Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning

Jack D Evans, François-Xavier Coudert

HAL Id: hal-02264216
## https://hal.science/hal-02264216

Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning

Jack D. Evans*[ID] and François-Xavier Coudert[ID]

Chimie ParisTech, PSL Research University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France

[S] *Supporting Information*

ABSTRACT: We show here that machine learning is a powerful new tool for predicting the elastic response of zeolites. We built our machine learning approach relying on geometric features only, which are related to local geometry, structure, and porosity of a zeolite, to predict bulk and shear moduli of zeolites with an accuracy exceeding that of force field approaches. The development of this model has illustrated clear correlations between characteristic features of a zeolite and elastic moduli, providing exceptional insight into the mechanics of zeolitic frameworks. Finally, we employ this methodology to predict the elastic response of 590 448 hypothetical zeolites, and the results of this massive database provide clear evidence of stability trends in porous materials.

Gradient Boosting Regressor
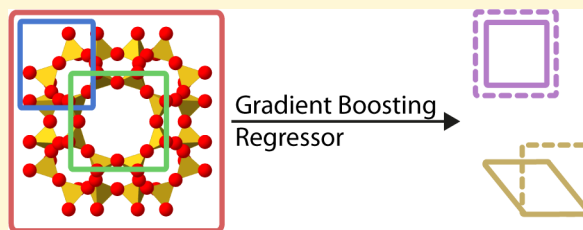
INTRODUCTION

Metastable phases of SiO_2, also known as siliceous zeolites, are typically formed by crystallization in the presence of templating agents.[1] This produces nanoporous structures that have applications in adsorption, separation, and gas storage.[2] More than 200 unique zeolite topologies have been observed from simple tetrahedral SiO_4 building blocks connected by their edges, though many fewer have been obtained as pure silica compounds.

This structural diversity of zeolitic frameworks induces a wide variety of their properties, including geometric characteristics (e.g., channel dimensionality, connectivity, and pore space), which are directly relevant for adsorption-related applications, but also in physical properties such as mechanical behavior. Indeed, among the large array of topologies observed in known zeolites, there are unique structures that display mechanical properties unusual in the inorganic world such as large-scale elastic anisotropy. This also includes a structure that displays complete auxeticity,[3] a rare property demonstrated by few materials, and even fewer crystalline structures.[4] Moreover, it has been proposed that mechanical stability, reflected in the elastic properties of the frameworks, plays a key role in the selection of feasible zeolite structures, i.e. experimentally synthesized structures are among the most mechanically stable potential frameworks.[5]

Commonly, the computational determination of mechanical properties of crystalline materials in the elastic regime involves the optimization and energy determination of many deformation modes, requiring greater computational cost compared to routine optimizations.[6] This can be performed using first-principles methods for the determination of energies at the quantum chemical level, and efficient computing techniques have allowed for the calculation of extensive databases of material properties. Recently, de Jong et al. reported the elastic

properties for 1181 inorganic solids[7] (most of them nonporous structures), and this database is being continually updated.[8] However, even with current advanced computing resources, high-throughput methods can be applied only to a small fraction of realized crystal structures as even for simple quaternary inorganic compounds there is an estimated 10^{10} combinations.[9] As an alternative, computationally cheap classical interatomic pair potentials can also be employed;[10] however, the resulting accuracy is expected to be significantly less.

Bypassing the need for energy-based calculations entirely, machine learning offers a pathway to accelerate materials discovery and optimization[11] by allowing large-scale high-throughput screening of hypothetical materials.[12] In particular, machine learning techniques have been used in many fields[13] including the pharmaceutical and material science industries.[14,15] It is especially powerful in predicting properties for which molecular simulations are complex, such as melting points, solubility and crystallization.[16−18] Recently, machine learning methods have also demonstrated their effectiveness for identifying porosity in organic molecular crystals, hydrogen storage performance, and xenon/krypton separation in the Materials Genome database.[19−21]

In this work, we use machine learning methods to predict the elastic properties of pure silica zeolites. Using an accurate training set of elastic properties calculated by density functional theory (DFT), the machine learning approach is able to predict bulk and shear moduli of zeolites with impressive accuracy. By comparison to five state-of-the-art widely used interatomic potentials, we show that machine learning based purely on the

structures gives better results than all classical potentials. Moreover, intrinsic feature selection performed by the machine learning algorithm has highlighted the influence of local, structural, and porosity features on the resulting elastic moduli. Finally, to demonstrate the efficiency of this model, we predicted the elastic response of 590 448 hypothetical zeolitic structures with a computational cost of mere hundreds of CPU hours, which are primarily required for descriptor calculation. The properties of this massive database provide important evidence to the physical and chemical intuition researchers use to understand the mechanical stability of zeolites and related nanoporous materials.

■ METHODS AND COMPUTATIONAL DETAILS

The application of machine learning methods requires a selection of descriptors which accurately delineate the diverse zeolite structures and their characteristic features and can be expected to relate to the mechanical properties. We chose here to base the machine learning purely on structural and geometric quantities, as our aim is the automated prediction of mechanical properties directly from a single structure without the need for any calculation of chemical properties.

There have been a number of previous studies based on machine learning to predict various properties of crystalline materials, and they use a broad range of different descriptors, ranging from simple geometric and lattice information to complex partial radial distribution functions of the internal coordinates.[22] In this study, we constructed a list of descriptors that characterize the local and structural features of the zeolite structure (Figure 1). All can be directly obtained from the
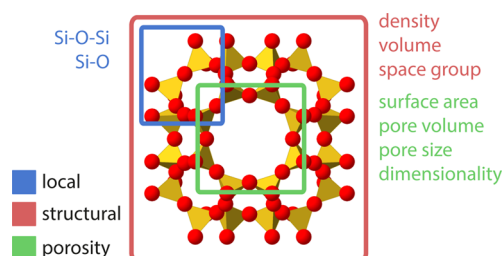

Figure 1. Summary of the descriptors used in the machine learning model, classified in local, structural, and porosity-related descriptors.

structure file in CIF format:[23] crystal density, unit cell volume, and space group. Geometric and topological descriptors have previously shown good performance for predicting the framework type of a zeolite.[24] In addition, inspired by a recent study,[25] we also used several statistical features of the distributions of Si−O bond lengths and Si−O−Si angles. This information is obtained from the CIF file using the `pymatgen` library.[26] We also used as descriptors a series of geometric parameters related to the nature of pore space such as internal surface area, pore volume, size, and channel dimensionality. These are also calculated by geometric methods from the structure files, at low computational cost using the `Zeo++` software.[27,28] A full list of all 32 descriptors can be found in the Supporting Information. Notably, this descriptor list does not include energy, and all descriptors can be calculated by efficient and cheap methods with approximately 1.5 CPU seconds required for each structure.

The training set employed for development of the machine learning model was previously reported by Coudert.[5] This set contains 121 pure silica zeolites which are a subset of the Database of Zeolite Structures. The elastic data was obtained from optimized structures using the B3LYP hybrid exchange−correlation functional[29] with empirical correction for dispersive interactions[30] and all electron basis sets for silicon and oxygen.[31] The reported geometry optimized structures were used to calculate the geometric parameters described previously. Furthermore, isotropic values of bulk and shear moduli

were obtained using Voigt−Ruess−Hill averages[32] computed by ELATE.[33]

To construct our predictive model, we used a gradient boosting regressor (GBR)[34,35] which is part of the Python `scikit-learn` package.[36] This method trains regression trees as an additive model in stepwise approach by optimizing arbitrary loss functions. At each stage, a regression tree is fit on the negative gradient of the loss function. GBR is both an accurate and effective method that has been used in many areas, including web-search ranking.[37] In particular, this method was chosen over other methods such as support vector machines,[38] as GBR models are considered robust, interpretable, and applicable for the small data set present in this study.[39]

Our implementation of GBR uses well-established selection criteria with 3-fold cross-validation, which was repeated 100 times to give a representation of the model accuracy. This approach gives the variance in accuracy. Hyperparameters, given in Table 1, were chosen to

Table 1. Hyperparameters Used for Training the GBR Model

parameter	value
estimators	1000
learning rate	0.01
minimum samples split	2
minimum samples per leaf	3
max depth	3
max features	square root of total features
subsample	0.4

provide high prediction accuracy and flexibility to minimize overfitting. Particularly, this was achieved by using a small learning rate and a shallow parameter for maximum depth of trees. Training deviance plots constructed during the learning procedure, as presented in the Supporting Information, demonstrate the accuracy and generality of the model.

The GBR model was trained to predict the Voigt−Reuss−Hill averages[32] of bulk and shear moduli (K and G, respectively), which describe the compressibility and response to shear of a material. Similar to a study by de Jong et al.,[25] we train and predict for $\log(K)$ and $\log(G)$, which prevents the model overweighting materials with stiffer moduli and allows for simple comparison between the studies. Although learning for K and G, not $\log(K)$ and $\log(G)$, does not appreciably affect the accuracy of the model, as demonstrated in the Supporting Information. Importantly, this model is extremely computationally efficient compared to conventional approaches. Training and prediction from the GBR model required ≈30 CPU seconds to predict the bulk and shear moduli of almost 600 000 hypothetical zeolitic structures.

■ RESULTS AND DISCUSSION

Model Accuracy. We find the machine learning model gives excellent accuracy when tested against the training set by cross-validation. The relatively high accuracy and low variance are demonstrated in Figure 2, where the DFT training set is plotted against the results of our model obtained by cross-validation for 121 pure silica zeolites. Root-mean-squared error (RMSE) was used to evaluate the performance of the model as it is an encouraged measure of dispersion.[40] Notably, RMSE of 0.102 ± 0.034 and 0.0847 ± 0.022 for $\log(K)$ and $\log(G)$, respectively, compare well to the predictive accuracy achieved for simple k-nary inorganic solids; 0.0750 and 0.1378.[25] This is excellent accuracy considering that the model is constructed from a small training set.

Although the level of absolute uncertainty in Figure 2 may seem high in some cases, it is useful to remind ourselves that the determination of elastic constants in porous materials by
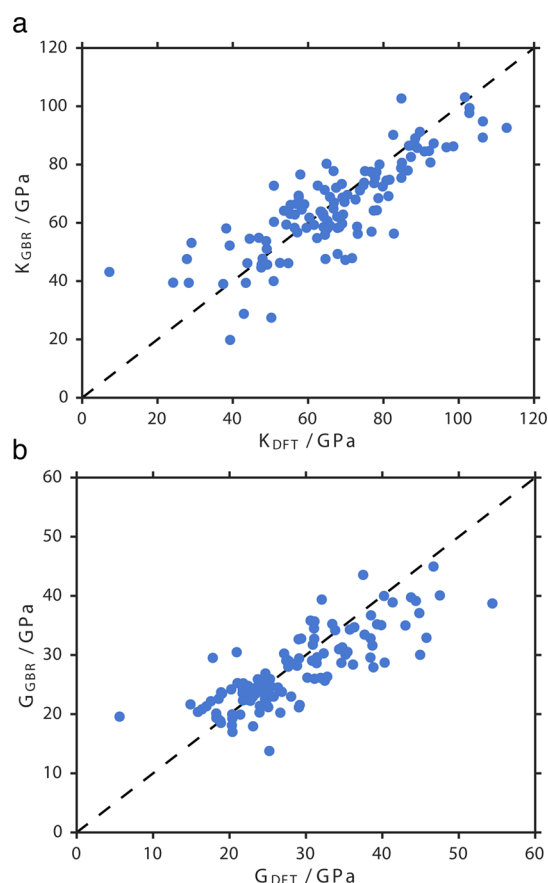
Figure 2. Comparison of cross-validated DFT training set with GBR predictions for bulk modulus K (a) and shear modulus G (b) of 121 pure silica zeolites.

calculations of structures, energies, and mechanical properties,[41] clearly illustrate the effectiveness of this machine learning approach. Similar predictions for the elastic moduli of pure-silica zeolites were reported by Siddorn and co-workers[3] using five different classical interatomic potentials: BKS,[42] Catlow,[43] Gale,[44] Sastre,[45] and Matsui models.[46] Figure 3 displays the comparison between the bulk and shear moduli calculated by DFT methods and the results obtained with the five classical potentials. We can see that the spread of the plots is similar or worse than the predictions obtained by our machine learning approach. Moreover, we can also see some trends in the force fields themselves with systematic errors: most force fields tend to systematically overestimate the bulk modulus, apart from the Matsui potential, which, however, systematically underestimates the shear modulus.

The capability of the machine learning method is evident by comparison of the RMSE observed for each of the force field models compared against the DFT data set presented in Table 2. This important comparison of different models for simulating

Table 2. RMSE for the GBR Model Reported in This Study and Conventional Model Potentials in Comparison to the DFT Dataset

method	log(K) RMSE	log(G) RMSE
GBR	0.102 ± 0.034	0.0847 ± 0.022
BKS	0.153	0.335
Catlow	0.131	0.0817
Gale	0.117	0.154
Sastre	0.130	0.179
Matsui	0.147	0.259

the elastic response of zeolites is noticeably lacking from the literature. The machine learning model provides excellent accuracy when compared to force fields for prediction of bulk modulus, with all force fields within the same RMSE range. Two models treat shear moduli (G) poorly: the BKS and Matsui potentials, while the Catlow potential shows the best agreement with the DFT data. Importantly, the GBR model

computational methods is a difficult challenge, and the values obtained are typically associated with a relatively high uncertainty. A comparison with conventional model potentials (force fields), widely used in the field of zeolite science for
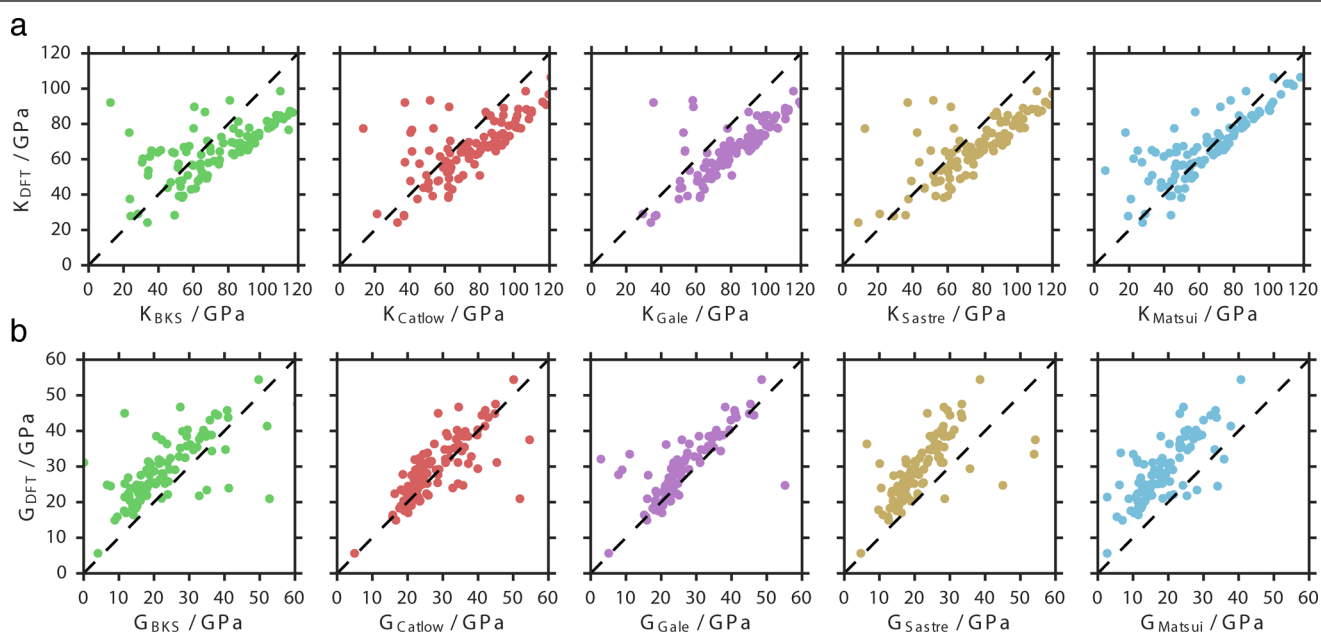

Figure 3. Comparison of DFT training set to the results obtained by conventional model potentials for K (a) and G (b).

described here is equally accurate to the best conventional approaches and shows greater accuracy in treating the shear modulus G.

Model Features. Descriptors often do not contribute equally to a machine learning model, and by interpreting a trained model, we can identify the crucial features. One advantage of the regression trees used by the GBR model is that it intrinsically executes feature selection performed during construction of the trees. This information can be readily interpreted to find the relative influence of each of the chosen descriptors. This information is plotted in Figure S2, showing the weight of the top contributing descriptors in the GBR model. We find that local descriptors significantly contribute to the model for both K and G: parameters related to the Si−O bonds and Si−O−Si angles have the largest weight in the model. This can be confirmed by direct correlation plots between specific descriptors and mechanical moduli. In particular, as illustrated in Figure 4, the harmonic mean of
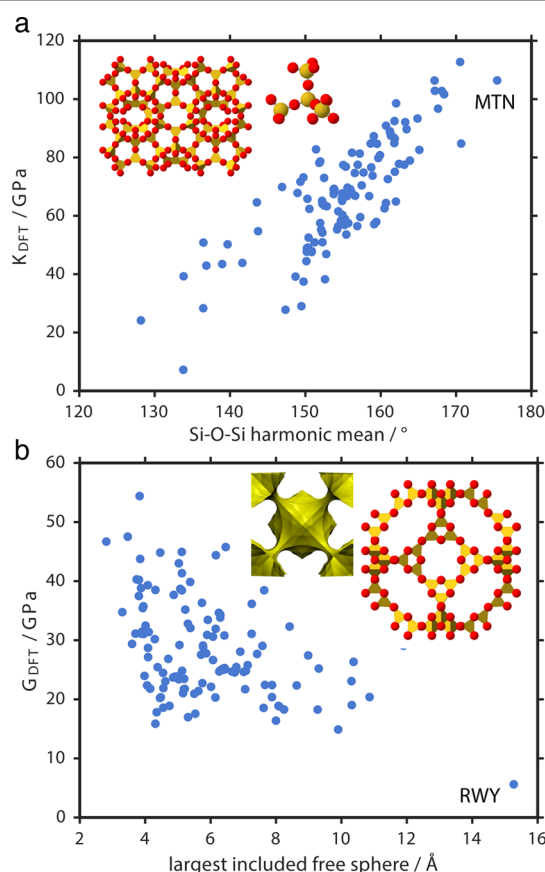

Figure 4. (a) Correlation between K and Si−O−Si harmonic mean in the DFT training set. (a, inset) An example of zeolite MTN which shows very linear Si−O−Si angles and large K. (b) Correlation between G and diameter of the largest included free sphere in the DFT training set. (b, inset) An example of the zeolite RWY which shows a large internal void and very low G.

the Si−O−Si angles is strongly correlated with K. Frameworks which contain mostly linear Si−O−Si angles such as MTN result in very stiff materials with large K. Furthermore, for prediction of G, it is observed that a porosity descriptor, the largest included free sphere, plays a crucial role. Structures containing large cages such as RWY, which has a cage greater than 15 Å in diameter, exhibit extremely low G.

The importance of Si−O−Si angles within a zeolite framework for the resulting elastic properties has previously been reported in the case of α-cristobalite.[47] In this study, we demonstrate that a strong correlation exists in over 100 zeolites. Moreover, this analysis of feature importances is in line with results obtained in 2004 by Monson and co-workers for five silica zeolites (SOD, LTA, CHA, MOR, and MFI).[48] We demonstrate that Si−O−Si angles, volume per SiO_2 unit, and space group all contribute to the bulk modulus K of a zeolite. While Si−O−Si angles and volume per SiO_2 unit are also crucial for predicting G, additional descriptors relating to Si−O bond lengths and porosity are influential.

Mechanical properties such as K and G are important considerations for the practical applications of zeolites as experimentally synthesized frameworks are among the most mechanically stable.[5] The results provided here demonstrate a clear correlation between several characteristic features of the framework, and these properties providing an unparalleled understanding of zeolites. Moreover, we expect similar correlations to be present in related families of microporous materials such as metal−organic frameworks and zeolitic imidazolate frameworks.

The GBR model, using these simple descriptors, provides excellent predictive performance without the need for complex descriptors or the use of explicit molecular simulation at the classical or quantum level. A preliminary study indicates, however, that this cannot be directly extended to predict the anisotropy of elastic properties or directional properties such as linear compressibility and Poisson's ratio. There, we expect that novel and more intricate descriptors may be required, including some global measure of the framework topology, connectivity, and directionality. In addition, we also investigated the possibility of using energy-based descriptors in the machine learning model such as the energy per SiO_2 unit relative to α-quartz. We find that the addition of this new descriptor to the model does not improve its accuracy in K and G in a statistically significant way. This means that relative energy (or enthalpy of formation) and elastic properties are not linked. It also means, on a more practical level, that by excluding energy from our model, we can still allow large-scale prediction of elastic moduli while using only local, structural, and porosity features, which are significantly cheaper to calculate than energy.

Application to a Hypothetical Database. The statistical model developed in this work is able to predict the elastic moduli of a zeolite in fractions of a second which, in turn, allows for the large-scale prediction of structural databases. To illustrate the effectiveness of our model we employed it to predict K and G for a hypothetical all-silica zeolite database described by Deem and co-workers, PCOD2.[49,50] This database contains over 500 000 frameworks obtained by a procedure that combines Monte Carlo simulation, simulated annealing, and refinement using interatomic potentials. These structures, available in CIF format, also contain the full stiffness tensor of second-order elastic constants, which were calculated with the GULP software[51] using the BKS interatomic potential. However, we find that the results obtained by this method have low reliability, with values of K varying between −27 000 and 20 500, and a significant fraction ($\simeq$8%) of structures being not mechanically stable (i.e., having at least one negative eigenvalue to the elastic tensor[52]). We thus used only the structural information in the database and relied on our machine learning approach to provide estimates for K and G with greater

accuracy. In addition, we conducted predictions for the known zeolite frameworks recognized by the International Zeolite Association (IZA) listed in the Supporting Information.

Descriptors were calculated for a total of 590 448 zeolite frameworks from the PCOD2 database and 195 zeolite frameworks from the IZA. Our GBR model was trained using the DFT training set with hyperparameters described previously. The predicted results for K and G give distributions illustrated in Figure 5, plotted against framework density. While
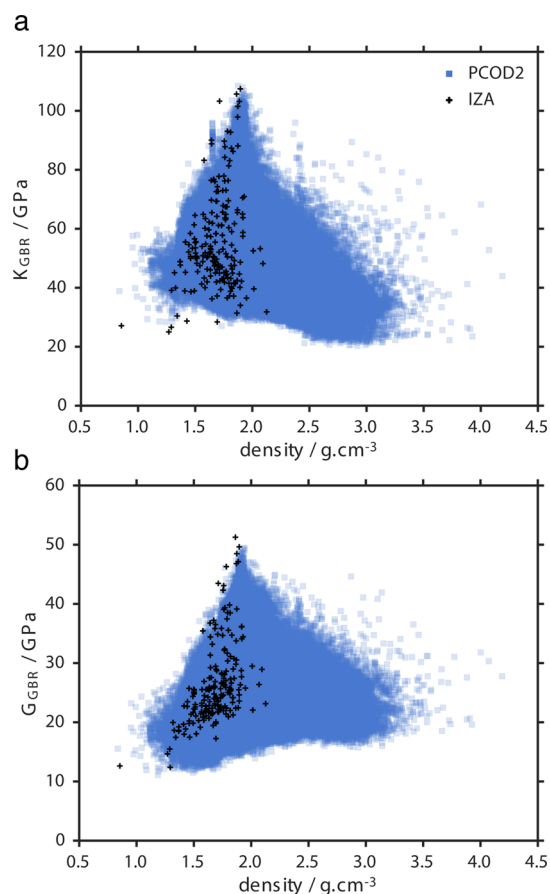

Figure 5. K (a) and G (b) of the PCOD2 database and recognized zeolite frameworks (IZA) as predicted by the GBR model. PCOD2 results are illustrated with opacity to demonstrate the population density.

the GBR algorithm is unable to extrapolate, we do not find this to be a limitation for this data set with predictions of few zeolites exhibiting moduli similar to the bounds of the training structures. Moreover, we find the distributions of geometric features, key to the GBR model, are similar for the data sets, as illustrated in the Supporting Information. The large-scale simulation of the PCOD2 database clearly demonstrates a lack of correlation between K, G, and density in the full space of hypothetical zeolites. Notably, most dense polymorphs show significantly less stiffness than those of low density. This relationship highlights the diverse mechanical response that can be achieved by the arrangement of tetrahedra. Furthermore, the lower K and G displayed by high-density zeolites may suggest a mechanical advantage to the formation of low density and microporous frameworks by SiO_2.

To examine the distributions of K and G, Gaussian kernel density estimates[53] were calculated, as displayed in Figure 6
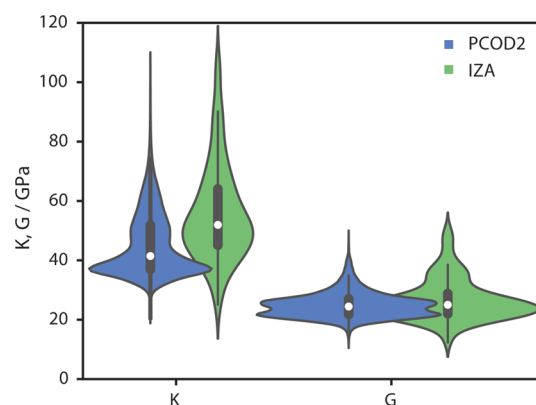

Figure 6. Violin plots of K and G for both the PCOD2 database and recognized zeolite frameworks (IZA), as predicted by the GBR model. Distributions are calculated by Gaussian kernel density estimates, and interquartile ranges are depicted by internal lines.

with corresponding interquartile ranges. Zeolite frameworks belonging to the PCOD2 database are observed to largely have K and G moduli less than 40 and 30 GPa, respectively. The skew of this distribution suggests most hypothetical structures are significantly less stiff and shear-resistant than the known zeolite frameworks recognized by the IZA.

There have been a number of studies discussing feasibility criteria of zeolite frameworks,[54] and it was previously reported that the known zeolite frameworks lie at the low energy-density edge of the hypothetical SiO_2 database distribution.[50] We can add to these criteria that, in general, feasible zeolite frameworks are characterized by stiff bulk moduli approximately greater than 40 GPa. Moreover, this large-scale analysis suggests that the limits of stiffness and shear resistance are highlighted within the known frameworks, BCT and BIK, with few hypothetical polymorphs found with similar mechanical resistance.

Finally, the mechanical properties of this large library of hypothetical zeolites can be leveraged with the calculated porosity features to discover relationships which were previously difficult to recognize owing to the relatively few zeolites for which elastic moduli have been calculated. There have been a number of studies examining porosity of large databases;[55−57] however, the mechanical properties of the materials are often neglected despite providing valued insight into the inherent stability of the structure.[58] Illustrated in Figure 7 is the accessible pore volume (for a probe size equivalent to He) plotted against G.

There is an observed limit of pore volume achieved by IZA zeolites of approximately 0.2 $cm^3 \cdot g^{-1}$ with the noticeable exception of the highly porous networks RWY and IRR. These structures display only moderate shear resistance <20 GPa. Generally, there is a distinct decrease in maximum shear moduli exhibited for frameworks of large pore volume. This provides clear evidence to experimental observations of structural instabilities of large pore zeolites and similar stability trends observed in related materials. Moreover, this relationship is likely the cause for the lack of extra-large-pore all-silica zeolites with heteroatoms and additional ions required for the synthesis of such structures.[59]

■ CONCLUSION

Here, we demonstrated a novel machine learning methodology and descriptors for the large-scale prediction of elastic bulk and shear moduli for all-silica zeolites. Our methodology combines
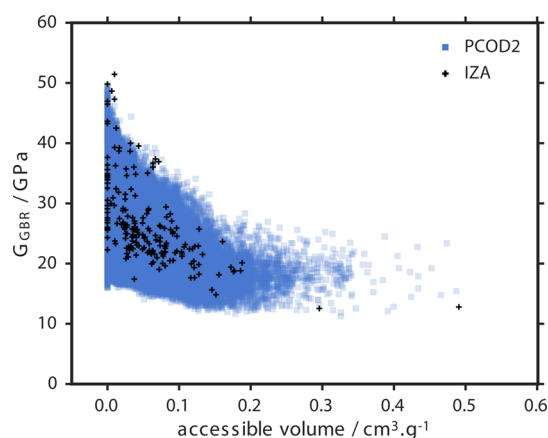
Figure 7. G of the PCOD2 database and recognized zeolite frameworks (IZA), as predicted by the GBR model, plotted against the accessible pore volume calculated for a probe radius of 1.3 Å. PCOD2 results are illustrated with opacity to demonstrate the population density.

GBR using regression trees and a set of local, structural, and porosity features to generalize zeolite materials. The development of this model has highlighted several correlations which link characteristic features of a zeolite to these important elastic properties. Subsequently, training this model with a DFT data set allows for accurate predictions of K and G for any zeolitic compound. We illustrated this power by calculating the moduli for 590 448 hypothetical zeolites. This large-scale study provided general features of SiO_2 polymorphs and the clear decrease in maximum shear resistance for increasing pore volume.

This study is focused entirely on the zeolite family; however, we believe that a machine learning approach is a key tool in the multiscale characterization of porous materials[19] to allow for the screening of millions of materials. Future studies will examine more complex features of zeolites to provide a model for prediction of anisotropy for the screening of new frameworks for applications as pressure sensors and nano-actuators. Furthermore, this approach can be readily modified to describe new and emerging databases of porous materials such as metal–organic frameworks and molecular crystals to provide new maps of energy–structure–function,[60] high-lighting new materials for an application which are also mechanically stable.

ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemmater.7b02532.

> Additional computational details (PDF)
> Descriptors and mechanical properties of the DFT training set (XLSX)
> Descriptors and predicted mechanical properties of the IZA database (XLSX)
> Descriptors and predicted mechanical properties of the hypothetical zeolite database (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: jack.evans@chimie-paristech.fr.

ORCID 🟢

Jack D. Evans: 0000-0001-9521-2601

François-Xavier Coudert: 0000-0001-5318-3910

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Auerbach, S.; Carrado, K.; Dutta, P. *Handbook of Zeolite Science and Technology*; CRC Press: Boca Raton, FL, 2003.

(2) Davis, M. E. Ordered porous materials for emerging applications. *Nature* **2002**, *417*, 813−821.

(3) Siddorn, M.; Coudert, F.-X.; Evans, K. E.; Marmier, A. A systematic typology for negative Poisson's ratio materials and the prediction of complete auxeticity in pure silica zeolite JST. *Phys. Chem. Chem. Phys.* **2015**, *17*, 17927−17933.

(4) Brańka, A. C.; Heyes, D. M.; Wojciechowski, K. W. Auxeticity of cubic materials. *Phys. Status Solidi B* **2009**, *246*, 2063−2071.

(5) Coudert, F.-X. Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16012−16018.

(6) Evans, J. D.; Fraux, G.; Gaillac, R.; Kohen, D.; Trousselet, F.; Vanson, J.-M.; Coudert, F.-X. Computational Chemistry Methods for Nanoporous Materials. *Chem. Mater.* **2017**, *29*, 199−212.

(7) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Ande, C. K.; van der Zwaag, S.; Plata, J. J.; Toher, C.; Curtarolo, S.; Ceder, G.; Persson, K. A.; Asta, M. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2015**, *2*, 150009.

(8) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

(9) Davies, D. W.; Butler, K. T.; Jackson, A. J.; Morris, A.; Frost, J. M.; Skelton, J. M.; Walsh, A. Computational Screening of All Stoichiometric Inorganic Materials. *Chem.* **2016**, *1*, 617−627.

(10) Kramer, G. J.; Farragher, N. P.; van Beest, B. W. H.; van Santen, R. A. Interatomic force fields for silicas, aluminophosphates, and zeolites: Derivation based onab initiocalculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 5068−5080.

(11) Le, T. C.; Winkler, D. A. Discovery and Optimization of Materials Using Evolutionary Approaches. *Chem. Rev.* **2016**, *116*, 6107−6132.

(12) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.

(13) Jordan, M. I.; Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255−260.

(14) Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* **2001**, *26*, 5−14.

(15) Nosengo, N. Can artificial intelligence create the next wonder material? *Nature* **2016**, *533*, 22−25.

(16) Karthikeyan, M.; Glen, R. C.; Bender, A. General Melting Point Prediction Based on a Diverse Compound Data Set and Artificial Neural Networks. *J. Chem. Inf. Model.* **2005**, *45*, 581−590.

(17) Lusci, A.; Pollastri, G.; Baldi, P. Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1563−1575.

(18) Wicker, J. G. P.; Cooper, R. I. Will it crystallise? Predicting crystallinity of molecular materials. *CrystEngComm* **2015**, *17*, 1927−1934.

(19) Evans, J. D.; Huang, D. M.; Haranczyk, M.; Thornton, A. W.; Sumby, C. J.; Doonan, C. J. Computational identification of organic porous molecular crystals. *CrystEngComm* **2016**, *18*, 4133−4141.

(20) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 2844−2854.

(21) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27*, 4459−4475.

(22) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO2Capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056−3060.

(23) Hall, S. R.; Allen, F. H.; Brown, I. D. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 655−685.

(24) Yang, S.; Lach-hab, M.; Vaisman, I. I.; Blaisten-Barojas, E. Identifying Zeolite Frameworks with a Machine Learning Approach. *J. Phys. Chem. C* **2009**, *113*, 21721−21725.

(25) de Jong, M.; Chen, W.; Notestine, R.; Persson, K.; Ceder, G.; Jain, A.; Asta, M.; Gamst, A. A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-nary Inorganic Polycrystalline Compounds. *Sci. Rep.* **2016**, *6*, 34256.

(26) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.

(27) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134−141.

(28) Martin, R. L.; Smit, B.; Haranczyk, M. Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials. *J. Chem. Inf. Model.* **2012**, *52*, 308−318.

(29) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(30) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(31) Nada, R.; Nicholas, J. B.; McCarthy, M. I.; Hess, A. C. Basis sets for ab initio periodic Hartree?Fock studies of zeolite/adsorbate interactions: He, Ne, and Ar in silica sodalite. *Int. J. Quantum Chem.* **1996**, *60*, 809−820.

(32) Hill, R. The Elastic Behaviour of a Crystalline Aggregate. *Proc. Phys. Soc., London, Sect. A* **1952**, *65*, 349−354.

(33) Gaillac, R.; Pullumbi, P.; Coudert, F.-X. ELATE: an open-source online application for analysis and visualization of elastic tensors. *J. Phys.: Condens. Matter* **2016**, *28*, 275201.

(34) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* **2001**, *29*, 1189−1232.

(35) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer Series in Statistics: Berlin, 2001.

(36) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(37) Mohan, A.; Chen, Z.; Weinberger, K. Web-search ranking with initialized gradient boosted regression trees. *Proceedings of the Learning to Rank Challenge*; PMLR: Haifa, Israel, 2011; pp 77−89.

(38) Li, H.; Liang, Y.; Xu, Q. Support vector machines and its applications in chemistry. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 188−198.

(39) Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. *ICML '06 Proceedings of the 23rd International Conference on Machine Learning*; ACM: New York, 2006; pp 161−168.

(40) Alexander, D. L. J.; Tropsha, A.; Winkler, D. A. Beware ofR2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* **2015**, *55*, 1316−1322.

(41) Combariza, A. F.; Gomez, D. A.; Sastre, G. Simulating the properties of small pore silicazeolites using interatomic potentials. *Chem. Soc. Rev.* **2013**, *42*, 114−127.

(42) van Beest, B. W. H.; Kramer, G. J.; van Santen, R. A. Force fields for silicas and aluminophosphates based onab initiocalculations. *Phys. Rev. Lett.* **1990**, *64*, 1955−1958.

(43) Sanders, M. J.; Leslie, M.; Catlow, C. R. A. Interatomic potentials for SiO2. *J. Chem. Soc., Chem. Commun.* **1984**, 1271−1273.

(44) Gale, J. D. Analytical Free Energy Minimization of Silica Polymorphs. *J. Phys. Chem. B* **1998**, *102*, 5423−5431.

(45) Sastre, G.; Gale, J. D. Derivation of an Interatomic Potential for Germanium- and Silicon-Containing Zeolites and Its Application to the Study of the Structures of Octadecasil, ASU-7, and ASU-9 Materials. *Chem. Mater.* **2003**, *15*, 1788−1796.

(46) Tsuneyuki, S.; Tsukada, M.; Aoki, H.; Matsui, Y. First-Principles Interatomic Potential of Silica Applied to Molecular Dynamics. *Phys. Rev. Lett.* **1988**, *61*, 869−872.

(47) Yeganeh-Haeri, A.; Weidner, D. J.; Parise, J. B. Elasticity of agr-Cristobalite: A Silicon Dioxide with a Negative Poisson's Ratio. *Science* **1992**, *257*, 650−652.

(48) Astala, R.; Auerbach, S. M.; Monson, P. A. Density Functional Theory Study of Silica Zeolite Structures: Stabilities and Mechanical Properties of SOD, LTA, CHA, MOR, and MFI. *J. Phys. Chem. B* **2004**, *108*, 9208−9215.

(49) Earl, D. J.; Deem, M. W. Toward a Database of Hypothetical Zeolite Structures. *Ind. Eng. Chem. Res.* **2006**, *45*, 5449−5454.

(50) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. Computational Discovery of New Zeolite-Like Materials. *J. Phys. Chem. C* **2009**, *113*, 21353−21360.

(51) Gale, J. D.; Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **2003**, *29*, 291−341.

(52) Mouhat, F.; Coudert, F.-X. Necessary and sufficient elastic stability conditions in various crystal systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 224104.

(53) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, 1986; Vol. 26.

(54) Li, Y.; Yu, J.; Xu, R. Criteria for Zeolite Frameworks Realizable for Target Synthesis. *Angew. Chem., Int. Ed.* **2013**, *52*, 1673−1677.

(55) Haldoupis, E.; Nair, S.; Sholl, D. S. Pore size analysis of > 250000 hypothetical zeolites. *Phys. Chem. Chem. Phys.* **2011**, *13*, 5053−5060.

(56) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* **2011**, *4*, 83−89.

(57) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nat. Mater.* **2012**, *11*, 633−641.

(58) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F.-X. Investigating the Pressure-Induced Amorphization of Zeolitic Imidazolate Framework ZIF-8: Mechanical Instability Due to Shear Mode Softening. *J. Phys. Chem. Lett.* **2013**, *4*, 1861−1865.

(59) Jiang, J.; Yu, J.; Corma, A. Extra-Large-Pore Zeolites: Bridging the Gap between Micro and Mesoporous Structures. *Angew. Chem., Int. Ed.* **2010**, *49*, 3120−3145.

(60) Pulido, A.; et al. Functional materials discovery using energy−structure−function maps. *Nature* **2017**, *543*, 657−664.