

High-throughput Calculation of Atomic Planar Density for Compounds

STERLING G. BAIRD* AND TAYLOR D. SPARKS

*Department of Materials Science & Engineering, University of Utah, Salt Lake City,
Utah 84112 United States. E-mail: sterling.baird@icloud.com*

Mathematica; CIF; Materials Project

Abstract

A large collection of element-wise planar densities for compounds obtained from Materials Project is calculated using brute force computational geometry methods. We demonstrate that the element-wise max lattice plane densities can be useful as machine learning features. The methods described here are implemented in an open-source Mathematica package hosted at <https://github.com/sgbaird/LatticePlane>.

1. Introduction

Atomic arrangement within planes and by extension, planar density (the number of atoms per unit area on a plane), is relevant to applications such as elasticity (Rabiei *et al.*, 2020), oxidation (Ahn *et al.*, 2011), surface-energy (Wang, 2020), thermoelectrics (Snyder & Toberer, 2008), nanoscale materials patterning (Liu *et al.*, 2016), and magnetism (Williams, 1937), all of which can exhibit anisotropy with respect to crystallographic direction: e.g. bulk modulus (Fine *et al.*, 1984; Holec *et al.*, 2012; Yu & Liu, 2019).

Traditionally, planar density calculations have been performed on an individual basis and manually (Fan, 2016). With the availability of consistent representation of arbitrary crystals in the form of crystallographic information files (CIFs) (Hall *et al.*, 1991), extensive databases that contain collections of CIFs such as Materials Project (Jain *et al.*, 2013) or Open Quantum Materials Database (Saal *et al.*, 2013), and advanced 3D computational geometry libraries in software such as Mathematica (Wolfram, 2021), MATLAB (MATLAB, 2021), and Python (Vollprecht, 2021), automated computation of planar densities of arbitrary crystals for arbitrary lattice planes is possible. Indeed, in this work, we present such a workflow using Mathematica’s excellent built-in computational geometry functions in conjunction with MaXrd (Ramsnes *et al.*, 2019) which facilitates both importing CIF files and dealing with crystal symmetry domain knowledge. This package is implemented on the full collection of Materials Project compounds as obtained via the Materials API (Ong *et al.*, 2015). This represents the largest collection of planar densities to our knowledge with lattice planes up to a min and max Miller index of -3 and 3, respectively.

While a neat analytical formula to describe lattice plane densities in arbitrary crystal structures may exist, our method is a brute force approach which relies on numerical geometrical computations. Additionally, this provides flexibility in treatment of the lattice sites (Section 2.3) which otherwise might require extensive tailoring or entirely new analytical formulas to describe. Indeed, a relevant analytical approach has been published (Fan, 2016); however, the correctness of this method has been questioned (Liu *et al.*, 2020) without further rebuttal as of July 2021. Additionally, our database has the benefit of containing both summed and element-wise atomic planar densities (i.e. planar densities of each of the respective periodic elements).

We describe the methods for calculating lattice plane density (Section 2) and demonstrate that this database can be used to supply features to machine learning models

to predict material properties via a case study of bulk modulus (Section 3).

2. Planar Density Methods

We describe our approach for calculating atomic planar densities for an arbitrary Miller plane in Section 2.1. When computing planar densities for multiple Miller planes and to enhance computational efficiency, we consider only the unique Miller planes with respect to crystallographic symmetry. We then cast these into the degenerate (full) representation as a post-processing step (Section 2.2). These methods have been incorporated into a Mathematica package called `LatticePlane` hosted at <https://github.com/sgbaird/LatticePlane>. A description of various model parameters is given in Section 2.3. Bulk downloading of Materials Project CIF files and target property data is handled via the Materials API (Ong *et al.*, 2015).

2.1. Calculation of Lattice Plane Density

Calculating the planar density for a Miller plane takes on a brute force approach using numerical computation and involves the following steps separated into Setup, Intersections, and Post-Processing categories:

1. Setup

- (a) Import CIF file via `MaXrd`
- (b) Create the unit cell using basis vectors and if applicable, expand into a supercell
- (c) Convert the HKL indices into an infinite plane region

2. Intersections

- (a) Compute intersection between the infinite plane and the unit cell as a polygon region

- (b) Compute area of the polygon
- (c) Filter out atoms further than a specified distance threshold from the polygon
- (d) Represent the subset of atoms element-wise within the unit cell as ball regions (or probability densities)
- (e) Compute intersection between each ball and the polygon (or integrate each probability density within the polygon)

3. Post-Processing

- (a) If treating the atoms as lattice sites/nodes, divide intersected area of each ball by its hemispherical radius
- (b) Sum intersected areas element-wise and divide by polygon area

2.2. Unique Miller Plane Indices

Each crystallographic point group exhibits different types of crystallographic symmetry. For computational efficiency, only the unique Miller planes for a particular point group are computed. These are then cast into the degenerate (full) representation spanning all integer combinations of the possible Miller indices using rules obtained via **MaXrd**. The number of unique Miller planes for several points groups and varying constraints on the max Miller index are given in Table 2. For example, computing all combinations of Miller indices up to a max of 3 for the $m\bar{3}m$ point group involves the computation of 19 unique Miller planes, whereas the triclinic case (no symmetry) involves computation of all 342 combinations. An example of the unique and full planes (truncated) for the $m\bar{3}m$ point group is shown in Figure 1a and b, respectively.

2.3. Mathematica Package Model Parameters

One of the benefits of using a brute force numerical approach to compute planar densities is that the model parameters can be freely tuned for a chosen application. Relevant `LatticePlane` model parameters are given in Table 1.

Atoms/sites can be treated as nodes, hard spheres, or probability distributions. While the code package implements hard sphere and probability distribution models, node density can effectively be obtained by treating the atoms as hard spheres, only considering atoms whose centers are within a tight (e.g. numerical) tolerance of the lattice plane of interest, and normalizing the intersection area by the area of the hard sphere for each atom. Node density values remain unchanged for larger supercells; however, planar densities of finite-radius hard sphere and probability distribution models when sub-hemisphere slices of atoms are considered (i.e. atoms whose centers do not lie exactly on the Miller plane) converge with increasing supercell size. Examples of intersections within a $3 \times 3 \times 3$ supercell are shown in Figure 2.

In terms of other model parameters, atom radius factor scales the covalent radius of the atoms and defaults to a value of 1. Additionally, increasing the max Miller index increases the computational cost (Section 2.2).

3. Potential for Machine Learning

Addition of element-wise max lattice plane density as machine learning features improved the machine learning results relative to training only on elemental presence (Figure 5). The R^2 fit improved from 0.600 ± 0.031 to 0.673 ± 0.029 . Mathematica’s built-in `Predict` function was used. Predictions of bulk modulus using only elemental presence are summarized in Figure 6 to facilitate comparison with the machine learning model which incorporated element-wise max lattice plane densities.

While the machine learning was performed with respect to a bulk property, the

dataset may be better suited to target properties which are planar in nature. An example of such an application is the Open Catalyst Project (Chanussot *et al.*, 2021).

4. Conclusion

We created a computational workflow which calculates element-wise atomic planar densities using crystallographic information files (CIFs) as inputs and applied this to a large database of compounds. We tested the usefulness of this dataset for machine learning applications and found that in a simple test case, the addition of max element-wise atomic planar density as a feature improves the predictive accuracy relative to learning on elemental presence only using the target property of bulk modulus.

5. Future Work

While we chose to focus on node density, incorporation of `pymatgen` (Ong *et al.*, 2013) charge decoration (Composition module) can help with smartly choosing radii for elements based on their covalent, ionic, and metallic radii for the hard sphere and probability density models and may offer better quality information for data science models. We hope to see this dataset used in crystallographic orientation dependent applications such as the Open Catalyst Project in pursuit of surrogate models for computationally complex physics-based simulations.

Acknowledgements

T.D.S. and S.G.B. gratefully acknowledge support from the National Science Foundation (CMMI-1562226 and DMR-1651668).

References

Ahn, J.-J., Jo, Y.-D., Kim, S.-C., Lee, J.-H. & Koo, S.-M. (2011). *Nanoscale Res Lett*, **6**(1), 235.

- Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., Palizhati, A., Sriram, A., Wood, B., Yoon, J., Parikh, D., Zitnick, C. L. & Ulissi, Z. (2021). *arXiv:2010.09990 [cond-mat]*.
- Fan, Q. (2016). *J Appl Crystallogr*, **49**(5), 1454–1458.
- Fine, M., Brown, L. & Marcus, H. (1984). *Scripta Metallurgica*, **18**(9), 951–956.
- Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Crystallogr A Found Crystallogr*, **47**(6), 655–685.
- Holec, D., Friák, M., Neugebauer, J. & Mayrhofer, P. H. (2012). *Phys. Rev. B*, **85**(6), 064101.
- Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G. & a. Persson, K. (2013). *APL Materials*, **1**(1), 011002.
- Liu, H., Hoepfner, S. & Schubert, U. S. (2016). *Adv. Eng. Mater.* **18**(6), 890–902.
- Liu, Y., Liu, Y. & Drew, M. G. B. (2020). *SN Appl. Sci.* **2**(4), 755.
- MATLAB (2021). *Version 9.10.0 (R2021a)*. Natick, Massachusetts: The MathWorks Inc.
- Ong, S. P., Cholia, S., Jain, A., Brafman, M., Gunter, D., Ceder, G. & Persson, K. A. (2015). *Computational Materials Science*, **97**, 209–215.
- Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A. & Ceder, G. (2013). *Computational Materials Science*, **68**, 314–319.
- Rabiei, M., Palevicius, A., Dashti, A., Nasiri, S., Monshi, A., Vilkauskas, A. & Janusas, G. (2020). *Materials*, **13**(19), 4380.
- Ramsnes, S., Larsen, H. B. & Thorkildsen, G. (2019). *J Appl Crystallogr*, **52**(1), 214–218.
- Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. (2013). *JOM*, **65**(11), 1501–1509.
- Snyder, G. J. & Toberer, E. S. (2008). *Nature Materials*, **7**(2), 105–114.
- Vollprecht, W., (2021). Scikit-geometry. scikit-geometry.
- Wang, L. (2020). *Metall Mater Trans A*, **51**(12), 6127–6131.
- Williams, H. J. (1937). *Phys. Rev.* **52**(7), 747–751.
- Wolfram, (2021). Mathematica, Version 12.3.1.
- Yu, F. & Liu, Y. (2019). *Computation*, **7**(4), 57.

iucr

Table 1. *Model parameters used in the Mathematica package **LatticePlane** and a list of corresponding properties.*

Model Parameters	Properties	Defaults
Method	Discrete (Hard Sphere or PDF)	Hard Sphere
Atom radius factor	Positive, continuous	1 (Method=Hard Sphere)
Supercell size	Positive integer	$1 \times 1 \times 1$
Max HKL index	Positive integer	3
Plane distance threshold	Positive, continuous	0.01 Angstroms

Table 2. *Number of unique Miller planes for 7 crystallographic point groups of varying levels of symmetry.*

Point Group (Crystal System)	# of Unique Miller Planes		
	Max HKL = 1	Max HKL = 2	Max HKL = 3
m-3m (Cubic)	3	9	19
6/mmm (Hexagonal)	5	17	39
32 (Trigonal)	8	34	90
4/mmm (Tetragonal)	5	17	39
222 (Orthorhombic)	8	34	90
m (Monoclinic)	17	74	195
1 (Triclinic)	26	124	342

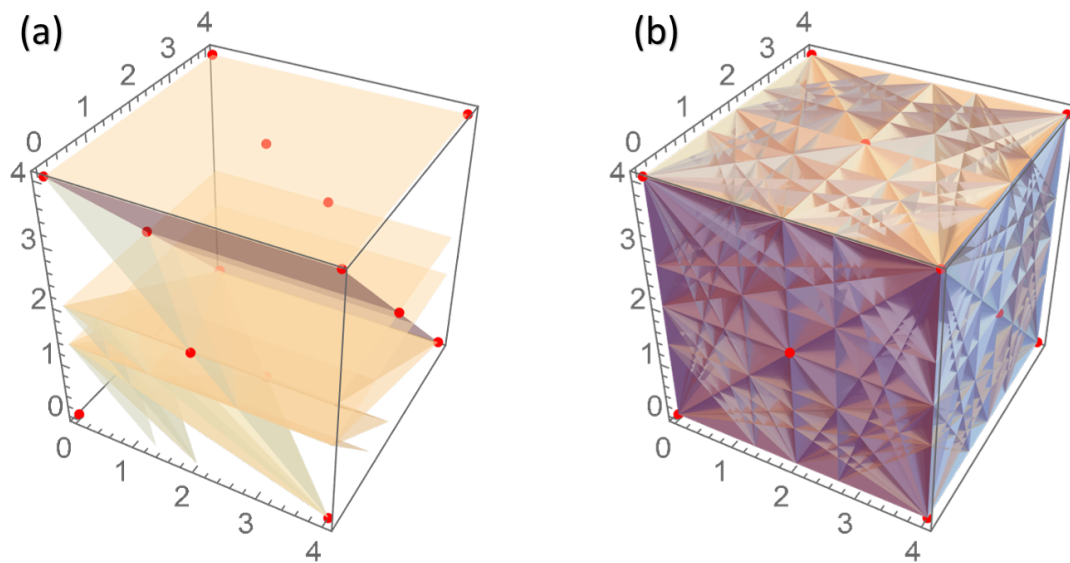


Fig. 1. (a) Nineteen unique and (b) 324 degenerate lattice planes for an Al $Fm\bar{3}m$ crystal structure. Atomic sites are plotted as red points. Shading of planes is for visual demonstration only. Axes are in units of Angstroms.

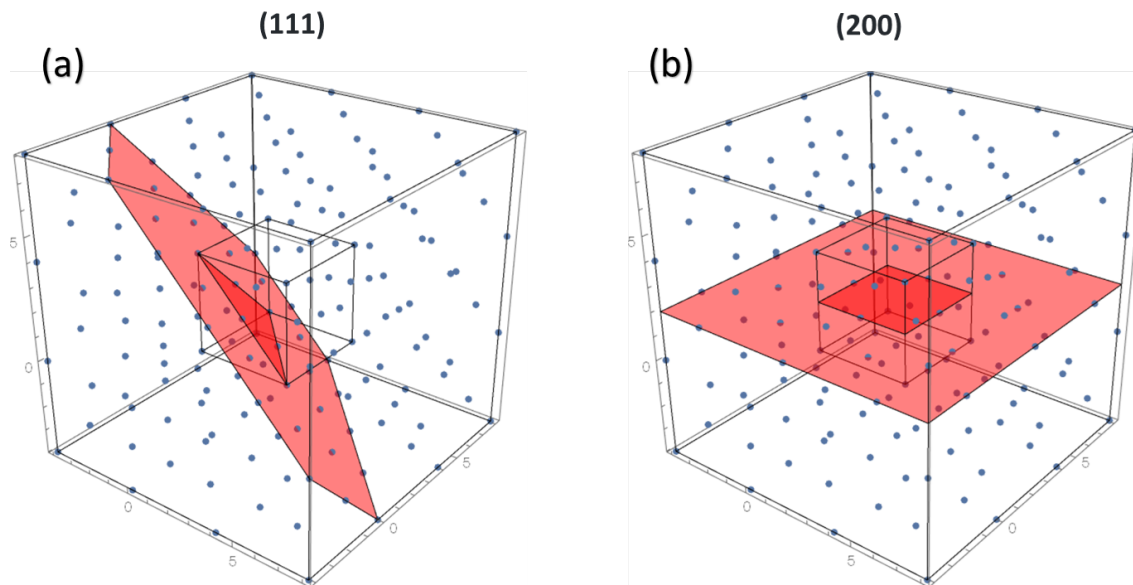


Fig. 2. Example of $3 \times 3 \times 3$ supercell intersections for $Fm\bar{3}m$ in (a) (111) and (b) (200) Miller planes.

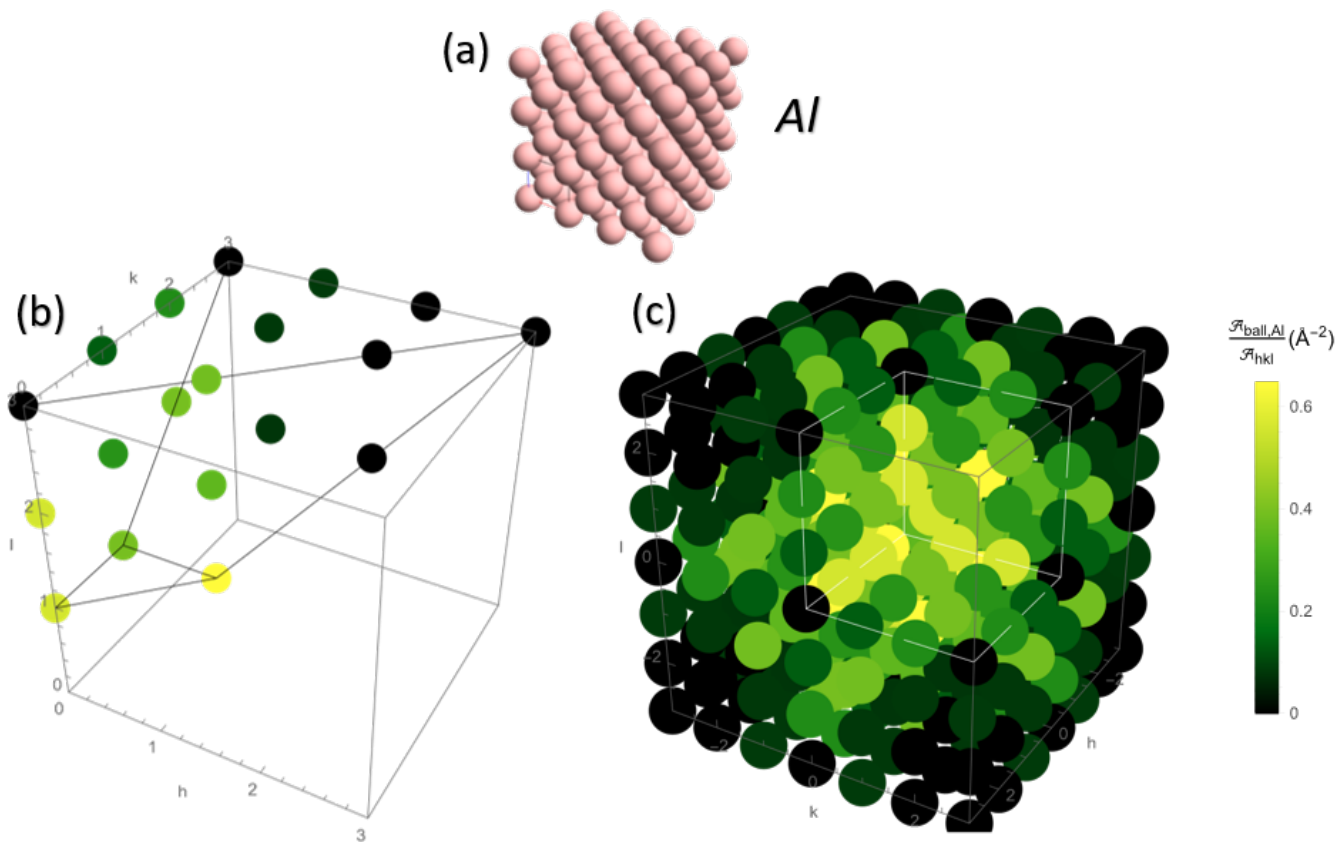


Fig. 3. (a) Crystal plot of $Fm\bar{3}m$ Al (b) unique Miller plane densities (c) degenerate Miller plane densities with positive-valued Miller planes (white wireframe box) removed for visualization purposes. An atom-to-plane distance threshold of 0.1 Angstroms and a max Miller index of 3 were used.

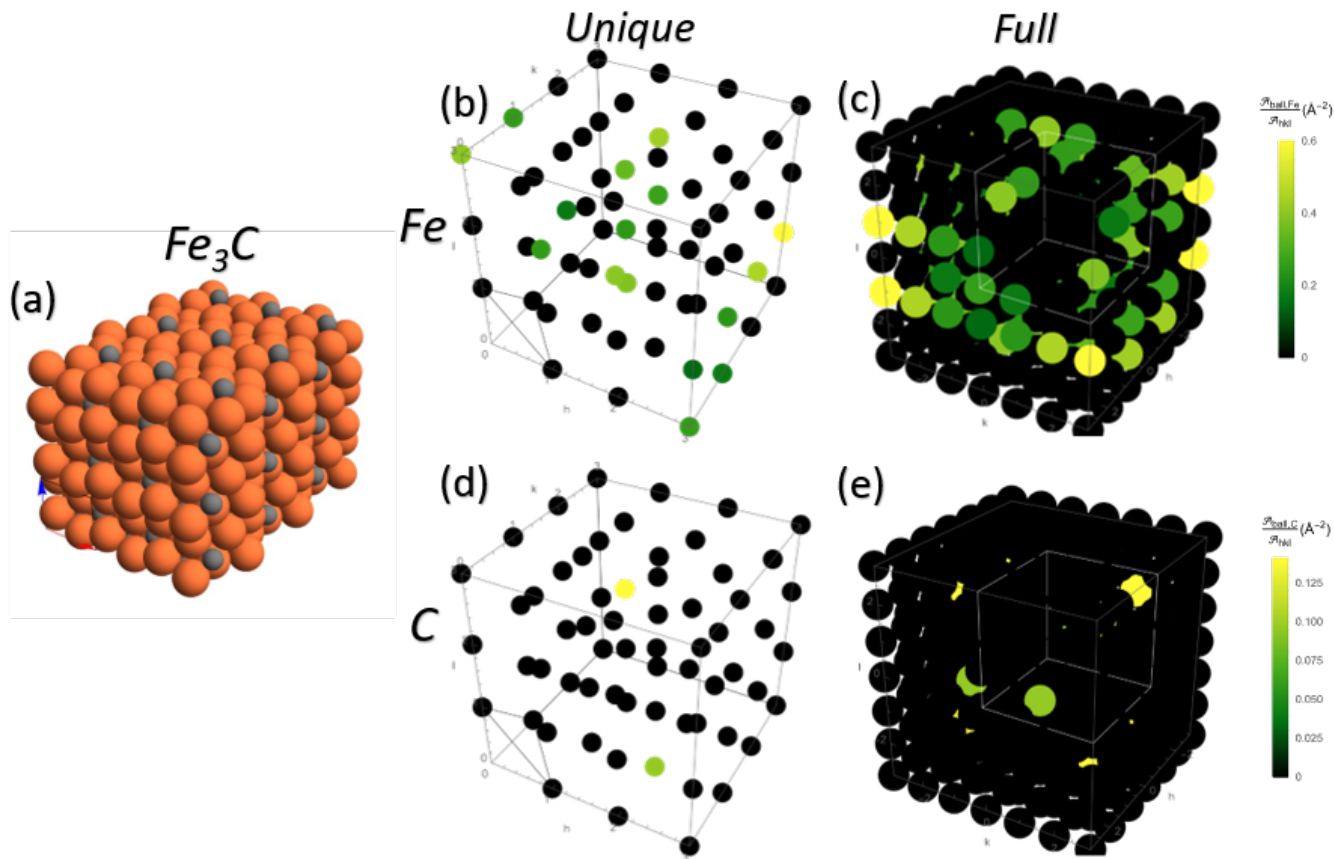


Fig. 4. (a) Crystal plot of $Pnma$ Fe_3C (b) unique Miller plane densities of Fe (c) degenerate Miller plane densities for Fe (d) unique Miller plane densities for C (e) degenerate Miller plane densities for C. For the degenerate (full) plots (c, e), positive-valued Miller planes were removed for visualization purposes. The removed data is bounded by a white wireframe box. An atom-to-plane distance cutoff of 0.1 Angstroms and a max Miller index of 3 were used.

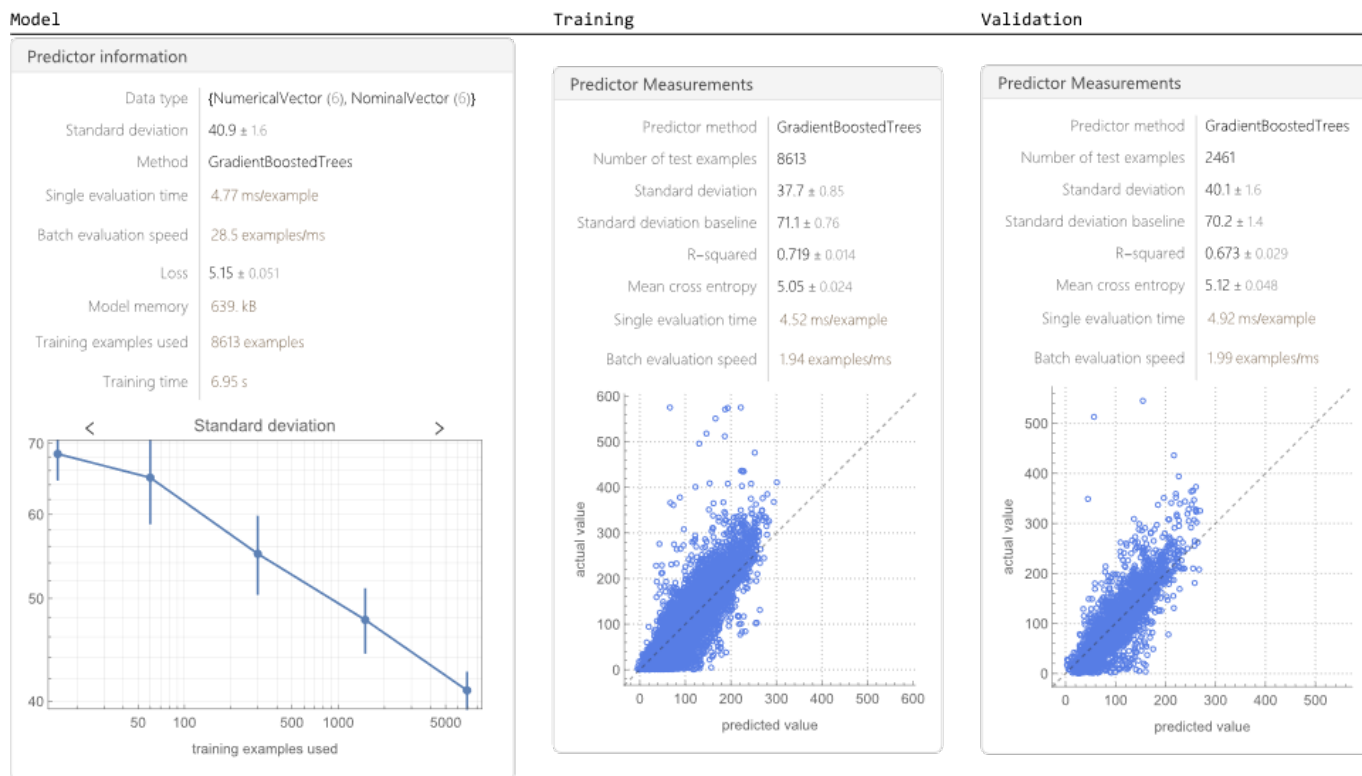


Fig. 5. Machine learning results using element-wise max lattice plane densities as numerical features and elements within the compound as nominal features. Compounds with less than 6 elements were zero-padded.

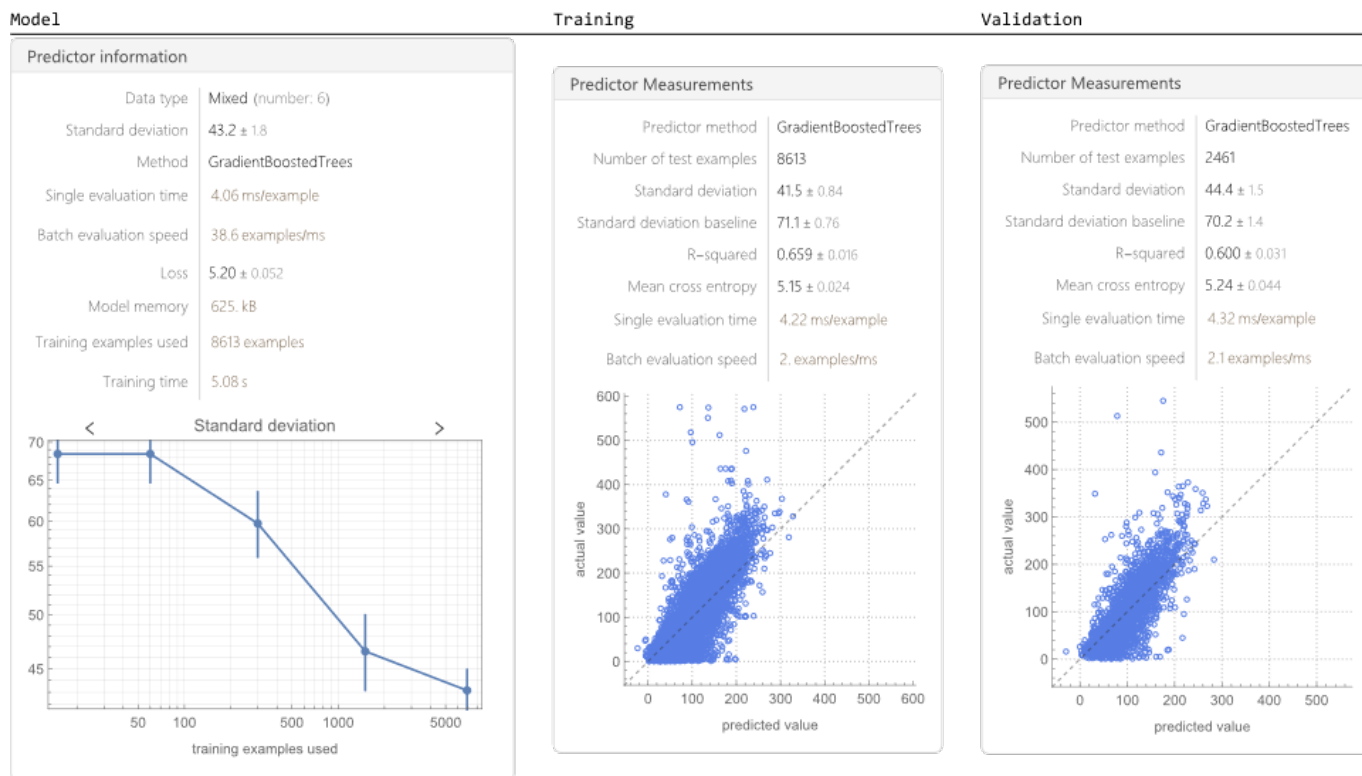


Fig. 6. Machine learning predictions of bulk modulus using only elemental presence as nominal data. Compounds with less than 6 elements were zero-padded.

Synopsis

A large database of CIF files is used in conjunction with computational geometry software to calculate element-wise planar densities up to a max HKL index of 3.