Combining *First-Principles* and Data Modeling for the Accurate Prediction of the Refractive Index of Organic Polymers

Mohammad Atif Faiz Afzal,[1,*] Chong Cheng,[1] and Johannes Hachmann[1,2,3,†]

[1] *Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, United States*
[2] *Computational and Data-Enabled Science and Engineering Graduate Program, University at Buffalo, The State University of New York, Buffalo, NY 14260, United States*
[3] *New York State Center of Excellence in Materials Informatics, Buffalo, NY 14203, United States*

Organic materials with a high index of refraction (RI) are attracting considerable interest due to their potential application in optic and optoelectronic devices. However, most of these applications require an RI value of 1.7 or larger, while typical carbon-based polymers only exhibit values in the range of 1.3–1.5. This paper introduces an efficient computational protocol for the accurate prediction of RI values in polymers to facilitate *in silico* studies that can guide the discovery and design of next-generation high-RI materials. Our protocol is based on the Lorentz-Lorenz equation and is parametrized by the polarizability and number density values of a given candidate compound. In the proposed scheme, we compute the former using *first-principles* electronic structure theory and the latter using an approximation based on van der Waals volumes. The critical parameter in the number density approximation is the packing fraction of the bulk polymer, for which we have devised a machine learning model. We demonstrate the performance of the proposed RI protocol by testing its predictions against the experimentally known RI values of 112 optical polymers. Our approach to combine *first-principles* and data modeling emerges as both a successful and highly economical path to determining the RI values for a wide range of organic polymers.

I. INTRODUCTION

Organic small molecules, oligomers, and polymers are emerging materials and of significant interest for numerous fields of application due to their unique or otherwise desirable properties [1]. Unlike most conventional inorganic materials, they are generally flexible, light-weight, mechanically stable on impact, easy to process, and inexpensive to produce [2, 3]. Perhaps most importantly, their properties can be tailored towards specific demands by controlling their molecular structure [4]. A particular area of interest is the application of organic materials in optic and optoelectronic devices [5], such as (image) sensors [6, 7], displays [8], and light sources (including organic light-emitting diodes) [9], in which they can be introduced *in situ* as microlenses [10], waveguides [11], microresonators [12], interferometers [13], anti-reflective coatings [14], optical adhesives [15], and substrates [16]. Some of the optical properties that are relevant for these applications are the refractive index (RI), Abbe number, birefringence, absorption spectrum, and color [17].

The RI value dictates the shape and size of many optical components, in particular those with lens function. Most of the aforementioned applications require materials with large RI values (i.e., larger than 1.7), and there are several applications that require very large ones (i.e., larger than 1.8) [18]. Unfortunately, the vast majority of organic polymers only offer RI values ranging from 1.3 to 1.5 [4] (compared to inorganic materials, which can feature values up to ∼4). The development of high-RI polymers has thus gained attention, and several approaches have been proposed to overcome the RI-value limitations of typical organic polymers. They include the notion to incorporate highly polarizable moieties, such as rigid aromatic fragments [19], heteroatoms [20, 21], or organometallics [22], into the polymer scaffold. Another strategy that has been pursued is to reinforce the polymer matrix with metal alkoxides (e.g., TiO_2, Fe_3O_4) [23, 24] or other high-RI molecules (e.g., ZnS, diamondoids) [25, 26]. While these approaches have resulted in a few systems with RI values between 1.6 and 1.8, most of them are of limited utility for practical applications due to a variety of materials, processability, or preparation issues [4]. Increasing the RI values of organic polymers beyond 1.8 has remained a completely elusive task and continues to be an important challenge in synthetic chemistry [20].

The traditional, experimentally focused discovery process for new materials is very time-, labor-, and resource-intensive, which limits the number and diversity of candidate compounds that can be explored. Progress thus tends to be slow and incremental, in particular for advanced materials systems, which require more and more intricate property profiles. However, chemical and materials research has been undergoing a significant transformation in recent years that can alleviate many of these shortcomings: After decades of continuous advances in methods, algorithms, and computer hardware, the fields of modeling and simulation have reached a tipping point, and they are finally at a stage where they can make accurate predictions for systems that are both realistic and relevant. Progress is now increasingly driven by computational studies, which have become crucial assets in the

* m27@buffalo.edu
† hachmann@buffalo.edu

pursuit of next-generation materials and chemistry. By making guiding predictions, they can significantly boost the efficiency of research endeavors, and uncover promising targets for investigations in the laboratory (see, e.g., Ref. [27–34]). The White House Materials Genome Initiative [35] underscores the value of integrated joint ventures between experimentalists and theoreticians in tackling complex discovery and design challenges and delivering revolutionary new materials.

A prerequisite for the computationally-driven development of a new materials is access to suitable (i.e., accurate and efficient) computational protocols for the target property within a compound space of interest. This paper presents such a protocol for the prediction of RI values of organic polymers. One of the distinctive feature of this protocol compared to prior work by others [36, 37] is that it fuses *first-principles* and data modeling.

In Sec. II, we introduce the physical foundations of the proposed protocol (Sec. II A), motivate a number of assumptions and approximations that are used (Sec. II B and II C), and discuss the details of the employed computational approach (Sec. II D). In Sec. III, we present and discuss results for the different components that comprise the protocol (Sec. III A, III B, and III C) as well as the overall protocol itself (Sec. III D). In each case we evaluate the predictive performance of our model by comparing its results with data from a validation set of experimentally known compounds. Sec. III E provides a discussion of the interplay between the physical parameters of our model. Our findings are summarized in Sec. IV.

II. BACKGROUND AND METHODS

A. Lorentz-Lorenz Equation

The RI value (n_r) is defined as the ratio between the speed of light in vacuum (c_0) and in a given material (c). For non-magnetic materials, the RI is thus the square root of its relative permittivity (ϵ_r), i.e.,

$$n_r = \frac{c_0}{c} = \sqrt{\epsilon_r}.$$

The permittivity is a function of the polarizability (α) and using the Lorentz local field approximation, it can be written as

$$\epsilon_r = \frac{1 + 2\alpha N/3\epsilon_0}{1 - \alpha N/3\epsilon_0},$$

where N is the number density, i.e., the number of molecules per volume. It follows that the RI is

$$n_r = \sqrt{\frac{1 + 2\alpha N/3\epsilon_0}{1 - \alpha N/3\epsilon_0}},$$

which is a version of the Lorentz-Lorenz equation (equivalent to the Clausius-Mossotti relation). It follows, that the Lorentz-Lorenz equation connects the macroscopic RI value of a bulk material to the electronic polarizability α and number density N of its molecular constituents. The Lorentz-Lorenz equation thus offers a route to calculating the RI value of a material *via* α and N, and we use it as the physical basis for the proposed computational protocol.

B. *First-Principles* Molecular Polarizability Calculations

The polarizability α of a compound can be obtained from quantum chemical linear response calculations. An array of electronic structure methods has been used to determine the polarizability values of various materials [38–44], including organic polymers [45–47]. The polarizability is generally a frequency-dependent (i.e., dynamic) property as shown in Fig. 1. The frequency-dependent polarizability is relatively hard to compute, as it formally involves solving the time-dependent Schrödinger equation and/or scanning through the range of relevant frequencies [48]. Consequently, only relatively few studies consider the polarizability dispersion in organic polymers [49, 50].

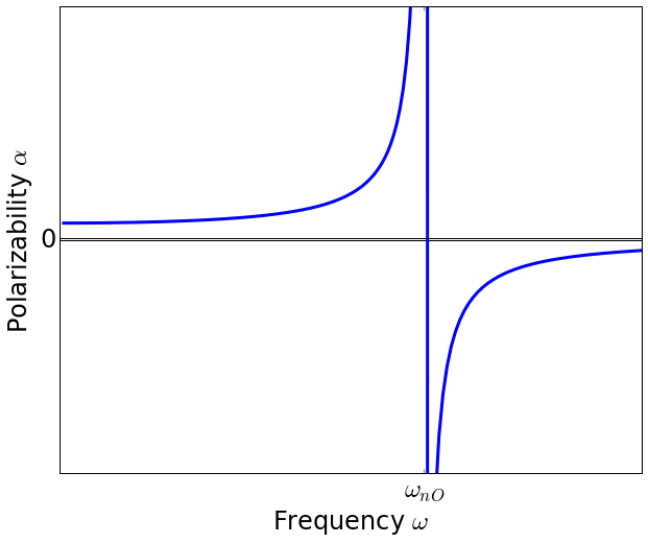

FIG. 1. Dispersion characteristic of the polarizability: Excited states are marked by singularities, and in the frequency range below, the polarizability converges asymptotically towards the constant static polarizability value.

From the frequency dependence of the polarizability follows that the RI is also a frequency-dependent property. However, the variation of both polarizability and RI in the visible frequency region is in fact often relatively small [26], as long as low-lying excited states are absent. Since the latter would render a material unsuitable for optical applications anyways, we generally do not consider materials that exhibit them. Large variations can be observed in the ultraviolet region where resonances

with the excited state manifold become a dominant feature, but since stability considerations prohibit organic polymers to be used for high-energy applications, this is also not a relevant concern. Below the frequency range of the excited states, the polarizability and RI value taper off monotonically (cf. Fig. 1). In most cases, they become essentially constant throughout the visible and infrared range and converge to an asymptotic value [26]. The asymptotic RI corresponds to the value that can be obtained from the static polarizability. The latter can be computed much more easily than the frequency-dependent value. It only requires a single linear response calculation without explicit time dependence, and is thus much less demanding in terms of computing time and numerical stability. We can conclude that the RI values obtained from static polarizability calculations form a close lower bound for the frequency-dependent values in the relevant spectral range. This approach has been used in the past and has given very good agreement with experimental results [37, 45, 46, 51, 52].

Another challenge is to perform the polarizability calculations for quasi-infinite polymers. Realistic systems are amorphous and may thus not be well represented by periodic boundary condition calculations, while non-periodic calculations on long-chain oligomer models are generally cost-prohibitive. However, for systems with a relatively short correlation length (e.g., due to finite conjugation and delocalization of the π-electron backbone), we can expect an early onset of extensivity in the optical properties. We can exploit this behavior through an extrapolation scheme. In this scheme, we perform a series of relatively simple monomer and small oligomer calculations until we observe a linear trend in the polarizability results, based on which we can project to the polymer limit.

The molecular polarizability calculations in the proposed protocol utilize Kohn-Sham density functional theory (DFT) for its advantageous trade-off of cost and accuracy [53]. The former includes its low-order polynomial scaling with system size and its relatively modest basis set demands (compared to high-level wavefunction methods). Given the molecular-level disorder in amorphous polymers, we forgo the expensive *first-principles* optimization of idealized geometries of our candidate compounds in favor of an inexpensive molecular mechanics approach. A simple, yet efficient way to identify (and exclude) compounds with potentially low-lying excited states is to assess the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The HOMO–LUMO gap is a first approximation for the lowest excitation, and it is readily obtained in DFT at no additional cost.

C. Data Model for the Number Density

The number density N for amorphous polymers is typically computed using classical molecular dynamics simulations. However, this approach is relatively cumbersome and computationally demanding. As an alternative, we pursue an approach based on the molecular volume (approximated by the van der Waals volume V_{vdW}), i.e.,

$$N = \frac{K_p}{V_{vdW}},$$

where K_p is the packing fraction in the bulk polymer, which has shown good agreement with experimental results in other work [54, 55]. There are a number of ways to compute the van der Waals volume, ranging from complicated electronic structure calculations with subsequent partitioning of the electron density to simplistic fragment methods [56, 57]. A benchmark study that we will detail elsewhere has shown that the differences in results from different methods are generally small. For the present work, we thus adopt the latter, i.e., we calculate V_{vdW} by adding tabulated atomic values [58] and subtracting off the overlap in the bonding region. The average packing fraction K_p for organic polymers is given in the literature as 0.68 [59], however, the actual values of different polymers show a significant spread and are known to range at least from 0.5 to 0.8. (K_p is generally also a function of the degree of polymerization, but except for shorter oligomers, this only plays a minor role.) As the average value of this critical parameter is thus essentially meaningless, we have devised a machine learning model to correlate the polymer structure with its packing fraction. Due to the relatively small volume of available training data, we chose a comparatively inflexible (but exceedingly fast) support vector regression (SVR) approach to avoid overfitting.

D. Computational Details

The polarizability calculations of the proposed protocol use an all-electron, restricted DFT framework with the PBE0 hybrid functional [60] in combination with the triple-ζ quality def2-TZVP basis set by the Karlsruhe group [61]. We include Grimme's D3 correction [62] to account for dispersion interaction. The proof-of-principle study shown in the following section was carried out using the ORCA 3.0.2 quantum chemistry program package [63] with default settings. We optimized the geometries of all monomers and oligomers using the universal force field (UFF) [64] as implemented in the OpenBabel software [65]. We calculated the van der Waals volumes using Slonimskii's method detailed in Ref. [66], for which we implemented a Python script. We generated the packing fraction model using SVR within a feature space of 43 constitutional descriptors on a training data set of 84 polymers with experimentally known K_p values compiled from the literature. The available data was divided into 80% training and 20% test set for cross-validation. The data modeling was performed using *ChemML* 0.9 [67], our program suite for machine learning and informatics in chemical and materials research. In this work,

ChemML employed the scikit-learn 0.17 SVR library [68] and descriptors from Dragon 7 [69]. Further details of the data model can be found in the Supplementary Material. The proof-of-principle study involved about 450 individual calculations, which we performed using *ChemHTPS* 0.6 [70], our program suite for automated virtual high-throughput screening in chemical and materials research.

III. RESULTS AND DISCUSSION

We developed the proposed RI protocol on two common non-conjugated polymers – polyethylene (PE) and polystyrene (PS) – as prototype systems. Subsequently, we performed a study of 112 non-conjugated polymers for which experimental RI values are known in order to validate the predictive performance of the RI protocol as well as its individual components.

A. Polarizabilities

The PBE0/def2-TZVP-D3 polarizability results for PE and PS from monomer to heptamer are shown in Fig. 2. The linear trend with respect to the number of monomer units n (due to extensivity) is easily recognized. The correlation coefficient R^2 for the linear regression is $\gg 0.99$. For all cases studied in this work, extensivity was observed for very short oligomer sequences, and we based our extrapolation scheme on the linear regression slope obtained from the monomer to tetramer results.

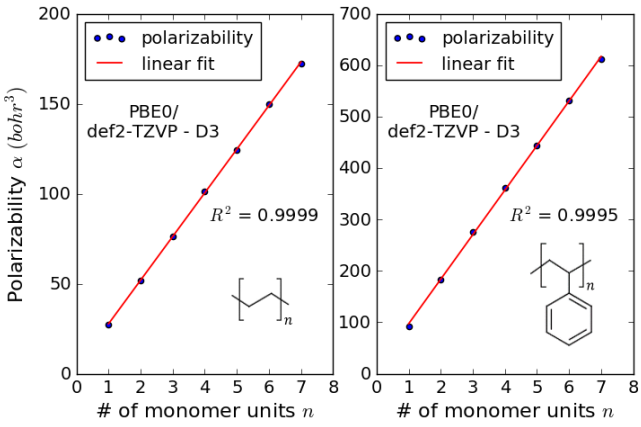

FIG. 2. Linear relationship between number of monomer units and polarizability for polyethylene (PE) and polystyrene (PS) as prototypes for non-conjugated polymers.

Note that for conjugated polymers with longer correlation lengths, the onset of extensivity can occur at significantly longer chain length, i.e., values for a sequence of shorter oligomers will not show a linear trend [71, 72]. The scheme can still be used in these cases, but it requires the calculation of longer oligomer sequences until a linear trend for extrapolation is found.

B. Number Densities and Densities

Using Slonimskii's method we could readily compute the van der Waals volumes V_{vdW} for the prototype systems PE and PS. Assuming the average packing fraction of $K_p = 0.68$, we obtain the number density values N as a function of the number of monomer units n shown in Fig. 3. The plots illustrate that N decreases monotonically with increasing number of monomer units, and the inverse $1/N \propto V_{vdW}$ is evidently extensive.

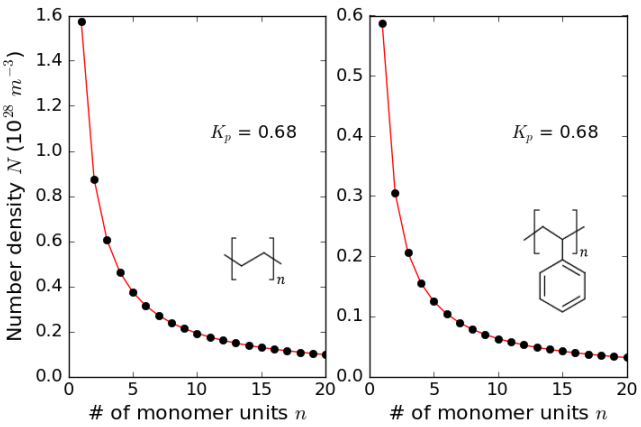

FIG. 3. Change of number density values for increasing degree of polymerization for the PE and PS prototype systems.

Note that we can use this approach to compute another property of interest in organic materials research, i.e., the density ρ of amorphous polymers in the bulk (*via* $\rho = N \cdot N_A/M$ with the Avogadro number N_A and molecular weight M). The density results for our prototype systems are presented in Fig. 4. The plots show that ρ increases and ultimately convergences towards asymptotically constant values. This finite size effect due to the terminal groups is typically of limited magnitude. The results of the oligomers with $n = 50$ offer a good representation of the polymer limit and can thus be used as the default for determining ρ. The resulting densities are in very good agreement with experimentally known values [73]. We can also use ρ to work backwards and obtain the actual K_p values. For PE and PS we obtain 0.64 and 0.66, respectively, i.e., using the average packing fraction of 0.68 happened to be a valid assumption in these particular cases.

C. Packing Fractions

As K_p is generally not known and the average value referenced in the literature [59] is of limited utility, we have devised an SVR data model that correlates the polymer structure with the packing fraction as outlined in Sec. II C and II D. Fig. 5 displays the range and distribution of K_p values for the 84 polymers for which we found experimental results. This data – ranging from 0.53 to

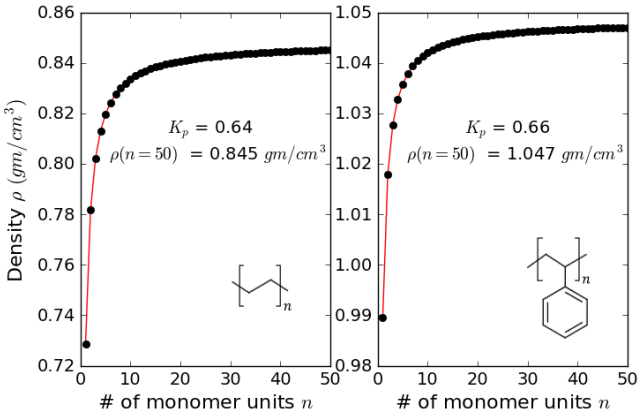

FIG. 4. Change of density for increasing degree of polymerization for the PE and PS prototype systems. Note the characteristic asymptotic convergence to a constant value.

0.79 with an average value of 0.67 – formed the basis for our data-derived K_p prediction model. (Note that the average K_p for our data set is nearly identical with the average $K_p = 0.68$ cited in Ref. [59]).

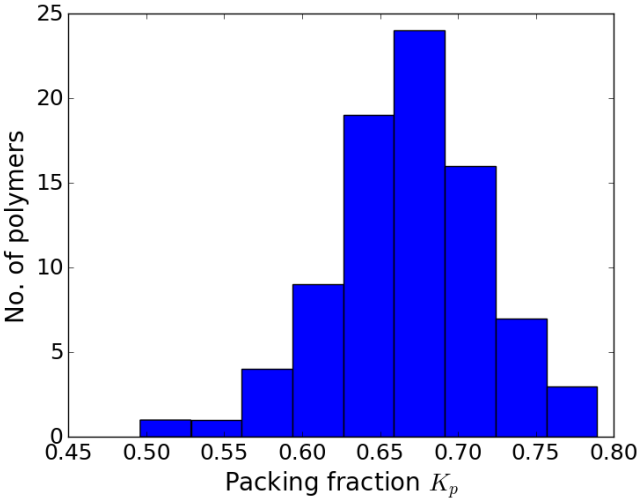

FIG. 5. Range and distribution of experimental packing fraction values for 84 polymers used in the creation of our data-derived prediction model.

The model gives an R^2 of 0.97 for the training and 0.87 for the test set. The performance drop for the latter is reasonable and acceptable given the small size of the available data set. The computational demand for the K_p prediction model is negligible and results for even large-scale compound libraries can be obtained in minutes on a single processor.

D. Refractive Indices

Given the modeling protocols and resulting data for α, V_{vdW}, K_p, and N, we use the Lorentz-Lorenz equation to make RI predictions. Using the α and N values obtained for our PE and PS prototype systems as shown in Figs. 2 and 3, we calculate the RI values and their variation with the number of monomer units n given in Fig. 6. The RI increases for longer oligomers before reaching a plateau for $n = 20$ to 30.

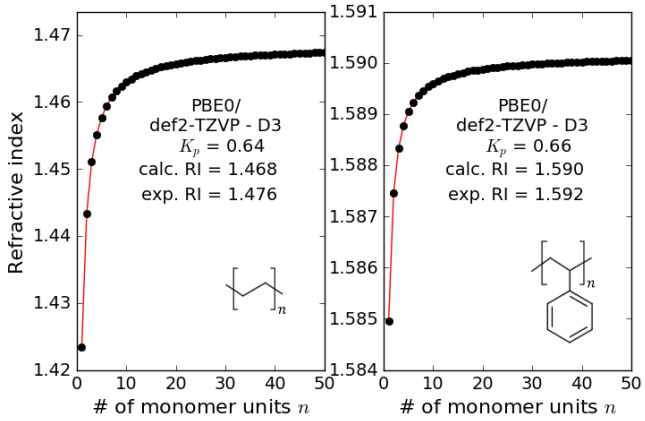

FIG. 6. Change of refractive index (RI) with increasing degree of polymerization for the PE and PS prototype systems. Note the characteristic asymptotic convergence to the constant value at the polymer limit, that is in excellent agreement with the experimental data.

Our modeling protocol predicts the RI values for PE and PS to be 1.468 and 1.590, respectively, which is in outstanding agreement with the experimental RI values of 1.476 and 1.592, respectively [73]. We further validate our modeling protocol by predicting the RI values of 112 polymers for which we could find experimental data for comparison (see Fig. 7).

The R^2 of 0.94 shows that the model is in very good agreement with the experimental RI values. The benchmark comparison gives a mean absolute deviation (MAD) of 0.010 (0.9%), a root mean square deviation (RMSD) of 0.018 (0.1%), and a maximum deviation (MaxD) of 0.045 or 3.0%, respectively, i.e., our modeling protocol is quite accurate and affords at least semi-quantitative predictions (in particular since typically only two decimals in the RI values are considered as significant). The average deviation (AD) is very small with +0.004 (+0.3%), i.e., our model is not significantly biased towards systematic over- or under-predictions. A result of particular importance for studies that focus on candidate rankings rather than quantitative predictions for individual candidates is that the trends in the data are generally well captured. We stress that the experimental RI data is independent of the data used for the creation of the SVR model for N, i.e., the SVR model was not biased towards providing good RI results.

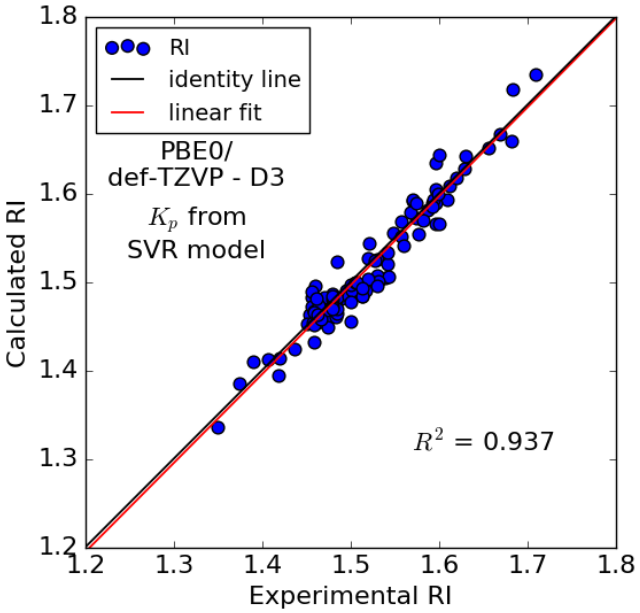

FIG. 7. Validation of the proposed RI prediction model (based on the data-derived model for the packing fraction) through comparison with 112 experimental data points.

For comparison, using the average packing fraction value of 0.68 – as is oftentimes cited in related work – instead of our SVR model leads to the results shown in Fig. 8. This model is considerably worse, as can be seen from the R^2 of 0.78, MAD of 0.019 (1.9%), RMSD of 0.026 (0.3%), MaxD of 0.139 and 6.9%, and AD of −0.009 (−0.6%).

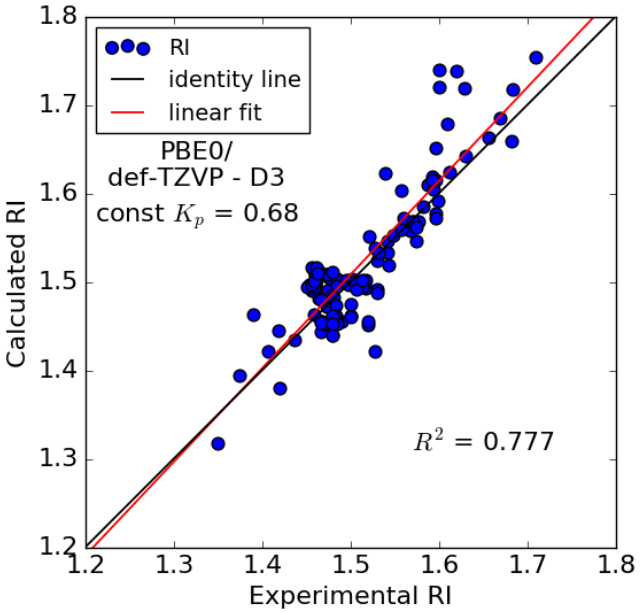

FIG. 8. Validation of an RI prediction model based on a constant average packing fraction.

E. Interplay between Polarizability and Number Density

As the Lorentz-Lorenz equation relies on α and N as input parameters, we analyze their interplay for the 112 polymers in our validation and benchmark data set. Fig. 9 shows the calculated α and N values as well as the contour lines for the resulting RI values in this parameter space.

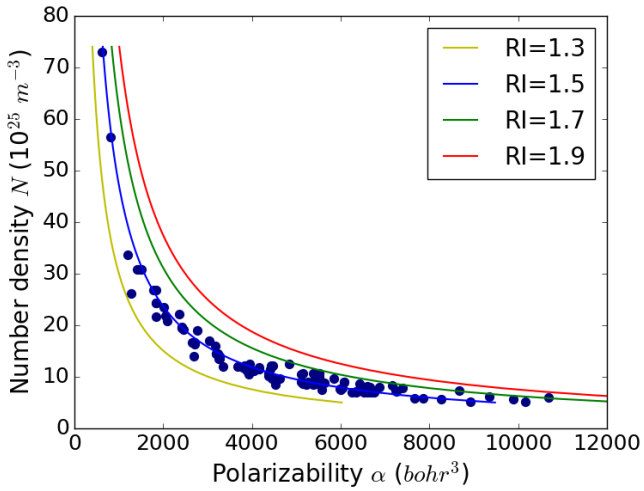

FIG. 9. Parameter space of polarizability α and number density N as well as resulting RI value domains with examples from our validation and benchmark data.

To achieve high RI values, a candidate compound must feature both large number density and polarizability values (as is also apparent from the structure of the Lorentz-Lorenz equation). Optimizing both properties simultaneously is a challenging task as extensivity couples α and N, i.e., the longer a polymer, the larger α (as α increases with the number of contributing monomer units), but the smaller N (as fewer molecules fit into a given volume element). A design strategy that can be derived from this notion is to incorporate highly polarizable moieties that have a limited effect on the number density. The data set at hand is tilted towards a larger spread in α while N is more clustered. It is worth noting that the RI regions in the α *vs* N parameter space are relatively narrow and most compounds in the data set group around the contour line for RI=1.5. The high-RI examples primarily stand out for large polarizability values rather than large number densities, which supports the before-mentioned design strategy.

IV. CONCLUSIONS

We have successfully developed an *in silico* modeling protocol for the accurate and efficient prediction of the RI values of organic polymeric materials and have demonstrated its performance by faithfully reproducing the ex-

perimentally known RI values of 112 compounds. Our work is an example for the synergistic benefits of fusing physical and data models, with the former providing the general structure of the approach (i.e., the Lorentz-Lorenz equation) and a significant part of the required input parameters (i.e., polarizabilities from *first-principles* quantum chemistry and van der Waals volumes from Slonimskii's method), while the latter provides rapid access to input data that is otherwise not readily available (i.e., an SVR model for the packing fraction and an extrapolation scheme for molecular results towards the polymer limit). A subset of our RI protocol can also be used to predict the density of amorphous polymers. Our work is furthermore an example for the great promise of applying machine learning and modern data science in chemical research. In subsequent work, we will utilize this new RI protocol to conduct virtual high-throughput screening studies on large-scale candidate libraries with the goal of accelerating the discovery of novel organic materials with unprecedented RI values.

ACKNOWLEDGMENTS

This work was supported by start-up funds provided through the University at Buffalo (UB). Computing time on the high-performance computing clusters '*Rush*', '*Alpha*', and '*Beta*' was provided by the UB Center for Computational Research (CCR). The work presented here is part of MAFA's PhD thesis. Electronic Supplementary Material accompanies this paper and is available through the journal website.

[1] T. Higashihara and M. Ueda, "Recent progress in high refractive index polymers," Macromolecules **48**, 1915–1929 (2015).

[2] C. L. Lu, C. Guan, Y. F. Liu, Y. R. Cheng, and B. Yang, "PbS/polymer nanocomposite optical materials with high refractive index," Chemistry of Materials **17**, 2448–2454 (2005).

[3] L. Zimmermann, M. Weibel, W. Caseri, and U. W. Suter, "High refractive-index films of polymer nanocomposites," Journal of Materials Research **8**, 1742–1748 (1993).

[4] J.-g. Liu and M. Ueda, "High refractive index polymers: fundamental research and practical applications," Journal of Materials Chemistry **19**, 8907–8919 (2009).

[5] T. Lei, J. Y. Wang, and J. Pei, "Roles of flexible chains in organic semiconducting materials," Chemistry of Materials **26**, 594–603 (2014).

[6] M. D. Angione, R. Pilolli, S. Cotrone, M. Magliulo, A. Mallardi, G. Palazzo, L. Sabbatini, D. Fine, A. Dodabalapur, N. Cioffi, and L. Torsi, "Carbon based materials for electronic bio-sensing," Materials Today **14**, 424–433 (2011).

[7] A. Voigt, U. Ostrzinski, K. Pfeiffer, J. Y. Kim, V. Fakhfouri, J. Brugger, and G. Gruetzner, "New inks for the direct drop-on-demand fabrication of polymer lenses," Microelectronic Engineering **88**, 2174–2179 (2011).

[8] S. Ummartyotin, J. Juntaro, M. Sain, and H. Manuspiya, "Development of transparent bacterial cellulose nanocomposite film as substrate for flexible organic light emitting diode (OLED) display," Industrial Crops and Products **35**, 92–97 (2012).

[9] C. Xiang and R. Ma, "Devices to increase OLED output coupling efficiency with a high refractive index substrate," (2017), US Patent 9,640,781.

[10] H. Nishiyama, J. Nishii, M. Mizoshiri, and Y. Hirata, "Microlens arrays of high-refractive-index glass fabricated by femtosecond laser lithography," Applied Surface Science **255**, 9750–9753 (2009).

[11] Y. Kokubun, N. Funato, and M. Takizawa, "Athermal waveguides for temperature-independent lightwave devices," IEEE Photonics Technology Letters **5**, 1297–1300 (1993).

[12] H. Wei and S. Krishnaswamy, "Direct laser writing polymer micro-resonators for refractive index sensors," IEEE Photonics Technology Letters **28**, 2819–2822 (2016).

[13] A. Rodrguez, G. Vitrant, P. A. Chollet, and F. Kajzar, "Optical control of an integrated interferometer using a photochromic polymer," Applied Physics Letters **79**, 461–463 (2001).

[14] S. Singaravalu, D. Mayo, H. Park, K. Schriver, and R. Haglund, "Anti-reflective polymer-nanocomposite coatings fabricated by RIR-MAPLE," in *SPIE LASE*, Vol. 8607 (International Society for Optics and Photonics, 2013) p. 860718.

[15] J.-B. Kim, J.-H. Lee, C.-K. Moon, S.-Y. Kim, and J.-J. Kim, "Highly enhanced light extraction from surface plasmonic loss minimized organic light-emitting diodes," Advanced Materials **25**, 3571–3577 (2013).

[16] E. Kim, H. Cho, K. Kim, T.-W. Koh, J. Chung, J. Lee, Y. Park, and S. Yoo, "A facile route to efficient, low-cost flexible organic light-emitting diodes: Utilizing the high refractive index and built-in scattering properties of industrial-grade PEN substrates," Advanced Materials **27**, 1624–1631 (2015).

[17] S.-S. Sun and L. R. Dalton, *Introduction to Organic Electronic and Optoelectronic Materials and Devices (Optical Science and Engineering Series)* (Boca Raton, FL, USA, 2008).

[18] H. Jintoku and H. Ihara, "The simplest method for fabrication of high refractive index polymer-metal oxide hybrids based on a soap-free process," Chemical Communications **50**, 10611–10614 (2014).

[19] R. Seto, T. Sato, T. Kojima, K. Hosokawa, Y. Koyama, G. I. Konishi, and T. Takata, "9,9 '-spirobifluorene-containing polycarbonates: Transparent polymers with high refractive index and low birefringence," Journal of Polymer Science Part A-Polymer Chemistry **48**, 3658–3667 (2010).

[20] J. J. Griebel, S. Namnabat, E. T. Kim, R. Himmelhuber, D. H. Moronta, W. J. Chung, A. G. Simmonds, K.-J. Kim, J. van der Laan, N. A. Nguyen, E. L. Dereniak, M. E. Mackay, K. Char, R. S. Glass, R. A. Norwood, and J. Pyun, "New infrared transmitting material via inverse

vulcanization of elemental sulfur to prepare high refractive index polymers," Advanced Materials **26**, 3014–3018 (2014).

[21] Y. Tojo, Y. Arakwa, J. Watanabe, and G. Konishi, "Synthesis of high refractive index and low-birefringence acrylate polymers with a tetraphenylethane skeleton in the side chain," Polymer Chemistry **4**, 3807–3812 (2013).

[22] W. F. Ho, M. A. Uddin, and H. P. Chan, "The stability of high refractive index polymer materials for high-density planar optical circuits," Polymer Degradation and Stability **94**, 158–161 (2009).

[23] B. T. Liu and P. S. Li, "Preparation and characterization of high-refractive-index polymer/inorganic hybrid films containing TiO2 nanoparticles prepared by 4-aminobenzoic acid," Surface & Coatings Technology **231**, 301–306 (2013).

[24] L. Parke, I. R. Hooper, R. J. Hicken, C. E. J. Dancer, P. S. Grant, I. J. Youngs, J. R. Sambles, and A. P. Hibbins, "Heavily loaded ferrite-polymer composites to produce high refractive index materials at centimetre wavelengths," APL Materials **1** (2013).

[25] Q. Y. Zhang, K. Su, M. B. Chan-Park, H. Wu, D. A. Wang, and R. Xu, "Development of high refractive ZnS/PVP/PDMAA hydrogel nanocomposites for artificial cornea implants," Acta Biomaterialia **10**, 1167–1176 (2014).

[26] D. R. Robello, "Moderately high refractive index, low optical dispersion polymers with pendant diamondoids," Journal of Applied Polymer Science **127**, 96–103 (2013).

[27] R. S. Sánchez-Carrera, S. Atahan, J. Schrier, and A. Aspuru-Guzik, "Theoretical characterization of the air-stable, high-mobility dinaphtho[2,3-b:23-f]thieno[3,2-b]-thiophene organic semiconductor," The Journal of Physical Chemistry C **114**, 2334–2340 (2010).

[28] A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. B. Mannsfeld, A. P. Zoombelt, Z. Bao, and A. Aspuru-Guzik, "From computational discovery to experimental characterization of a high hole mobility organic crystal." Nature Communication **2**, 437 (2011).

[29] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, "The Harvard Clean Energy Project: Large-scale computational screening and design of organic photovoltaics on the world community grid," The Journal of Physical Chemistry Letters **2**, 2241–2251 (2011).

[30] R. Olivares-Amaya, C. Amador-Bedolla, J. Hachmann, S. Atahan-Evrenk, R. S. Sánchez-Carrera, L. Vogt, and A. Aspuru-Guzik, "Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics," Energy & Environmental Science **4**, 4849–4861 (2011).

[31] C. Amador-Bedolla, R. Olivares-Amaya, J. Hachmann, and A. Aspuru-Guzik, "Organic Photovoltaics," in *Informatics Mater. Sci. Eng. Data-driven Discov. Accel. Exp. Appl.*, edited by K. Rajan (Oxford, 2013) Chap. 17, pp. 423–442.

[32] J. Hachmann, R. Olivares-Amaya, A. Jinich, A. L. Appleton, M. A. Blood-Forsythe, L. R. Seress, C. Roman-Salgado, K. Trepte, S. Atahan-Evrenk, S. Er, S. Shrestha, R. Mondal, A. Sokolov, Z. Bao, and A. Aspuru-Guzik, "Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry - the Harvard Clean Energy Project," Energy & Environmental Science **7**, 698–704 (2014).

[33] E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, and A. Aspuru-Guzik, "What is high-throughput virtual screening? A perspective from organic materials discovery," Annual Reviews of Materials Research **45**, 195–216 (2015).

[34] S. A. Lopez, E. O. Pyzer-Knapp, G. N. Simm, T. Lutzow, K. Li, L. R. Seress, J. Hachmann, and A. Aspuru-Guzik, "The Harvard organic photovoltaic dataset," Sci. Data **3**, 160086 (2016).

[35] National Science and Technology Council, "Materials Genome Initiative for Global Competitiveness," Tech. Rep. (Washington, DC: National Science and Technology Council, 2011).

[36] H. Redmond and J. E. Thompson, "Evaluation of a quantitative structure-property relationship (QSPR) for predicting mid-visible refractive index of secondary organic aerosol (SOA)," Physical Chemistry Chemical Physics **13**, 6872–6882 (2011).

[37] S. S. Park, S. Lee, J. Y. Bae, and F. Hagelberg, "Refractive indices of liquid-forming organic compounds by density functional theory," Chemical Physics Letters **511**, 466–470 (2011).

[38] S. Ando, "DFT calculations on refractive index dispersion of fluoro-compounds in the DUV-UV-visible region," Journal of Photopolymer Science and Technology **19**, 351–360 (2006).

[39] X. Rocquefelte, F. Goubin, H.-J. Koo, M.-H. Whangbo, and S. Jobic, "Investigation of the origin of the empirical relationship between refractive index and density on the basis of first principles calculations for the refractive indices of various TiO2 phases," Inorganic chemistry **43**, 2246–2251 (2004).

[40] B. Jensen and A. Torabi, "Quantum theory of the dispersion of the refractive index near the fundamental absorption edge in compound semiconductors," IEEE Journal of Quantum Electronics **19**, 448–457 (1983).

[41] M. Rabah, B. Abbar, Y. Al-Douri, B. Bouhafs, and B. Sahraoui, "Calculation of structural, optical and electronic properties of ZnS, ZnSe, MgS, MgSe and their quaternary alloy $Mg_{1-x}Zn_xS_ySe_{1-y}$," Materials Science and Engineering: B **100**, 163–171 (2003).

[42] B. Amrani, T. Benmessabih, M. Tahiri, I. Chiboub, S. Hiadsi, and F. Hamdache, "First principles study of structural, elastic, electronic and optical properties of CuCl, CuBr and CuI compounds under hydrostatic pressure," Physica B: Condensed Matter **381**, 179–186 (2006).

[43] A. H. Reshak and W. Khan, "Electronic structure, optical and thermoelectric transport properties of layered polyanionic hydrosulfate LiFeSO4OH: Electrode for Li-ion batteries," Journal of Alloys and Compounds **591**, 362–369 (2014).

[44] S. Azam and A. H. Reshak, "Electronic structure of 1,3-dicarbomethoxy4,6-benzenedicarboxylic acid: Density functional approach," International Journal of Electrochemical Science **8**, 10359–10375 (2013).

[45] V. Ksianzou, R. K. Velagapudi, B. Grimm, and S. Schrader, "Polarization-dependent optical characterization of poly(phenylquinoxaline) thin films," Journal of Applied Physics **100** (2006).

[46] C. D. Zeinalipour-Yazdi and D. P. Pullman, "Quantitative structure - property relationships for longitudinal, transverse, and molecular static polarizabilities in

polyynes," Journal of Physical Chemistry B **112**, 7377–7386 (2008).

[47] X. L. Yu, B. Yi, and X. Y. Wang, "Prediction of refractive index of vinyl polymers by using density functional theory," Journal of Computational Chemistry **28**, 2336–2341 (2007).

[48] Z. F. Rao and R. F. Zhou, "Electronic structure and optical properties of resin," Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy **105**, 618–622 (2013).

[49] C. K. Rowan and I. Paci, "Optical properties of Ag/polyvinylidene fluoride nanocomposites: A theoretical study," Journal of Physical Chemistry C **115**, 8316–8324 (2011).

[50] A. Lenz, H. Kariis, A. Pohl, P. Persson, and L. Ojamae, "The electronic structure and reflectivity of PEDOT:PSS from density functional theory," Chemical Physics **384**, 44–51 (2011).

[51] M. E. Azim-Araghi, J. Baedi, and L. M. Goodarzi, "Electrical and optical properties of an organic semiconductor metal-free phthalocyanine (C32H18N8)," European Physical Journal-Applied Physics **58**, 30201 (2012).

[52] S. Lee and S. S. Park, "Dielectric properties of organic solvents from non-polarizable molecular dynamics simulation with electronic continuum model and density functional theory," Journal of Physical Chemistry B **115**, 12571–12576 (2011).

[53] F. Neese, "Prediction of molecular properties and molecular spectroscopy with density functional theory: From fundamental theory to exchange-coupling," Coord. Chem. Rev. **253**, 526–563 (2009).

[54] J. G. Liu, Y. Nakamura, T. Ogura, Y. Shibasaki, S. Ando, and M. Ueda, "Optically transparent sulfur-containing polyimide-TiO(2) nanocomposite films with high refractive index and negative pattern formation from poly(amic acid)-TiO(2) nanocomposite film," Chemistry of Materials **20**, 273–281 (2008).

[55] Y. Terui and S. Ando, "Coefficients of molecular packing and intrinsic birefringence of aromatic polyimides estimated using refractive indices and molecular polarizabilities," Journal of Polymer Science Part B-Polymer Physics **42**, 2354–2366 (2004).

[56] T. Lu and F. Chen, "Multiwfn: A multifunctional wavefunction analyzer," Journal of Computational Chemistry **33**, 580–592 (2012).

[57] Y. H. Zhao, M. H. Abraham, and A. M. Zissimos, "Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds," The Journal of Organic Chemistry **68**, 7368–7373 (2003).

[58] A. Bondi, "van der Waals volumes and radii," The Journal of Physical Chemistry **68**, 441–451 (1964).

[59] A. Askadskiĭ, *Computational Materials Science of Polymers* (Cambridge Int Science Publishing, 2003).

[60] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," The Journal of Chemical Physics **110**, 6158–6170 (1999).

[61] F. Weigend, R. Ahlrichs, K. A. Peterson, T. H. Dunning, R. M. Pitzer, and A. Bergner, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Physical Chemistry Chemical Physics **7**, 3297 (2005).

[62] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," The Journal of Chemical Physics **132**, 154104 (2010).

[63] F. Neese, "The ORCA program system," Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 73–78 (2012).

[64] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," Journal of the American Chemical Society **114**, 10024–10035 (1992).

[65] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," Journal of Cheminformatics **3**, 33–33 (2011).

[66] G. Slonimskii, A. Askadskii, and A. Kitaigorodskii, "The packing of polymer molecules," Polymer Science USSR **12**, 556–577 (1970).

[67] J. Hachmann and M. Haghighatlari, "*ChemML* 0.9 (2017),".

[68] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research **12**, 2825–2830 (2011).

[69] Talete srl, "DRAGON (Software for Molecular Descriptor Calculation)," (2011).

[70] J. Hachmann, W. S. Evangelista, and M. A. F. Afzal, "*ChemHTPS* 0.6 (2017),".

[71] H. Villar, M. Dupuis, J. Watts, G. Hurst, and E. Clementi, "Structure, vibrational spectra, and IR intensities of polyenes from ab initio SCF calculations," Journal of Chemical Physics **88**, 1003–1009 (1988).

[72] D. H. Mosley, J. G. Fripiat, B. Champagne, and J.-M. André, "Ab initio investigation of the static polarizability of planar and twisted infinite polythiophene chains," International Journal of Quantum Chemistry **52**, 451–467 (1994).

[73] J. Bicerano, *Prediction of Polymer Properties* (CRC Press, 2002).