom substitution on $\varepsilon_{HL}$ within the HMO framework.[60] For alternant compounds, such as **1**–**8**, $\varepsilon_{HL}$ is unaffected by changing $\alpha_\ell$. The main effect comes from changing $\beta_{\ell,k}$ and is expected to be largest for bonds that feature a bonding interaction in the HOMO and an anti-bonding interaction in the LUMO. A lowered $\beta_{\ell,k}$ leads to decreased bonding interactions in the HOMO and consequently a higher HOMO energy. For the LUMO, decreasing the anti-bonding interactions by a lowered $\beta_{\ell,k}$ leads to a lowering of the energy. The net effect by raising the HOMO and lowering the LUMO is a decrease in the $\varepsilon_{HL}$ gap. We can, therefore, expect optimization to favor atom pairs with a low $\beta_{\ell,k}$ for bonds that feature a bonding interaction in the HOMO and an anti-bonding interaction in the LUMO. Inspection of the molecular orbitals of the optimized frameworks (cf. Fig. S1) indeed reveals that these bonds are dominantly between two N atoms, which feature the by far lowest $\beta_{\ell,k}$ at 0.159 (the next lowest is for P–P at 0.539). Control optimization with the original parameter set by van Catledge[10] instead gives molecules with P–P at those bonds (cf. Fig. S2), consistent with the fact that $\beta_{P,P} = 0.63$ is the lowest for this parameter set.

For this proof of principle work, we picked two distinct molecular target properties. However, based on the framework employed, a significant number of alternative properties could also be predicted and, thus, used for inverse design via gradient-based optimization. Additionally, any combined objective that is derived from multiple target properties can equally be optimized for via the same types of algorithms out of the box. This is particularly interesting for properties where Hückel models are known to provide reasonable prediction accuracies, such as HOMO–LUMO gaps. The use of gradient-based optimization algorithms enables fast convergence toward the closest local optimum solution reducing the number of evaluations and leading to significantly increased computation time. This is particularly important as one of the main bottlenecks in current approaches to inverse molecular design is the number of property evaluations needed to find an optimal structure.[61] Going beyond single-objective optimization, one possible extension of our presented approach is targeting multiple objectives via genuine gradient-based multi-objective optimization, for example, both $\varepsilon_{HL}$ and $\langle \tilde{\alpha} \rangle$. The standard approach to perform multi-objective optimization is via property concatenation into a single function to use standard single-objective algorithms, where algorithms like Bayesian optimization are used.[62–66] However, gradient-based multi-objective optimization algorithms[67] have been developed and they, together with automatic differentiation, could be employed for both parameter optimization and inverse molecular design in order to explore the corresponding Pareto fronts in a systematic manner. For more details about multi-objective optimization and the Pareto front, we refer the reader to Refs. 62, 66, and 68.

From a conceptual point of view, representing chemical structure subspaces in a parameterized form can greatly facilitate inverse

Published under an exclusive license by AIP Publishing

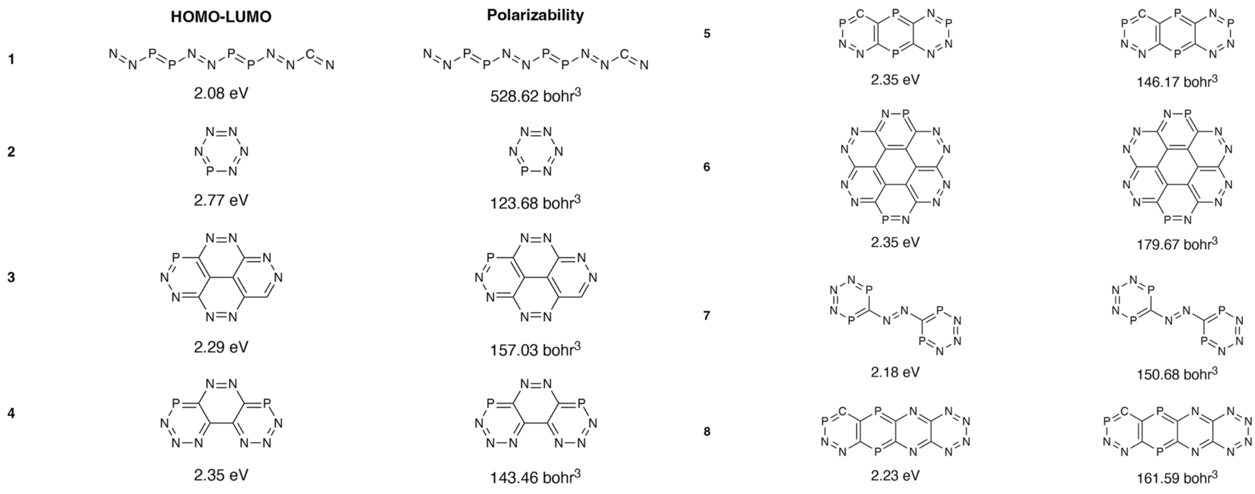

**FIG. 6.** The molecular structures with the lowest HOMO–LUMO gap and maximum polarizability for each of the eight molecules considered (Fig. 1).

design[16] as it allows the use of well-established approaches for parameter optimization to be used for the design of molecules. This is particularly effective when used in combination with AD due to its numerical stability and computational efficiency compared to alternative means to compute gradients. To put this in context, Xiao *et al.* estimated the speed-up of optimization over brute force enumeration to be ~$10^2$ for ten variable atoms and ~$10^3$ for 12 variable atoms.[16] Consequently, this also makes the molecular size that can still be feasibly treated in such an approach larger and thus, essentially, expands the chemical subspace the generative model can explore. However, one of the main downsides of the approach implemented in this work is the reliance on fixed molecular frameworks, which is common for alchemical formulations[69] strongly limiting the structural space considered in the optimization. Simple extensions would be (i) the combination of methods to change the molecular framework without relying on gradients with the method presented here to modify the atom identities within the respective framework or (ii) differentiable supermatrix structure where atom vacancies are allowed.[70] Ideally, future extensions should aim to find prudent ways allowing for framework modifications based on gradients as this potentially can lead to a dramatic reduction in the number of structure optimization steps and, thus, the number of property evaluations necessary. The extended Hückel model is also compatible with the proposed methodology, even with ML learned parameters,[38] by considering a description of the overlap integrals between different atom types, similar to Eq. (2).

## B. Parameter optimization

Another important task that might sometimes be underappreciated in computational chemistry is model parameter optimization. Here, we leverage the flexibility of AD and optimize all free parameters of the Hückel model in the same way as it is done for modern ML algorithms. Originally, the Hückel model is solely based on electronic interactions between nearest-neighbor atoms, which is typically also referred to as the tight-binding approximation

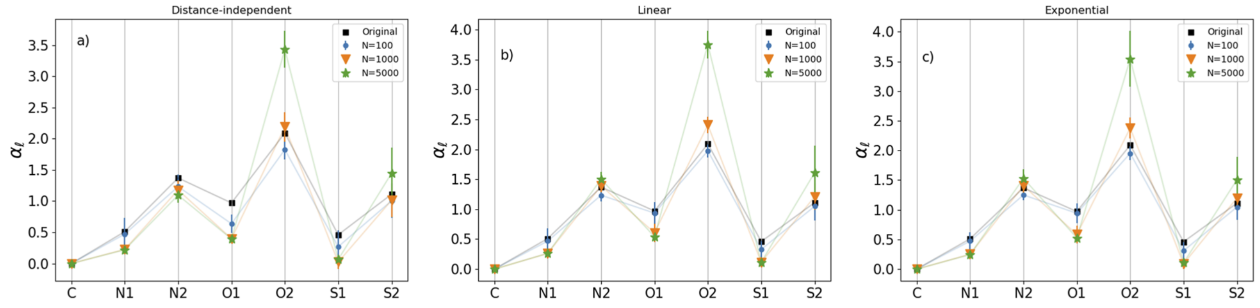

**FIG. 7.** Optimized $\alpha_\ell$ parameters for different Hückel models: (a) $\beta^{0}_{\ell,k}$, (b) $\beta^{lr}_{\ell,k}$ [Eq. (9)], and (c) $\beta^{\mathrm{exp}}_{\ell,k}$ [Eq. (8)]. All parameters reported were averaged over ten different training datasets and different numbers of training molecules. Colored symbols and bars represent the mean and standard deviation of the optimized parameters averaged over ten different datasets. The reference parameters (■-symbol) were taken from Ref. 10 and used as the initial parameters for all optimizations. We refer the reader to the main text for the optimization details.

Published under an exclusive license by AIP Publishing

[cf. $\beta_{\ell,k}$ parameters in Eq. (1)]. Beyond the standard Hückel model, one can introduce atomic distance-dependence of the corresponding interactions via $\beta_{\ell,k} = \beta^0_{\ell,k} g(\mathbf{R}_{\ell,k})$. For example, based on previous work by Longuett-Higgins and Salem,[71] $\beta_{\ell,k}$ has an exponential dependence on $\mathbf{R}_{\ell,k}$,

$$\beta^{\text{exp}}_{\ell,k} = -\beta^0_{\ell,k} \exp^{-\frac{\Delta R_{\ell,k}}{y_{\ell,k}}}. \tag{8}$$

A second functional form, which is based on the work of Su, Schrieffer, and Heeger[72,73] on conducting polymers, uses a linear distance-dependence of the interactions,

$$\beta^{lr}_{\ell,k} = -\beta^0_{\ell,k}\left(1 - y^{-1}_{\ell,k}\Delta R_{\ell,k}\right). \tag{9}$$

For both expressions [see Eqs. (8) and (9)], $\Delta R_{\ell,k}$ is the difference with respect to the reference bond length distance $R^0_{\ell,k}$, and $y_{\ell,k}$ is a length scale parameter. By including $R^0_{\ell,k}$ and $y_{\ell,k}$ in the set of parameters for the Hückel model, the complete set of parameters becomes $\boldsymbol{\theta} = [\alpha_\ell, \beta^0_{\ell,k}, y_{\ell,k}, R^0_{\ell,k}]$.

For this work, all initial $\alpha_\ell$ and $\beta^0_{\ell,k}$ parameters were taken from the work of van Catledge,[10] and the initial $R^0_{\ell,k}$ parameters were approximated from tables of standard bond lengths.[74] The length scale parameters ($y_{\ell,k}$) were initially set to 0.3 Å, which corresponds to the value that has been used for C–C in the literature.[75,76]

We used a subset of the GDB-13 dataset[77] that only consists of molecules with $\pi$-systems for fitting our model parameters (see the supplementary material for details on how the dataset was generated). Note that some molecules in the dataset could have n–$\pi^*$ transitions as their lowest excited state. We used a pool of 60 000 molecules, randomly sampled 100, 1000, and 5000 molecules from this set, and used 1000 additional molecules as the validation set

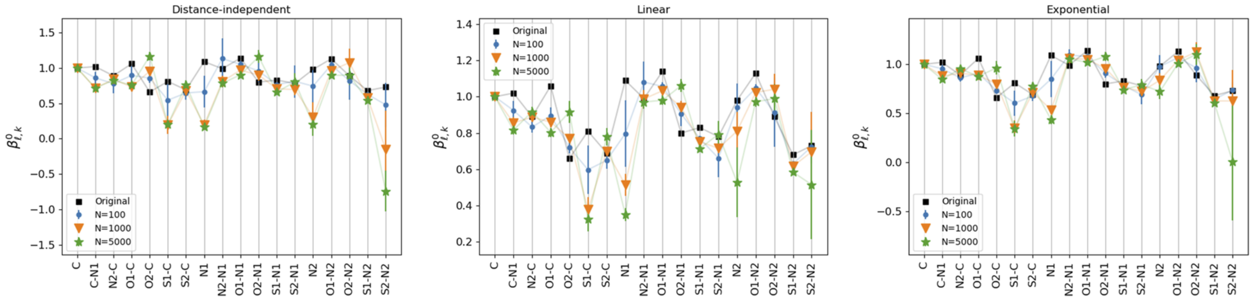

**FIG. 8.** Optimized $\beta^0_{\ell,k}$ parameters for different Hückel models: (a) $\beta^0_{\ell,k}$, (b) $\beta^{lr}_{\ell,k}$ [Eq. (9)], and (c) $\beta^{\text{exp}}_{\ell,k}$ [Eq. (8)]. All parameters reported were averaged over ten different training datasets and different numbers of training molecules. Colored symbols and bars represent the mean and standard deviation of the optimized parameters averaged over ten different datasets. The reference parameters (■-symbol) were taken from Ref. 10 and used as the initial parameters for all optimizations. We refer the reader to the main text for the optimization details.

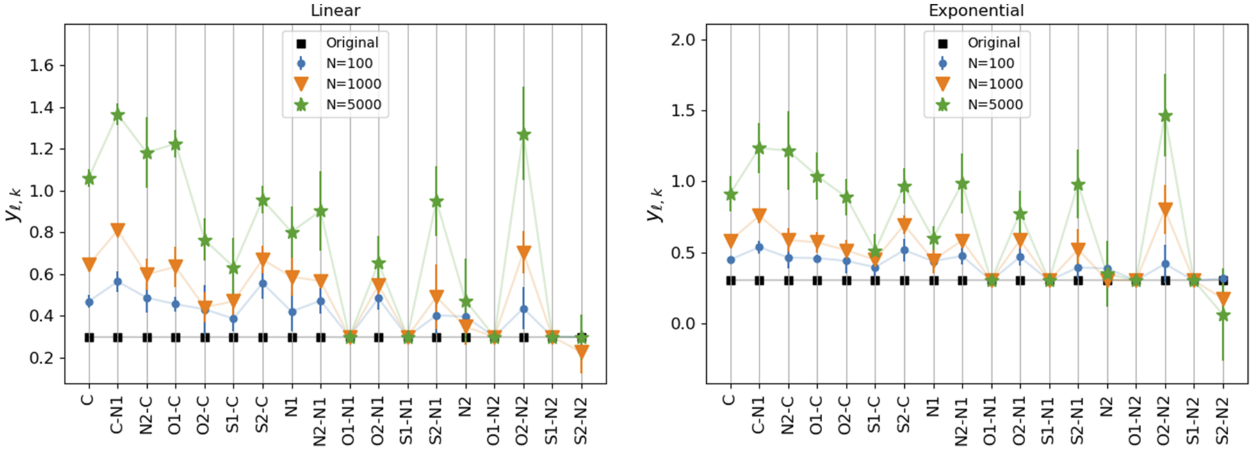

**FIG. 9.** Optimized $y_{\ell,k}$ parameters for different Hückel models: (a) $\beta^{lr}_{\ell,k}$ [Eq. (9)] and (b) $\beta^{\text{exp}}_{\ell,k}$ [Eq. (8)]. All parameters reported were averaged over ten different training datasets and different numbers of training molecules. Colored symbols and bars represent the mean and standard deviation of the optimized parameters averaged over ten different datasets. The reference parameters (■-symbol) were set to 0.3 Å and used as the initial parameters for all optimizations. We refer the reader to the main text for the optimization details.

Published under an exclusive license by AIP Publishing

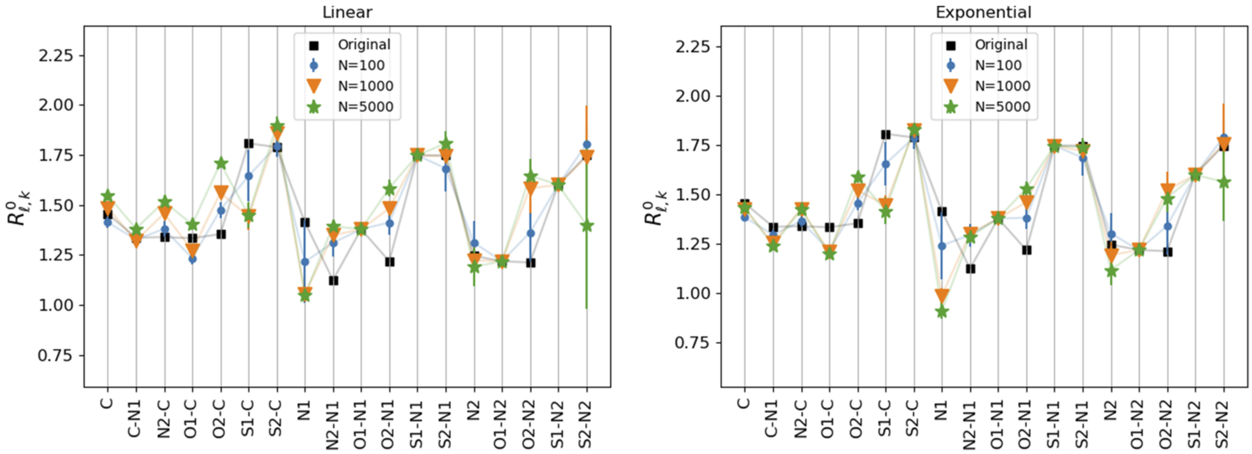

**FIG. 10.** Optimized $R^0_{\ell,k}$ parameters for different Hückel models: (a) $\beta^{lr}_{\ell,k}$ [Eq. (9)] and (b) $\beta^{exp}_{\ell,k}$ [Eq. (8)]. All parameters reported were averaged over ten different training datasets and different numbers of training molecules. Colored symbols and bars represent the mean and standard deviation of the optimized parameters averaged over ten different datasets. The reference parameters (■-symbol) were taken from Ref. 74 and used as the initial parameters for all optimizations. We refer the reader to the main text for the optimization details.

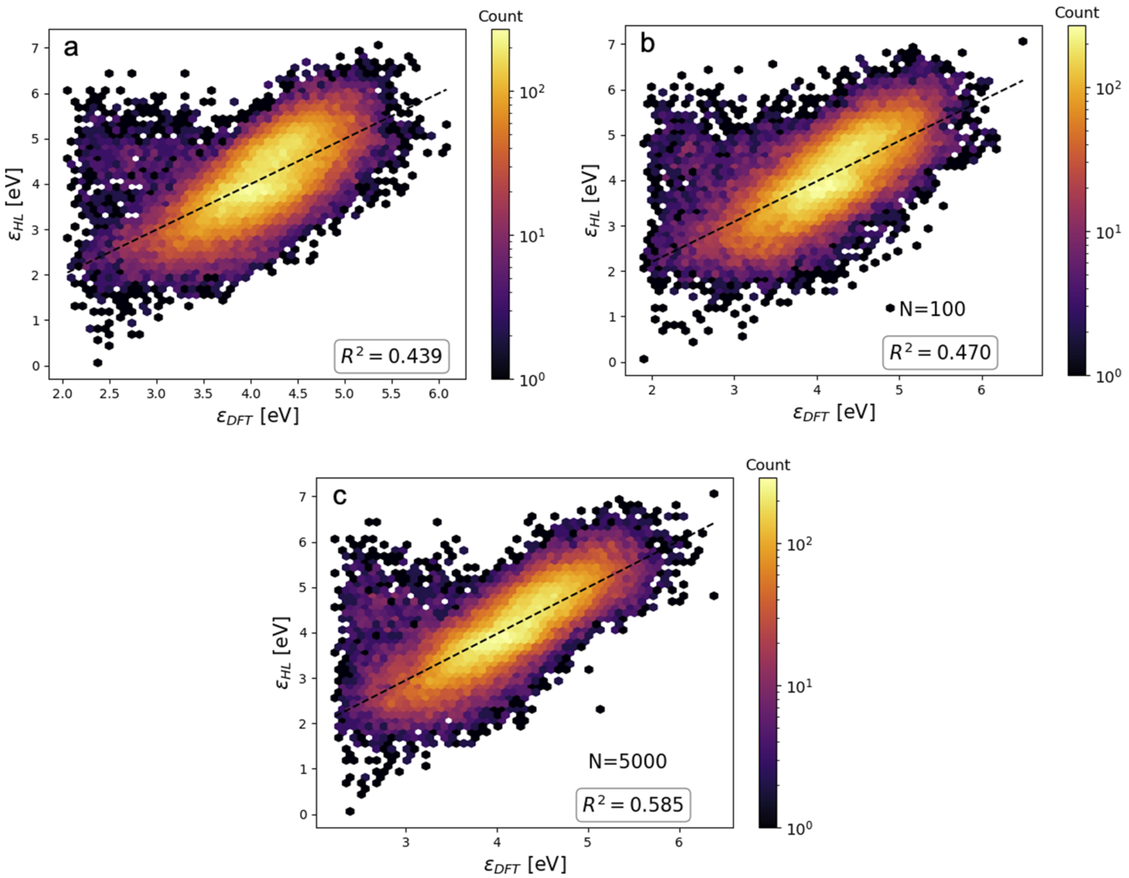

**FIG. 11.** $\varepsilon_{HL}$ predicted with the Hückel models and with the DFT level for 40k molecules not considered during training. The atom–atom interaction of the Hückel model is described by a distance-independent parameter, $\beta^0_{\ell,k}$. Results computed with a Hückel with parameters taken from the literature (a), and parameters optimized with N = 100 (b) and N = 5000 (c) data points. We refer the reader to the main text for the optimization details.

Published under an exclusive license by AIP Publishing

to monitor the optimization procedure. To optimize $\boldsymbol{\theta}$, we used the mean squared error as a loss function,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i}^{N} \left(\hat{\varepsilon}_{HL}(\mathcal{M}_i) - \varepsilon_{HL}(\boldsymbol{\theta};\mathcal{M}_i)\right)^2, \tag{10}$$

where $\mathcal{M}_i$ is a single molecule of the training set, and $\hat{\varepsilon}_{HL}$ is the vertical excitation energy between the ground state and the first excited singlet state computed at the TDA-SCS-$\omega$PBEPP86/def2-SVP level of theory.[78,79] At the Hückel level of theory, this excitation energy simply corresponds to the HOMO–LUMO gap ($\varepsilon_{HL}$) due to the disregard for electron correlation. To compare the prediction of $\varepsilon_{HL}$ with the DFT reference values properly, we linearly transformed the results of the Hückel model using two additional parameters, $w_0$ and $w_1$ (as in the below equation),

$$\varepsilon_{HL}(\boldsymbol{\theta};\mathcal{M}_i) = w_1 \times \varepsilon_{HL}(\alpha_\ell, \beta^0_{\ell,k}, y_{\ell,k}, R^0_{\ell,k};\mathcal{M}_i) + w_0, \tag{11}$$

where $\boldsymbol{\theta}$ jointly represents all parameters of the model, i.e., $\boldsymbol{\theta} = [\alpha_\ell, \beta^0_{\ell,k}, y_{\ell,k}, R^0_{\ell,k}, w_0, w_1]$.

For the optimization of all free parameters, we used the `AdamW` optimization algorithm,[54] as implemented in the `Optax` library,[80] with a learning rate of 0.02 and a weight decay of $10^{-4}$. Notably, we considered various training scenarios that included different values for the weight decay and the regularization of different sets of the Hückel parameters. However, we found no impact on the accuracy of the model. The initial model parameters were gathered from Refs. 10 and 74.

We optimized the parameters of three different Hückel models, (i) the original one where $\beta_{\ell,k}$ is distance-independent ($\beta_{\ell,k} = \beta^0_{\ell,k}$) and both (ii) the exponential [Eq. (8)] and (iii) the linear [Eq. (9)] distance-dependence functional forms. We want to emphasize that any other analytic form for $\beta_{\ell,k}$ could be considered and AD makes any of these expressions fully differentiable. Following the convention in the literature, we scaled the parameters $\beta^0_{\ell,k}$ and $\alpha_\ell$ with respect to the carbon atom parameters according to $\alpha_\ell = \alpha_\ell - \alpha_C$ and $\beta^0_{\ell,k} = \beta^0_{\ell,k}/\beta^0_{C,C}$. Notably, at least for our results in this work, we found that including a regularization term in the loss function did not impact the accuracy of the model. Finally, we found 20 epochs

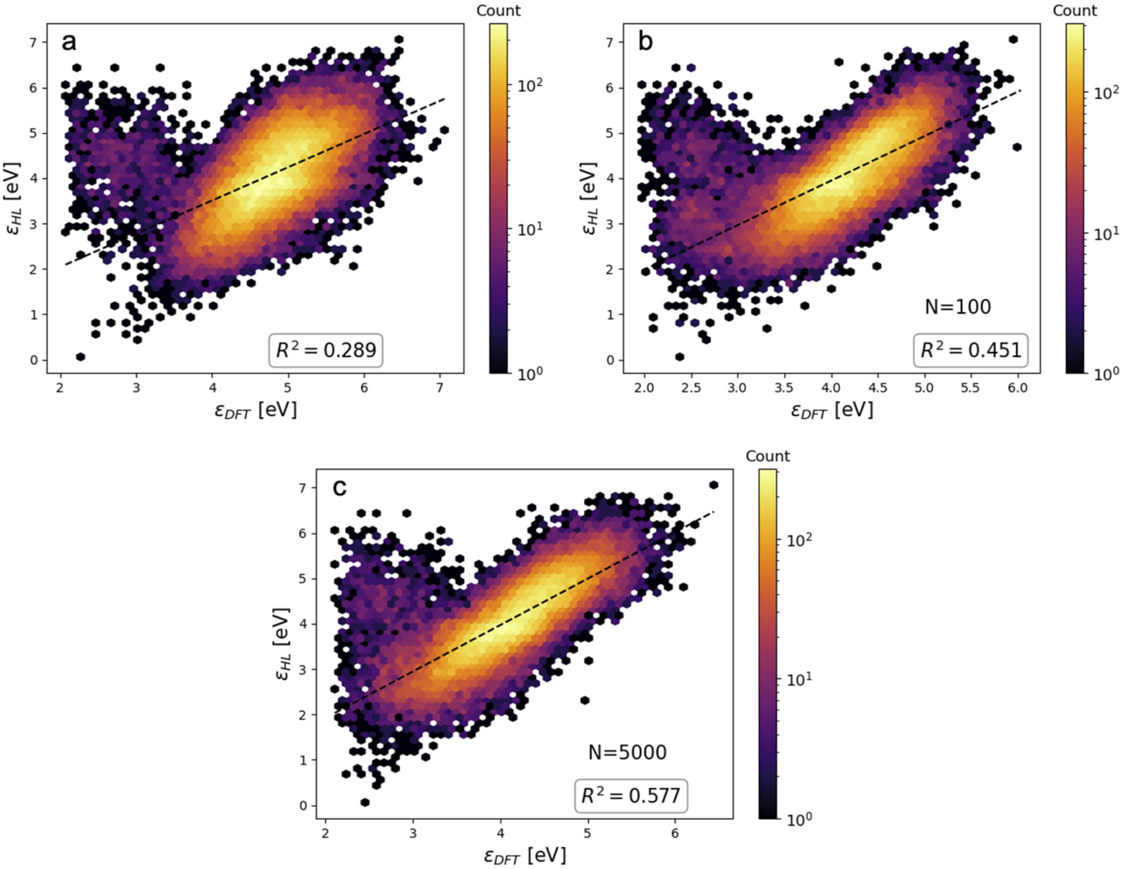

**FIG. 12.** $\varepsilon_{HL}$ predicted with the Hückel models and with the DFT level for 40k molecules not considered during training. The atom–atom interaction of the Hückel model is described by a distance-dependent parameter, $\beta^{lr}_{\ell,k}$ [Eq. (9)]. Results computed with a Hückel with parameters taken from the literature (a), and parameters optimized with N = 100 (b) and N = 5000 (c) data points. We refer the reader to the main text for the optimization details.

Published under an exclusive license by AIP Publishing

to be enough to minimize the loss function when the parameters are initialized with values from Refs. 10 and 74.

In Figs. 7–10, we display the optimized values of the parameters for the three different Hückel models considered. From the optimized parameters, we observe that $\alpha_O$ (i.e., the 2p orbital energy parameter for oxygen), for all three models, is the one that differs the most from the literature.[10,74] While there are no good reference data for $y_{\ell,k}$ to compare to, we observe, nevertheless, that the C–C parameter value changes considerably from the initial value of 0.3. For the $\beta^0_{\ell,k}$ parameters, only the values for N–C resemble the literature values. Furthermore, the optimal values of $R^0_{\ell,k}$ change the least from the values in Ref. 74.

Using these optimized parameters, we predicted $\varepsilon_{HL}$ for 40 000 additional test set molecules and compared the results with DFT reference data. The direct comparison is depicted in Figs. 11–13. By optimizing all parameters there is a significant improvement in the prediction of $\varepsilon_{HL}$ using our semiempirical model. Notably, we also found that considering a larger training dataset ($N > 25$ K) does not have a significant impact in the accuracy of the Hückel model, which suggests that either the corresponding molecules do not provide any additional information with respect to the relevant interaction parameters or that the model already is close to its best expected performance and cannot be improved further. These results justify the modern approaches where learning parameters for semiempirical methods is carried out with ML models.[36–40] Another important observation in that regard is that the analytical form of $g(\mathbf{R}_{\ell,k})$ in $\beta_{\ell,k}$ does not impact the accuracy of the model when optimized parameters are used. This suggests that none of these analytical forms of $\beta_{\ell,k}$ sufficiently impact the underlying physics relevant for predicting excitation energies in these molecules. Hence, we speculate that non-nearest neighbor $\pi$–electron interactions or an explicit account of $\sigma$–$\pi$ interactions as in extended Hückel methods would be key for more accurate models.

Next, we also optimized the parameters with respect to the polarizability for the distance-independent Hückel model. Here, the gradients needed for training are of third-order, e.g., $\frac{\partial}{\partial \alpha_\ell}\frac{\partial^2 \tilde{\alpha}_{kk}}{\partial F_k^2}$ or

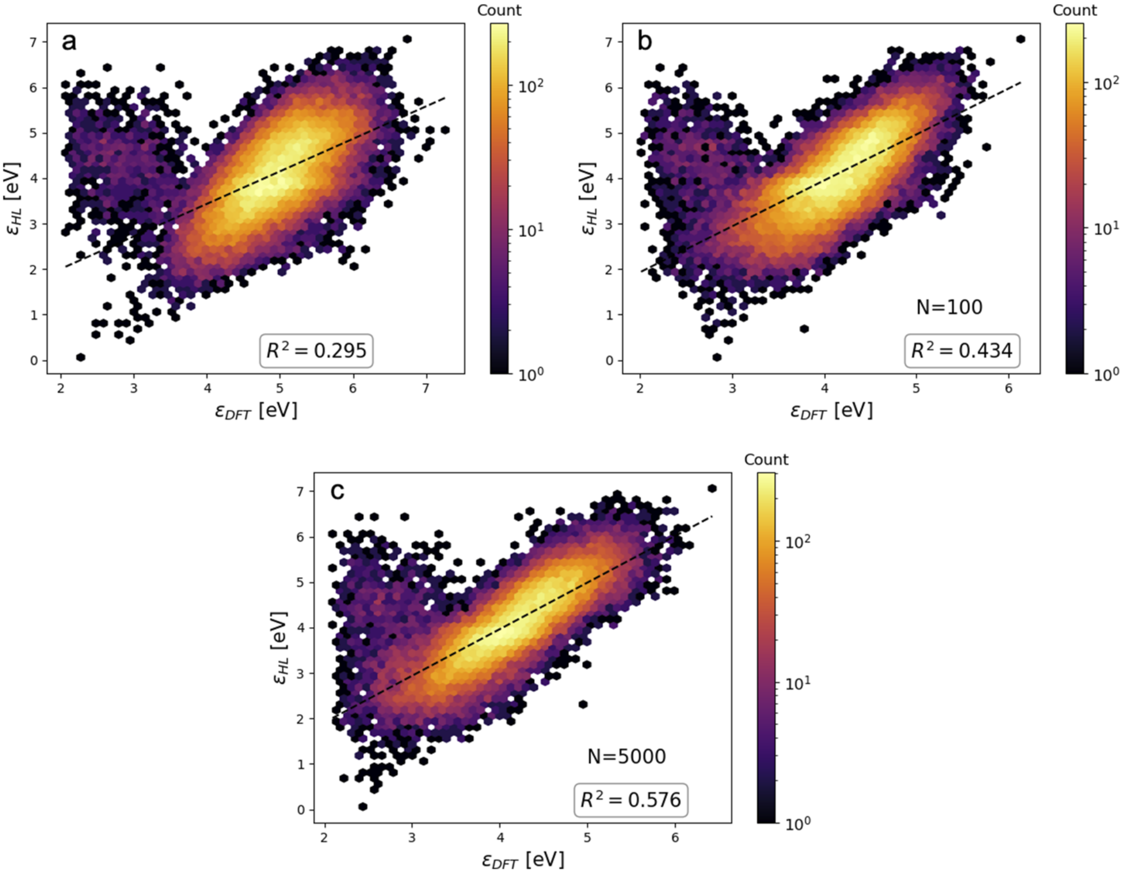

**FIG. 13.** $\varepsilon_{HL}$ predicted with the Hückel models and with the DFT level for 40k molecules not considered during training. The atom–atom interaction of the Hückel model is described by a distance-dependent parameter, $\beta^{\text{exp}}_{\ell,k}$ [Eq. (8)]. Results computed with a Hückel model with parameters taken from the literature (a), and parameters optimized with N = 100 (b) and N = 5000 (c) data points. We refer the reader to the main text for the optimization details.

Published under an exclusive license by AIP Publishing

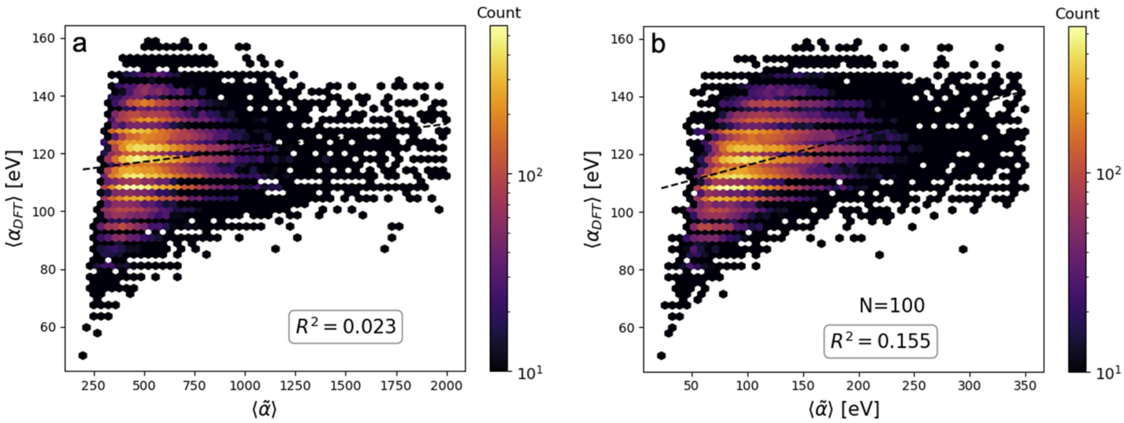

**FIG. 14.** $\langle \tilde{\alpha} \rangle$ predicted with the Hückel models and with the DFT level for 40k molecules not considered during training. The atom–atom interaction of the Hückel model is described by a distance-independent parameter, $\beta^0_{\ell,k}$. Results computed with a Hückel model with parameters taken from the literature (a) and (b) parameters optimized with N = 100 data points. We refer the reader to the main text for the optimization details.

$\frac{\partial}{\partial \beta_{k,\ell}} \frac{\partial^2 \tilde{\alpha}_{kk}}{\partial F_k^2}$, and can be computed more efficiently via AD, illustrating the potential of this approach. The molecular polarizabilities that were used as reference data were computed using `dftd4` (version 3.4.0)[81–83] via the default methodology summing atomic polarizabilities. Even though we observe a higher prediction accuracy when the optimized parameters are used compared to the model before parameter refinement, as depicted in Fig. 14, the accuracy of the model still remains relatively low and does not improve anymore when more training data are used. We suspect that this prediction task is particularly challenging for the Hückel model as the simulated polarizability only corresponds to the contribution from $\pi$-electrons, while that of the reference data accounts for all the electrons in the molecules. While we expect a significant portion of the molecular polarizabilities to stem from the $\pi$-electrons, the contributions of the $\sigma$-electrons cannot be neglected and can dominate this property. Nevertheless, this proof-of-concept application example demonstrates the operational ease of conducting parameter refinement of a given physics-based prediction model based on reference data, even when derivative properties are targeted.

## IV. CONCLUSIONS

In this work, we demonstrate the power of automatic differentiation to enable the efficient use of physics-inspired models for gradient-based optimization problems in the realm of molecular chemistry via semiempirical Hückel models. In particular, we showcase inverse molecular design via an alchemical problem formulation using fixed molecular frameworks. This allows us to perform structure optimization requiring only a very small number of intermediate structures to find local minima with respect to the properties of interest utilizing gradients with respect to atom identities at specific sites. While our approach is currently limited to a fixed molecular framework, performing optimizations over the molecular composition space alone is far from trivial. Compared to various alternative approaches, our implementation shows a remarkably high molecular sampling efficiency due to efficient utilization of gradient information in combination with powerful gradient-based optimization algorithms. Additionally, we showcase the ease of generating calibrated physics-based property prediction models using high quality reference training data of relatively modest size, again allowing for quick convergence of model parameters. This is particularly important as most physical models that rely on empirical parameters, such as semiempirical quantum chemistry models and density functional approximations, are still largely optimized by hand, making the corresponding procedures tedious. Compared to emerging approaches where neural networks are used to dynamically predict semiempirical parameters on a per-molecule basis,[38] our "static" parameterization has the potential of a broader applicability domain,[84] which is crucial for inverse design. Thus, we believe that our work will serve as an inspiration for the field of computational chemistry in order to adopt the readily available AD capabilities of mature ML programming frameworks allowing us to accelerate the construction of ever more accurate physics-based property simulation models.

## SUPPLEMENTARY MATERIAL

The data that support the findings of this study are available within the supplementary material.

## ACKNOWLEDGMENTS

R.P. acknowledges funding through a Postdoc.Mobility fellowship by the Swiss National Science Foundation (SNSF, Project No. 191127). A.A.-G. thanks Anders G. Frøseth for his generous support. A.A.-G. acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program. We also thank the SciNet HPC Consortium for support regarding the use of the Niagara supercomputer. SciNet is funded by the Canada Foundation for Innovation, the Government of Ontario, Ontario Research Fund-Research Excellence, and the University of Toronto.

Published under an exclusive license by AIP Publishing

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Rodrigo A. Vargas–Hernández**: Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Kjell Jorner**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Robert Pollice**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Alán Aspuru–Guzik**: Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material and in GitHub (inverse molecular design) https://github.com/RodrigoAVargasHdz/huxel_molecule_desing and (parameter optimization) https://github.com/RodrigoAVargasHdz/huxel.

## REFERENCES

[1] R. S. Mulliken, "Molecular scientists and molecular science: Some reminiscences," J. Chem. Phys. **43**, S2–S11 (1965).

[2] P.-O. Löwdin, "Present situation of quantum chemistry," J. Phys. Chem. **61**, 55–68 (1957).

[3] W. Thiel, "Semiempirical quantum–chemical methods," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**(2), 145–157 (2014).

[4] E. Hückel, "Quantentheoretische beiträge zum benzolproblem," Z. Phys. **70**, 204–286 (1931).

[5] E. Hückel, "Quanstentheoretische beiträge zum benzolproblem. II," Z. Phys. **72**, 310–337 (1931).

[6] E. Hückel, "Quantentheoretische beiträge zum problem der aromatischen und ungesättigten verbindungen. III," Z. Phys. **76**, 628–648 (1932).

[7] E. Hückel, "Die freien radikale der organischen chemie. IV," Z. Phys. **83**, 632–668 (1933).

[8] A. Streitwieser, *Molecular Orbital Theory for Organic Chemists* (Wiley, New York, 1961).

[9] B. Andes Hess, L. J. Schaad, and C. W. Holyoke, "On the aromaticity of annulenones," Tetrahedron **28**, 5299–5305 (1972).

[10] F. A. van Catledge, "A Pariser-Parr-Pople-based set of Hueckel molecular orbital parameters," J. Org. Chem. **45**, 4801–4802 (1980).

[11] I. Fleming, *Molecular Orbitals and Organic Chemical Reactions: Reference Edition* (Wiley, Chichester, United Kingdom, 2010).

[12] P. Klán and J. Wirz, *Photochemistry of Organic Compounds: From Concepts to Practice* (Wiley, 2009).

[13] C. S. Anstöter and J. R. R. Verlet, "A Hueckel model for the excited-sate dynamics of a protein chromophore developed using photoelectron imaging," Acc. Chem. Res. **55**(9), 1205–1213 (2022).

[14] J. A. N. F. Gomes and R. B. Mallion, "Aromaticity and ring currents," Chem. Rev. **101**(5), 1349–1384 (2001).

[15] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," Science **361**, 360–365 (2018).

[16] D. Xiao, W. Yang, and D. N. Beratan, "Inverse molecular design in a tight-binding framework," J. Chem. Phys. **129**(4), 044106 (2008).

[17] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," J. Mach. Learn. Res. **18**, 1–43 (2018); available at https://jmlr.org/papers/v18/17-468.html.

[18] T. Tamayo-Mendoza, C. Kreisbeck, R. Lindh, and A. Aspuru-Guzik, "Automatic differentiation in quantum chemistry with applications to fully variational Hartree–Fock," ACS Cent. Sci. **4**(5), 559–566 (2018).

[19] X. Zhang and G. K.-L. Chan, "Differentiable quantum chemistry with PySCF for molecules and materials at the mean-field level and beyond," J. Chem. Phys. **157**, 204801 (2022).

[20] U. Ekström, L. Visscher, R. Bast, A. J. Thorvaldsen, and K. Ruud, "Arbitrary-order density functional response theory from automatic differentiation," J. Chem. Theory Comput. **6**(7), 1971–1980 (2010).

[21] M. F. Kasim, S. Lehtola, and S. M. Vinko, "DQC: A python program package for differentiable quantum chemistry," J. Chem. Phys. **156**, 084801 (2021).

[22] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, "Kohn-Sham equations as regularizer: Building prior knowledge into machine-learned physics," Phys. Rev. Lett. **126**, 036401 (2021).

[23] M. F. Kasim and S. M. Vinko, "Learning the exchange-correlation functional from nature with fully differentiable density functional theory," Phys. Rev. Lett. **127**, 126403 (2021).

[24] L. Zhao and E. Neuscamman, "Excited state mean-field theory without automatic differentiation," J. Chem. Phys. **152**(20), 204112 (2020).

[25] A. Yachmenev and S. N. Yurchenko, "Automatic differentiation method for numerical construction of the rotational-vibrational Hamiltonian as a power series in the curvilinear internal coordinates using the Eckart frame," J. Chem. Phys. **143**(1), 014105 (2015).

[26] A. S. Abbott, B. Z. Abbott, J. M. Turney, and H. F. Schaefer, "Arbitrary-order derivatives of quantum chemical methods via automatic differentiation," J. Phys. Chem. Lett. **12**(12), 3232–3239 (2021).

[27] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, "PennyLane: Automatic differentiation of hybrid quantum-classical computations," arXiv:1811.04968 (2020).

[28] F. Pavošević and S. Hammes-Schiffer, "Automatic differentiation for coupled cluster methods," arXiv:2011.11690 (2020).

[29] G. F. von Rudorff, "Arbitrarily accurate quantum alchemy," J. Chem. Phys. **155**(22), 224103 (2021).

[30] N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev, and S. Tretiak, "Extending machine learning beyond interatomic potentials for predicting molecular properties," Nat. Rev. Chem. **6**, 653 (2022).

[31] R. A. Vargas-Hernández, R. T. Q. Chen, K. A. Jung, and P. Brumer, "Fully differentiable optimization protocols for non-equilibrium steady states," New J. Phys. **23**, 123006 (2021).

[32] R. A. Vargas-Hernández, R. T. Q. Chen, K. A. Jung, and P. Brumer, "Inverse design of dissipative quantum steady-states with implicit differentiation," arXiv:2011.12808 (2020).

[33] N. Yoshikawa and M. Sumita, "Automatic differentiation for the direct minimization approach to the Hartree–Fock method," J. Phys. Chem. A **126**(45), 8487–8493 (2022).

[34] S. Bac, A. Patra, K. J. Kron, and S. Mallikarjun Sharada, "Recent advances toward efficient calculation of higher nuclear derivatives in quantum chemistry," J. Phys. Chem. A **126**(43), 7795–7805 (2022).

[35] K. Inui and Y. Motome, "Inverse Hamiltonian design by automatic differentiation," arXiv:2203.07157 (2022).

Published under an exclusive license by AIP Publishing

[36] R. Steiger, C. H. Bischof, B. Lang, and W. Thiel, "Using automatic differentiation to compute derivatives for a quantum-chemical computer program," Future Gener. Comput. Syst. **21**(8), 1324–1332 (2005).

[37] G. Zhou, B. Nebgen, N. Lubbers, W. Malone, A. M. N. Niklasson, and S. Tretiak, "Graphics processing unit-accelerated semiempirical Born Oppenheimer molecular dynamics using PyTorch," J. Chem. Theory Comput. **16**, 4951–4962 (2020).

[38] T. Zubatiuk, B. Nebgen, N. Lubbers, J. S. Smith, R. Zubatyuk, G. Zhou, C. Koh, K. Barros, O. Isayev, and S. Tretiak, "Machine learned Hückel theory: Interfacing physics and deep neural networks," J. Chem. Phys. **154**(24), 244108 (2021).

[39] C. H. Pham, R. K. Lindsey, L. E. Fried, and N. Goldman, "High-accuracy semiempirical quantum models based on a minimal training set," J. Phys. Chem. Lett. **13**(13), 2934–2942 (2022).

[40] G. Zhou, N. Lubbers, K. Barros, S. Tretiak, and B. Nebgen, "Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics," Proc. Natl. Acad. Sci. U. S. A. **119**(27), e2120333119 (2022).

[41] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).

[42] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, and S. Wanderman-Milne (2018). "JAX: Composable transformations of Python+NumPy programs," Github. http://github.com/google/jax.

[43] B. Ramsundar, D. Krishnamurthy, and V. Viswanathan, "Differentiable physics: A position piece," arXiv:2109.07573 (2021).

[44] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," Software available from tensorflow.org, 2015

[45] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada* (Curran Associates, Inc., 2019), Vol. 32; available at https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[46] K. Yates, in *Hückel Molecular Orbital Theory*, edited by K. Yates (Academic Press, 1978), pp. 27–87.

[47] J. Pople and D. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill Series in Advanced Chemistry (McGraw-Hill, 1970).

[48] H. D. Cohen and C. C. J. Roothaan, "Electric dipole polarizability of atoms by the Hartree-Fock method. I. Theory for closed-shell systems," J. Chem. Phys. **43**(10), S34–S39 (1965).

[49] J. M. André, J. Delhalle, J. G. Fripiat, G. Hennico, and L. Piela, "Calculations of molecular polarizabilities from electric-field-variant atomic orbitals: An analysis of the problem and its application to the hydrogen molecule and to the alkane series," Int. J. Quantum Chem. **34**(S22), 665–678 (1988).

[50] H. A. Kurtz, J. J. P. Stewart, and K. M. Dieter, "Calculation of the nonlinear optical properties of molecules," J. Comput. Chem. **11**(1), 82–87 (1990).

[51] D. Jacquemin, B. Champagne, and J. M. André, "Electron correlation effects upon the static (hyper)polarizabilities of push-pull conjugated polyenes and polyynes," Int. J. Quantum Chem. **65**(5), 679–688 (1997).

[52] H. A. Kurtz and D. S. Dudis, "Quantum mechanical methods for predicting nonlinear optical properties," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (John Wiley & Sons, Hoboken, NJ, 2007), pp. 241–279.

[53] M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert, "Efficient and modular implicit differentiation," arXiv:2105.15183 (2021).

[54] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in International Conference on Learning Representations, 2019.

[55] Vector of length *d* where all its entries are equal to 0 except one that has a value of 1, e.g., [0000100] or [0100].

[56] F. Eisenlohr, "Eine neuberechnung der atomrefraktionen. I," Z. Phys. Chem. **75U**(1), 585–607 (1911).

[57] A. L. von Steiger, "Ein beitrag zur summationsmethodik der molekularrefraktionen, besonders bei aromatischen kohlenwasserstoffen," Ber. Dtsch. Chem. Ges. (A, B Ser.) **54**(6), 1381–1393 (1921).

[58] P. Schwerdtfeger and J. K. Nagle, "2018 table of static dipole polarizabilities of the neutral elements in the periodic table," Mol. Phys. **117**(9-12), 1200–1225 (2019).

[59] A. Krawczuk, D. Pérez, and P. Macchi, "PolaBer: A program to calculate and visualize distributed atomic polarizabilities based on electron density partitioning," J. Appl. Crystallogr. **47**, 1452–1458 (2014).

[60] E. Heilbronner and H. Bock, *The HMO Model and its Application: Basis and Manipulation* (Wiley, 1976).

[61] W. Gao, T. Fu, J. Sun, and C. W. Coley, "Sample efficiency matters: A benchmark for practical molecular optimization," arXiv:2206.12411 (2022).

[62] F. Häse, L. M. Roch, and A. Aspuru-Guzik, "Chimera: Enabling hierarchy based multi-objective optimization for self-driving laboratories," Chem. Sci. **9**, 7642–7655 (2018).

[63] R. J. Hickman, M. Aldeghi, F. Häse, and A. Aspuru-Guzik, "Bayesian optimization with known experimental and design constraints for chemistry applications," Digital Discovery **1**, 732 (2022).

[64] M. Seifrid, R. J. Hickman, A. Aguilar-Granda, C. Lavigne, J. Vestfrid, T. C. Wu, T. Gaudin, E. J. Hopkins, and A. Aspuru-Guzik, "Routescore: Punching the ticket to more efficient materials development," ACS Cent. Sci. **8**(1), 122–131 (2022).

[65] H. Tao, T. Wu, S. Kheiri, M. Aldeghi, A. Aspuru-Guzik, and E. Kumacheva, "Self-driving platform for metal nanoparticle synthesis: Combining microfluidics and machine learning," Adv. Funct. Mater. **31**(51), 2106725 (2021).

[66] R. A. Vargas-Hernández, C. Chuang, and P. Brumer, "Multi-objective optimization for retinal photoisomerization models with respect to experimental observables," J. Chem. Phys. **155**(23), 234109 (2021).

[67] J.-A. Désidéri, "Multiple-gradient descent algorithm (MGDA) for multiobjective optimization," C. R. Math. **350**(5), 313–318 (2012).

[68] M. Kaisa, *Nonlinear Multiobjective Optimization*, Volume 12 of International Series in Operations Research and Management Science (Kluwer Academic Publishers, Boston, 1999).

[69] G. F. von Rudorff and O. A. von Lilienfeld, "Alchemical perturbation density functional theory," Phys. Rev. Res. **2**, 023220 (2020).

[70] C. Shen, M. Krenn, S. Eppel, and A. Aspuru-Guzik, "Deep molecular dreaming: Inverse machine learning for *de-novo* molecular design and interpretability with surjective representations," Mach. Learn.: Sci. Technol. **2**, 03LT02 (2021).

[71] H. Longuet-Higgins and L. Salem, "The alternation of bond lengths in long conjugated chain molecules," Proc. R. Soc. London, Ser. A **251**, 172–185 (1959).

[72] W. P. Su, J. R. Schrieffer, and A. J. Heeger, "Solitons in polyacetylene," Phys. Rev. Lett. **42**, 1698–1701 (1979).

[73] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W.-P. Su, "Solitons in conducting polymers," Rev. Mod. Phys. **60**, 781–850 (1988).

[74] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, and R. Taylor, "Tables of bond lengths determined by X-ray and neutron diffraction. I. Bond lengths in organic compounds," J. Chem. Soc., Perkin Trans. 2 **1987**, S1–S19.

[75] L. Z. Stolarczyk and T. M. Krygowski, "Augmented Hueckel molecular orbital model of $\pi$-electron systems: From topology to metric. I. General theory," J. Phys. Org. Chem. **34**(3), e4154 (2021).

[76] J. H. Kwapisz and L. Z. Stolarczyk, "Applications of Hückel-Su-Schrieffer-Heeger method," Struct. Chem. **32**, 1393–1406 (2021).

[77] L. C. Blum and J.-L. Reymond, "970 million druglike small molecules for virtual screening in the chemical universe database GDB-13," J. Am. Chem. Soc. **131**(25), 8732–8733 (2009).

[78] M. Casanova-Páez and L. Goerigk, "Time-dependent long-range-corrected double-hybrid density functionals with spin-component and spin-opposite scaling: A comprehensive analysis of singlet–singlet and singlet–triplet excitation energies," J. Chem. Theory Comput. **17**(8), 5165–5186 (2021).

[79] F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. **7**, 3297–3305 (2005).

[80] .M. Hessel, D. Budden, F. Viola, M. Rosca, E. Sezener, and T. Hennigan (2020), "Optax: Composable gradient transformation and optimisation, in JAX!," Github. http://github.com/google/jax

[81] E. Caldeweyher, C. Bannwarth, and S. Grimme, "Extension of the D3 dispersion coefficient model," J. Chem. Phys. **147**(3), 034112 (2017).

[82] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-charge dependent London dispersion correction," J. Chem. Phys. **150**(15), 154122 (2019).

[83] E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme, "Extension and evaluation of the D4 London-dispersion model for periodic systems," Phys. Chem. Chem. Phys. **22**, 8499–8512 (2020).

[84] T. Hanser, C. Barber, J. F. Marchaland, and S. Werner, "Applicability domain: Towards a more formal definition," SAR QSAR Environ. Res. **27**, 865–881 (2016).

Published under an exclusive license by AIP Publishing