# Machine Learning Energies of 2 Million Elpasolite ($ABC_2D_6$) Crystals

Felix A. Faber,[1] Alexander Lindmaa,[2] O. Anatole von Lilienfeld,[1,3,*] and Rickard Armiento[2,†]
[1]Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[2]Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden
[3]General Chemistry, Free University of Brussels, Pleinlaan 2, 1050 Brussels, Belgium
(Received 24 August 2015; published 20 September 2016)

Elpasolite is the predominant quaternary crystal structure ($AlNaK_2F_6$ prototype) reported in the Inorganic Crystal Structure Database. We develop a machine learning model to calculate density functional theory quality formation energies of all $\sim 2 \times 10^6$ pristine $ABC_2D_6$ elpasolite crystals that can be made up from main-group elements (up to bismuth). Our model's accuracy can be improved systematically, reaching a mean absolute error of 0.1 eV/atom for a training set consisting of $10 \times 10^3$ crystals. Important bonding trends are revealed: fluoride is best suited to fit the coordination of the $D$ site, which lowers the formation energy whereas the opposite is found for carbon. The bonding contribution of the elements $A$ and $B$ is very small on average. Low formation energies result from $A$ and $B$ being late elements from group II, $C$ being a late (group I) element, and $D$ being fluoride. Out of $2 \times 10^6$ crystals, 90 unique structures are predicted to be on the convex hull—among which is $NFAl_2Ca_6$, with a peculiar stoichiometry and a negative atomic oxidation state for Al.

DOI: 10.1103/PhysRevLett.117.135502

Elpasolite ($AlNaK_2F_6$) is a glassy, transparent, luster, colorless, and soft quaternary crystal in the $Fm3m$ space group that can be found in the Rocky Mountains, Virginia, or the Apennines. The elpasolite crystal structure (see Fig. 1) is not uncommon; it is the most abundant prototype in the Inorganic Crystal Structure Database [1,2]. Some elpasolites emit light when exposed to ionic radiation, which makes them interesting material candidates for scintillator devices [3,4]. One could use first-principles methods such as density functional theory (DFT) [5,6] to computationally predict the existence and basic properties of every elpasolite. Unfortunately, even when considering crystals composed of only main group elements (columns I to VIII) the sheer number of the $\sim 2 \times 10^6$ possible combinations makes DFT based screening challenging—if not prohibitive. Recently, computationally efficient machine learning (ML) models were introduced for predicting molecular properties with similar accuracy as DFT [7,8]. Requiring only milliseconds per prediction, they represent an attractive alternative when it comes to the combinatorial screening of millions of crystals. While some ML model variants have already been proposed for solids [9–11], a generally applicable ML scheme with the DFT accuracy of formation energies is still amiss.

In this Letter we introduce a newly developed ML model, which we use to investigate the formation energies of all $\sim 2 \times 10^6$ elpasolites made from all main-group elements up to Bi. The resulting estimates enable the identification of a new elemental order of descending elpasolite formation energy, crystals with peculiar atomic charges, the 250 elpasolites with the lowest formation energies, as well as 128 new crystal structures predicted to lie on the convex hull among which is $NFAl_2Ca_6$, an elpasolite with unusual composition and atomic charges. The ML model achieves the same, or better, accuracy with respect to DFT as DFT in comparison to experimental data, and can be generalized to any crystalline material.

The ML model is based on kernel ridge regression [12–14], which maps the nonlinear energy difference between the actual DFT energy and an inexpensive approximate baseline model into a linear feature space [15]. More specifically, we construct a ML model of the energy difference between the crystal energy and the sum of static, atom-type dependent, averaged atomic energy contributions $\epsilon_{It}$, obtained through the fitting of each atomic species $t$ in all main group elements up to Bi. The ML model is a sum of weighted exponentials in similarity $d$ between the query and training crystal. The total energy-predicting model function reads

$$E(\mathbf{x}) = \sum_{I}^{N'} \epsilon_{It} + \sum_{i}^{N} \alpha_i e^{-d_i/\sigma}, \tag{1}$$

where $N'$ is the number of atoms/unit cell (ten in the case of elpasolites), and the second sum runs over all $N$ training instances. $\{\alpha_i\}$ are the weights obtained through linear regression, and $\sigma$ is the global exponential width, regulating

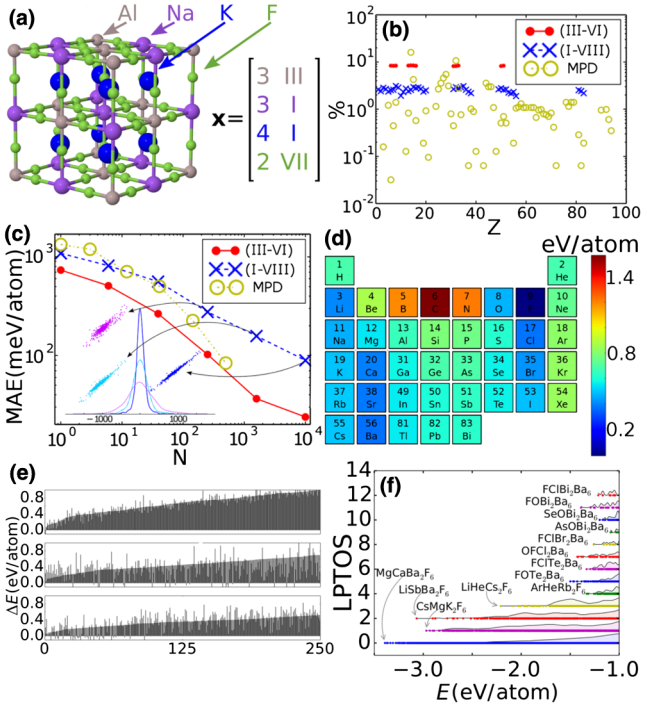


FIG. 1. (a) Illustration of elpasolite crystal ($AlNaK_2F_6$ structure). The 4-tuple $x = (x_1, \ldots, x_4)$ representation of atomic sites is specified. (b) Frequency of elements (defined by nuclear charge $Z$) for the three data sets studied. (c) Mean absolute out-of-sample prediction error as a function of the training set size for the three data sets studied. Inset: error distributions and DFT versus ML scatter plots for the three training set sizes for the (I–VIII) data set. (d) Estimated mean energy contribution of each element to formation of any elpasolite crystal. The color code reflects the new elemental elpasolite order. (e) Lowest 250 ML model predicted formation energies of elpasolites in ascending order from the (III–VI) (top) and (I–VIII) (middle and bottom) data sets. Results in the top and middle panel correspond to ML models trained on 2000 examples. The bottom panel results correspond to a ML model trained on $10 \times 10^3$ crystals. Validating DFT energies are shown aside. (f) Distributions of the absolute lowest possible total oxidation states (LPTOSs) in energies. The formulas indicate the lowest lying crystals.

the length scale of the problem. The similarity $d_i$ is the Manhattan distance, i.e., $d_i = \|\mathbf{x} - \mathbf{x}_i\|_1$. While various crystal structure representations $\mathbf{x}$ have previously been proposed [9–11,16–18], we have found the following representation to yield superior performance: $\mathbf{x}$ is an $(n \times 2$-tuple) that encodes any stoichiometry within a given crystal prototype. For quaternary ($n = 4$) elpasolites, each $x_{1-4}$ refers to the four representative sites; the atom type for each site is represented by its row (principal quantum number 2 to 6) and column (number of valence electrons) I to VIII in the periodic table, and the sites are ordered according to the Wyckoff sequence of the crystal. As such, $\mathbf{x}$ implicitly represents the (local) energy minimum structure for a system restricted to this prototype—without explicitly encoding precise coordinates, lattice constants, or other (approximate) solutions to Schrödinger's equation. This representation is not restricted to the elpasolite structure; it can be used for any fixed crystal symmetry: below we also briefly discuss test results for small size ML models applied to ternary crystals.

For training and evaluation, we have generated DFT formation energies for two data sets of elpasolites (for computational details see the Supplemental Material [19]): one small, (III–VI), made up of only 12 elements, C, N, O, Al, Si, P, S, Ga, Ge, As, Sn, and Sb, and one large, (I–VIII), containing all main-group elements up to Bi. Since (III–VI) only comprises $\sim 12 \times 10^3$ possible permutations, we have obtained the complete list of formation energies. (I–VIII) consists of $10 \times 10^3$ structures, i.e., 0.5% of the total number of $2 \times 10^6$ possible crystals. The (I–VIII) data set has been generated through the random selection of elpasolites while ensuring an unbiased composition. To verify that the ML model is general and not only restricted to elpasolites, we have also included a materials project [28] data set consisting of $\sim 0.5 \times 10^3$ ternary crystals in the $ThCr_2Si_2$ ($I4/mmm$) prototype and made up of 84 different atom types. The distribution of the chemical elements in the data sets is shown in Fig. 1(b). The numerical results on display in Fig. 1(c) indicate a systematic improvement of the predictive accuracy of the ML model with increasing training set size, for all three data sets. The inset details the normally distributed errors and scatter plots, which systematically improve with training set size for the models trained on the (I–VIII) data sets. For a $10 \times 10^3$ training set, the ML model reaches a mean absolute error (MAE) of 0.1 eV/atom compared to the reference, i.e., semilocal DFT. DFT, in turn, has an estimated MAE of $\sim$0.19 eV/atom compared to experiments on heats of formation for general chemistries with filled $d$ shells [29]. Other groups report DFT errors on the order of 0.1 eV/atom for transition metal oxides and elemental solids [30,31]. The converging performance for training on nearly all crystals of the (III–VI) data set suggests that our crystal representation of elpasolite structures [Fig. 1(a)] accounts for all the necessary degrees of freedom. While the errors decay systematically and linearly on a log-log plot, the learning rate levels off as $N$ approaches $10 \times 10^3$. This is due to the employed relaxation convergence threshold of $\pm 10$ meV/atom in the DFT calculations. Any inductive model must fail to go below this level, and only numerically more precise reference numbers would mitigate this trend. In all validation tests dealing with energy predictions for random out-of-sample crystals, the ML model performance meets the expectations set in Fig. 1(c). For example, drawing 100 crystals at random from the (III–VI) and (I–VIII) data sets, the ML models perform as expected when compared to the result from validating the DFT calculations (cf. Fig. S3 [19]). ML models trained on (III–VI) and (I–VIII) reach a MAE of 0.1 eV/atom at roughly 2.5% and 0.5% of the

total number of crystals respectively, suggesting that the machine "efficiency" increases with the number of possible combinations.

Having established the performance of the ML model, we have subsequently used the $10 \times 10^3$ training set model (I–VIII) for investigation of the elpasolite universe. Estimated formation energies for all $2 \times 10^6$ elpasolites are featured in Fig. 2. The formation energies are clearly dominated by the chemical identity of position 4, followed by position 3 but according to a different pattern. The chemical identity at position 1 and 2 has the smallest influence and a very similar impact (also illustrated in Fig. S1 of Ref. [19].) Because of the effective degeneracy of positions 1 and 2, all inner matrices in Fig. 2 appear largely symmetric. Figure 1(d) shows the average contribution of each element to the formation energies estimated by the $10 \times 10^3$ ML model. These average contributions per element are used to order the elements in Fig. 2 to yield the smoothest elpasolite map. Arranging the elements by their nuclear charge, or by their Pettifor order [32], results in a much more oscillatory map or stripelike pattern due to the underlying periodicities (cf. Ref. [19]). This elpasolite order is dominated by the element identity in position 4 [compare Fig. 1(d) to Fig. S1 of Ref. [19]: its breakdown is small as illustrated for pairwise energy contributions in Fig. S5 of Ref. [19].

Figure 1(d) visualizes the bonding emergent from the geometry and bond coordination of the elpasolite crystal structure (see also the figures in Supplemental Material [19]). Fluorine and carbon are at the respective ends of the global scale of low and high formation energies. But also alkaline metals, alkaline earth metals, and oxygen contribute to lowering the formation energy. On average, the formation energies of elpasolites involving halogens, alkaline metals, and noble gases increase as the periodic table is descended. The opposite holds for all other elements, except oxygen, boron, carbon and nitrogen, which all have a noticeably higher average formation energy than any other element. A saddle point can also be observed in the midst of the periodic table as well as two valleys along the halogen and alkaline earth rows. Site-specific resolution indicates that fluorine fits best with the bond coordination

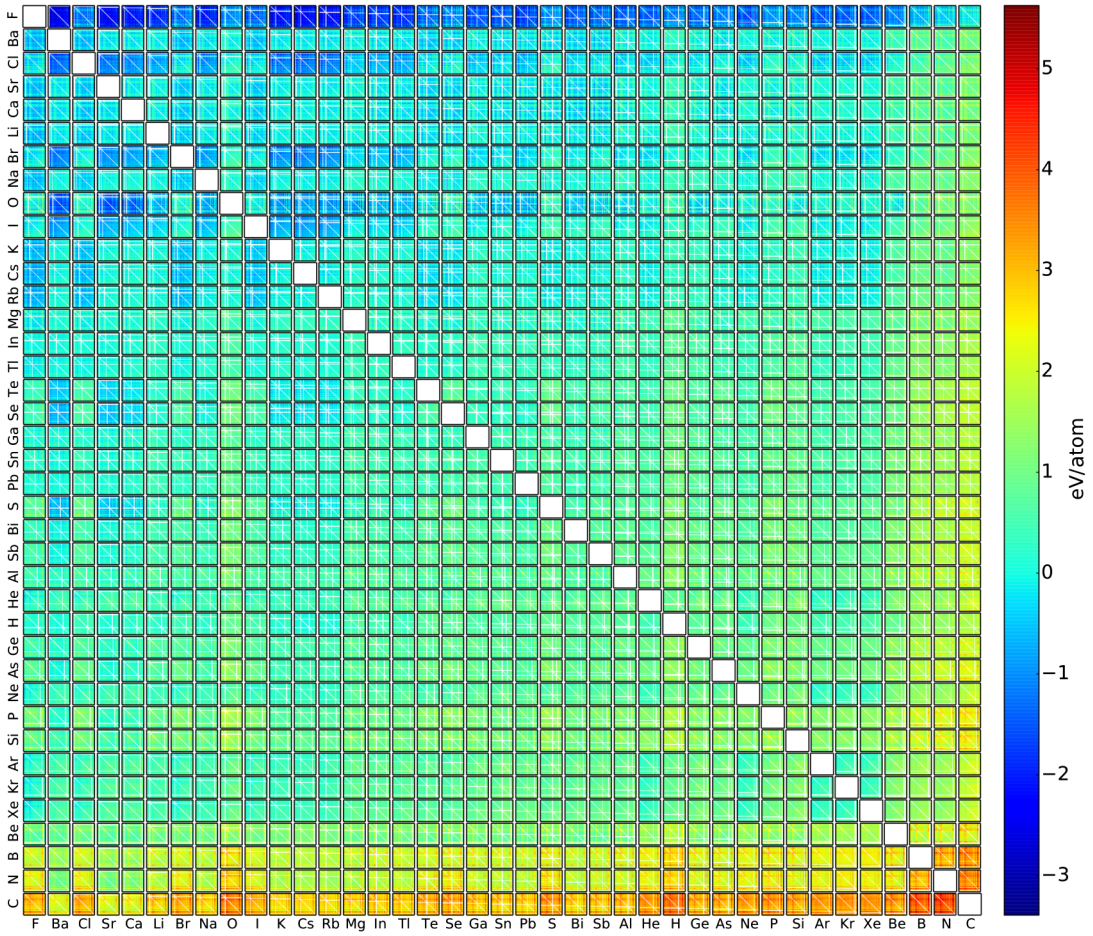


FIG. 2. Formation energies for all $2 \times 10^6$ elpasolites made up of all main-group elements up to Bi predicted by the $10 \times 10^3$ ML model. The outer vertical and horizontal axis corresponds to the $x_4$ and $x_3$ symmetry position, respectively. The inner vertical and horizontal axis corresponds to the $x_2$ and $x_1$ symmetry position, respectively. The elemental sequence follows the elpasolite order of Fig. 1(d). The white pixels correspond to subspaces of ternary, binary, or elementary nonelpasolite crystals.

of sites 1, 2, and 4, whereas the same does not apply to later halogens (not shown in the Letter; see the Supplemental Material [19]). In contrast, as the element on site 3 goes down column II in the periodic table, the formation energy is successively lowered, with Ca, Sr, and Ba contributing more than any halogen atom. On sites 1 and 2, the formation energy generally increases the most for heavy noble gases. On sites 3 and 4, it is carbon followed by neighboring B and N that increase the formation energy the most. The accuracy of such linear single atom energy models based on these scales, however, is not on par with the ML model, and—maybe more importantly—cannot be improved systematically through increasing training set sizes but rather converges to a finite residual error.

In order to achieve an accuracy of $\pm 0.1$ eV/atom for the elpasolites, a relatively large training set of $10 \times 10^3$ is needed. This is likely due to the sparsity of crystals at the opposite ends of the high and low formation energy spectrum; this results in a less predictive ML model for crystals in these regions, which is demonstrated in Fig S6 in Ref. [19]. Nevertheless, the $10 \times 10^3$ ML model readily identifies a larger set of lowest lying elpasolites for which the actual DFT minima can be obtained through subsequent DFT based screening. This is shown in Fig. 1(e) where the 250 crystals with the lowest ML predicted formation energies are shown in ascending order (with further details on these systems in the Supplemental Material [19]). Subsequent screening with DFT indicates the 26th crystal $CaSrCs_2F_6$ (out of $2 \times 10^6$) to be at the global formation energy minimum at $-3.44$ eV/atom, closely followed by a near-degenerate isomer $SrCaCs_2F_6$. The DFT energies of the next two degenerate pairs $CaSrRb_2F_6$ and $SrCaRb_2F_6$, and $CaBaCs_2F_6$ and $BaCaCs_2F_6$ correspond to $-3.41$, and $-3.39$ eV/atom, respectively. Overall, the elpasolites with the lowest formation energies, $ABC_2D_6$, correspond to $A$ and $B$ being late elements from group II, and $C$ and $D$ being a late element from group I and fluoride, respectively. Populating the four sites with elements from groups III, II, I, and VIII, respectively, differs from the experimentally established stoichiometry $AlNaK_2F_6$. In fact, the lowest DFT energy crystal with a group-III element is $CsAlRb_2F_6$ (in 69th position) with $-3.09$ eV/atom (ML energy: $-2.96$ eV/atom, see the Supplemental Material [19]).

We have also used our predictions to analyse the atomic oxidation states in elpasolites. In particular, we have found that roughly 6% of the crystals with formation energies below $-1$ eV/atom exhibit unusual atomic charges: they have a low formation energy despite the fact that no combination of conventional atomic charges would result in a neutral system. In order to identify these crystals, we have used the absolute value of the lowest possible total oxidation state (LPTOS) that could possibly be realized using the list of typical atomic oxidation states on display in Table III in Ref. [19]. The lowest lying crystals have a LPTOS of 0 ($-3$ to $-3.44$ eV/atom formation energies). However, already at $-3$ eV/atom crystals with a LPTOS of 2 or 1 start to occur. At formation energies of $\sim -1.25$ eV/atom and higher, the number of crystals with a nonzero LPTOS increases rapidly, with a LPTOS as high as 12. The corresponding crystal frequency distributions are shown in Fig. 1(e), along with the formulas for the mutually lowest lying crystals. Interestingly, the number of crystals with a zero LPTOS increases monotonically with formation energy, while for nonzero LPTOS crystals the distribution is oscillatory.

To demonstrate the usefulness of our ML model we have identified thermodynamically stable elpasolites. To this end, we first selected all those 274 213 elpasolites with negative ML formation energies, and without rare gas elements. Since stability depends on the energy difference to any possible polymorph or competing segregated phases [33,34], we have queried the available DFT formation energies stored in the materials project (MP) [28]. Some elpasolites, such as the archetypical $AlNaK_2F_6$, are already stored in the MP. Using the DFT results stored in the MP for competing quaternary, ternary, and binary phases, we have constructed phase diagrams [34] for all the 274 213 crystals. For each crystal, there are on average $\sim$12 competing phases stored in the MP; there is only one combination of elements (Cs, Li, Na, Rb) for which no binary or ternary competing phases have been stored. For 2133 out of the 274 213 crystals, the resulting stabilization energy is below the known convex hull of stability (for more details, see Ref. [19]). Subsequent validation using DFT instead of ML confirms 128 out of them to be stable. Out of these 38 are polymorphs ($ABC_2D_6$ versus $BAC_2D_6$), resulting in 90 unique stoichiometries. Such a reduction ($2133 \rightarrow 128$) in the number of crystal candidates is to be expected since sorting the crystals by ML energies being lower than the convex hull systematically favors those with negative ML formation energy errors. We note that this does not amount to proof that the 90 crystals are stable: the MP database is not exhaustive. This implies that other new competing phases and materials, with even stronger stabilization, might still be discovered in the future. Also, the intrinsic error of the employed DFT method within the MP might still alter the outcome with respect to experiment. As such, the 90 new elpasolite DFT energies represent new upper bounds on the convex hull at the corresponding compositions. They have been submitted to the MP database, and most of them have been made available for further studies (see Table V in Ref. [19] for the list of the 90 structures).

Among these elpasolites, metals, semiconductors, and insulators are roughly distributed equally. All structures with an earth alkaline metal in crystal position 4 have a low or zero band gap. We have noted an intriguing yet stable metal, $NFAl_2Ca_6$ (MP ID: mp-989399, No. 20 in Table V in Ref. [19]) with Ca at position 4, instead of F or Cl. The Bader charge analysis [35–37] (Table I) indicates an exotic

TABLE I. Calculated atomic charges in $NFAl_2Ca_6$ elpasolite using different methods (obtained using SIESTA[43]).

| Method | N | F | Al | Ca |
|---|---|---|---|---|
| Bader | −2.00 | −0.98 | −2.13 | 1.20 |
| Hirshfeld | −0.63 | −0.36 | −1.05 | 0.52 |
| Voronoi deformation density | −0.81 | −0.29 | −1.13 | 0.56 |

negative oxidation state for Al (-II), previously only reported for Al in substantially larger Zintl phase unit cells ($Sr_{14}[Al_4]_2Ge_3$) [38]. Since Bader charges sometimes yield nonintuitive results [39,40], we calculated Hirshfeld [41] and Voronoi deformation density [39,42] charges (Table I) which confirm the negative oxidation state, albeit reduced by one unit (-I). The calculated phonon spectra of $NFAl_2Ca_6$ also indicate stability [19].

In conclusion, we have developed and used ML models of formation energies to investigate all possible elpasolites made up of main-group elements. We have presented numerical results for $\sim 2 \times 10^6$ formation energies. The ML model is only implicitly dependent on spatial coordinates, through the reference data used for training. No spatial coordinates are needed for new queries, yet for a training set of $10 \times 10^3$ crystals the model reaches $\pm 0.1$ eV/atom—comparable to the DFT accuracy for solids. The results have been used to identify the most strongly bound elpasolites as well as to investigate the energy and bonding trends at crystal structure sites, leading to a new "elpasolite order" of elements, consistent with the bonding physics in the elpasolite crystal structure. We identified and added 128 structures (90 unique stoichiometries) to the convex hull of the MP database. A charge analysis for the metallic elpasolite $NFAl_2Ca_6$ indicates a negative atomic oxidation state of Al. This outcome directly demonstrates that ML methods can be used for the discovery of stable as well as unconventional chemistries. Because of the low computational cost of the ML model one can now also afford to remove human bias by considering also those structures that previously would have been excluded due to "chemical intuition." Our results suggest that ML models hold great promise for the computational screening of polymorphs, other crystal structure symmetries, solid mixtures, phase transitions, or defects at an unprecedented rate and extent. Other crystal properties than energies could also be considered.

The authors thank G. Hart, R. Ramakrishnan, and R. Ramprasad for comments, A. Jain, D. Winston, P. Huck, and K. A. Persson for helpful input on the stability calculations and for validating elpasolites for inclusion in the MP, and B. Huang and R. Sarmiento-Pérez for help with further validation of the stability and the charges of $NFAl_2Ca_6$. O. A. v. L. acknowledges funding from the Swiss National Science Foundation (Grant No. PP00P2_138932). This material is based upon work supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF under Grant No. FA9550-15-1-0026. This research was partly supported by the NCCR MARVEL, funded by the Swiss National Science Foundation. This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under Project No. mr14. R. A. acknowledges funding by the Swedish Research Council Grant No. 621-2011-4249, a Linnaeus Environment grant (LiLi-NFM) and the Swedish e-Science Research Centre (SeRC). Calculations have been performed at the Swedish National Infrastructure for Computing (SNIC).

*anatole.vonlilienfeld@unibas.ch
†rickard.armiento@liu.se

[1] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, Acta Crystallogr., Sect. B: Struct. Sci. **58**, 364 (2002).
[2] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, J. Chem. Inf. Comput. Sci. **23**, 66 (1983).
[3] P. Yang, F. P. Doty, M. A. Rodriguez, M. R. Sanchez, X. Zhou, and K. S. Shah, in Symposium L—Nuclear Radiation Detection Materials—2009, MRS Online Proc. Libr. **1164-L11-05** (2009).
[4] K. Biswas and M.-H. Du, Phys. Rev. B **86**, 014102 (2012).
[5] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[6] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[7] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[8] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, New J. Phys. **15**, 095003 (2013).
[9] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, Phys. Rev. B **89**, 205118 (2014).
[10] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, Phys. Rev. B **89**, 094104 (2014).
[11] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, Int. J. Quantum Chem. **115**, 1094 (2015).
[12] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, IEEE Trans. Neural Networks **12**, 181 (2001).
[13] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT, Cambridge, MA, 2002).
[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2011).
[15] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, J. Chem. Theory Comput. **11**, 2087 (2015).
[16] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Sci. Rep. **3**, 2810 (2013).
[17] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. **114**, 105503 (2015).
[18] T. D. Huan, A. Mannodi-Kanakkithodi, and R. Ramprasad, Phys. Rev. B **92**, 014106 (2015).
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.135502, which includes Refs. [18–27], for details on computations, additional

analysis of the ML model output, and the ML model source code.
[20] R. Armiento *et al.*, The High-Throughput Toolkit (httk), http://httk.openmaterialsdb.se/.
[21] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[22] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[23] G. Kresse and J. Furtmüller, *Vienna Ab Initio Simulation Package, Users Guide* (The University of Vienna, Vienna, 2007).
[24] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[25] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[26] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, Crit. Rev. Solid State Mater. Sci. **39**, 1 (2014).
[27] O. Shyue Ping, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Comput. Mater. Sci. **68**, 314 (2013).
[28] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).
[29] S. Lany, Phys. Rev. B **78**, 245207 (2008).
[30] G. Hautier, S. Ong, A. Jain, C. Moore, and G. Ceder, Phys. Rev. B **85**, 155208 (2012).
[31] A. E. Mattsson, R. Armiento, J. Paier, G. Kresse, J. M. Wills, and T. R. Mattsson, J. Chem. Phys. **128**, 084714 (2008).
[32] D. Pettifor, *Bonding and Structure of Molecules and Solids* (Oxford University Press, New York, 2002).
[33] A. R. Akbarzadeh, V. Ozoli, and C. Wolverton, Adv. Mater. **19**, 3233 (2007).
[34] S. P. Ong, L. Wang, B. Kang, and G. Ceder, Chem. Mater. **20**, 1798 (2008).
[35] W. Tang, E. Sanville, and G. Henkelman, J. Phys. Condens. Matter **21**, 084204 (2009).
[36] E. Sanville, S. D. Kenny, R. Smith, and G. Henkelman, J. Comput. Chem. **28**, 899 (2007).
[37] G. Henkelman, A. Arnaldsson, and H. Jònsson, Comput. Mater. Sci. **36**, 354 (2006).
[38] W. Marco and C. Röhr, Z. Naturforsch. B **62**, 1227 (2014).
[39] C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends, and F. M. Bickelhaupt, J. Comput. Chem. **25**, 189 (2004).
[40] R. Sarmiento-Prez, T. F. Cerqueira, I. Valencia-Jaime, M. Amsler, S. Goedecker, S. Botti, M. A. Marques, and A. H. Romero, New J. Phys. **15**, 115007 (2013).
[41] F. L. Hirshfeld, Theor. Chim. Acta **44**, 129 (1977).
[42] F. M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, and E. J. Baerends, Organometallics **15**, 2923 (1996).
[43] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys. Condens. Matter **14**, 2745 (2002).