# Finding Unprecedentedly Low-Thermal-Conductivity Half-Heusler Semiconductors via High-Throughput Materials Modeling

Jesús Carrete, Wu Li, and Natalio Mingo[*]

*CEA-Grenoble, 17 Rue des Martyrs, Grenoble 38054, France*

Shidong Wang, and Stefano Curtarolo[†]

*Center for Materials Genomics, Materials Science, Electrical Engineering, Physics and Chemistry,*
*Duke University, Durham, North Carolina 27708, USA*

The lattice thermal conductivity ($\kappa_\omega$) is a key property for many potential applications of compounds. Discovery of materials with very low or high $\kappa_\omega$ remains an experimental challenge due to high costs and time-consuming synthesis procedures. High-throughput computational prescreening is a valuable approach for significantly reducing the set of candidate compounds. In this article, we introduce efficient methods for reliably estimating the bulk $\kappa_\omega$ for a large number of compounds. The algorithms are based on a combination of machine-learning algorithms, physical insights, and automatic *ab initio* calculations. We scanned approximately 79,000 half-Heusler entries in the AFLOWLIB.org database. Among the 450 mechanically stable ordered semiconductors identified, we find that $\kappa_\omega$ spans more than 2 orders of magnitude—a much larger range than that previously thought. $\kappa_\omega$ is lowest for compounds whose elements in equivalent positions have large atomic radii. We then perform a thorough screening of thermodynamical stability that allows us to reduce the list to 75 systems. We then provide a quantitative estimate of $\kappa_\omega$ for this selected range of systems. Three semiconductors having $\kappa_\omega < 5 \text{ Wm}^{-1}\text{ K}^{-1}$ are proposed for further experimental study.

Subject Areas: Computational Physics, Condensed Matter Physics,
Materials Science

## I. INTRODUCTION

High-throughput (HT) computational materials science is a rapidly expanding area of materials research. It merges a plethora of techniques from a variety of disciplines. These include the kinetics and thermodynamics of materials, solid-state physics, artificial intelligence, computer science, and statistics [1]. The application of HT has recently led to new insights and novel compounds in different fields [2–9]. Despite the importance of thermal transport properties for many crucial technologies, there are to date no high-throughput investigations into lattice thermal conductivity.

Here, we seek to address this challenge. We concentrate on the lattice thermal conductivity of half-Heusler (HH) compounds, as they have great promise for applications as thermoelectric materials [10–13]. Half-Heusler compounds are ternary solids. Their crystalline structure consists of two atoms (*A* and *B*), located in equivalent positions in a rock-salt structure. A third atom (*X*) sits in an inequivalent

[*]natalio.mingo@cea.fr
[†]stefano@duke.edu

position, filling half of the octahedrally coordinated sites [Fig. 1a].

Experimental studies have reported the thermoelectric figure of merit for a small set of these systems and their alloys [14–18]. Theoretical electronic characterizations have been performed for 36 candidates [19]. It has been speculated that their high thermal conductivity, close to $10 \text{ Wm}^{-1}\text{ K}^{-1}$, could limit thermoelectric performance [20,21]. At room temperature, the lattice thermal conductivity $\kappa_\omega$ represents the largest contribution to the total conductivity.

Promising thermoelectric figures of merit have been reported both for *n*-type (1.5 at 700 K [22]) and for *p*-type (0.8 at 1000 K [17]) half-Heuslers. Such values are comparable to the best thermoelectric materials proposed thus far [23]. Those values, however, were not found in ordered half-Heuslers but rather in alloyed or nanostructured systems. Furthermore, finding ordered compounds with very low $\kappa_\omega$ is advantageous, as their electronic mobilities are expected to be higher than in alloys. In addition, alloying the already low-$\kappa_\omega$ ordered compounds would lower $\kappa_\omega$ even further.

The pool of candidate compounds analyzed in this article is larger than in previous investigations. All possible half-Heusler compounds from all combinations of nonradioactive elements in the periodic table are considered, as

FIG. 1 (color online).    (a) Prototype Half-Heusler structure with primitive vectors and a conventional cell. (b) Elements considered in this study.

included in the AFLOWLIB.org consortium repository [24] [Fig. 1b]. The formation enthalpies of the fully relaxed structures are obtained through density functional theory within the AFLOW high-throughput framework [25].

From a total of 79,057 entries, those with positive formation enthalpies are removed. When several half-Heuslers are related by permutations of elements, only the lowest-enthalpy configurations are considered. Finally, zero-gap compounds are removed from the list. For the surviving subset of 995 compounds, the second-order force constants are characterized with full phonon dispersion curves. This allows further reduction of the set to a total of 450 mechanically stable semiconductors. Although these requirements do not guarantee global thermodynamical stability, metastable compounds with long lifetimes have been synthesized and used [26]. Hence, their inclusion should not be disregarded *a priori*.

For the 450 resulting stable half-Heuslers, we compute a large set of structural, electronic, and harmonic properties. In principle, one could directly compute $\kappa_\omega$ for all the compounds. The computational requirements for this approach would be prohibitive. To solve this issue, our strategy is to obtain $\kappa_\omega$ for a smaller subset of systems. We use physical insights and machine learning techniques to predict the remaining values. Cross validation shows that the approach is reliable for rapidly identifying low-$\kappa_\omega$ compounds.

Once the main factors correlated with a low thermal conductivity are identified for the 450-HH library, we use the thermodynamical information in the AFLOWLIB.org database to test the stability of these HHs against more than 110,000 phases. All competing ternary compounds from the Inorganic Crystal Structure Database (ICSD) [27] and all binaries in that database sharing two elements with each HH are included. The final list of thermodynamically stable compounds contains 75 entries. For these, we devise and implement a novel approach to compute the lattice thermal conductivity. Our accuracy is better than 50% of the exact calculation and has a much lower computational cost. This allows us to provide estimates of $\kappa_\omega$ that can be compared with experiment for 75 thermodynamically stable compounds.

## II. PREDICTING BULK LATTICE THERMAL CONDUCTIVITIES

The general expression for $\kappa_\omega$ at temperature $T$ is [28]

$$\kappa_\omega = \frac{1}{k_B T^2 V} \sum_\lambda n_0(n_0 + 1)|v_\lambda^{(z)}|^2 \hbar^2 \omega_\lambda^2 \tau_\lambda, \qquad (1)$$

where $\lambda$ denotes the phonon branch index $\alpha$ and wave vector $\mathbf{q}$, $k_B$ is Boltzmann's constant, $n_0$ is the Bose-Einstein distribution, $\omega_\lambda$ is the frequency of phonons, $v_\lambda^{(z)}$ is the phonon group velocity in the transport direction $z$, and $\tau_\lambda$ is the relaxation time. The relaxation time is determined by third-order derivatives of the total energy with respect to the atomic displacements of any three atoms $i$, $j$, and $k$ in directions $a$, $b$, and $c$ ($\Phi_{ijk}^{abc}$, the anharmonic force constants) in a large supercell [29].

In ordered half-Heuslers, the dominant source of scattering is due to three-phonon processes, and we can calculate thermal conductivities with the full *ab initio* anharmonic characterization [30,31]. For CoSbZr, one of the most thoroughly studied half-Heuslers, we obtain $\kappa_\omega = 25.0$ Wm$^{-1}$K$^{-1}$. The value is very close to a previous theoretical estimate ($\sim$22 Wm$^{-1}$K$^{-1}$) for monocrystalline CoSbZr [19] and slightly higher than the experimental values [32,33] (between 15 and 20 Wm$^{-1}$K$^{-1}$). Synthesized samples of Refs. [32,33] may contain microstructures and imperfections not considered in our work. Despite their accuracy, "full *ab initio* calculations" of $\kappa_\omega$ [Eq. (1)] are prohibitive for HT studies because of the computational requirements of the derivatives giving $\tau_\lambda$.

In this section, we present two different approaches circumventing the limitation. The first method is based on the empirical observation that the force constants show a high degree of transferability between compounds sharing crystal structure [34]. This suggests that a single set of anharmonic force constants could be used to get an estimate of the bulk $\kappa_\omega$. We call this thermal conductivity calculated with "transferred" forces $\kappa_{transf}$ (see Table II).

We want to preserve the choice between equivalent positions for maximizing transferability. Thus, instead of

taking the anharmonic force constants of a particular half-Heusler compound, we choose those of $Mg_2Si$. This compound shares the half-Heusler lattice with sites $A$ and $B$ occupied by Mg atoms. For cross validation, we also fully compute the anharmonic force constants of 32 half-Heusler systems. These are randomly selected with uniform probability inside the convex hull of Fig. 2a, to ensure a wide variety of harmonic and anharmonic features. Comparison between $\kappa_\omega$ and $\kappa_{transf}$ indicates that, although the latter has limited quantitative precision, the qualitative agreement is very good, with a Spearman rank correlation coefficient of 0.93. Hence, the descriptor can be effectively used to separate compounds having high or low $\kappa_\omega$. Note that we chose the Spearman rank correlation [35] instead of the usual Pearson one. The former is invariant under any



**(a)**

**(b)**

**(c)**

FIG. 2 (color online). (a) Joint scatter plot of $c_v$ at 300 K and $P_3$, colored according to our low- and high-$\kappa_\omega$ classification based on $\kappa_{forest}$ (see text); the convex hull of the point set is also included for guidance. (b) Frequency densities of the estimators of thermal conductivity at 300 K $\kappa_{transf}$ and $\kappa_{forest}$ as defined in the text. (c) "Violin plot" showing the distribution of $a_{latt}$ within the low- and high-$\kappa_\omega$ classes.

monotonic transformation of one or both variables and takes values $\pm 1$ for any strict monotonic (not just linear) dependence.

The second proposed approach is based on a completely different direction: We use "random-forest regression" by leveraging the 32 fully calculated $\kappa_\omega$ as a training set. We can then employ the fitted model to predict the remaining conductivities. We call these predictions $\kappa_{forest}$ (see Table II). Random forests [36] are a family of general classification and regression algorithms and are well adapted to dependent input data. They have already been successfully applied to numerous problems [37,38], including compound classification [39]. Here, the 32 compounds represent only around 7% of the mechanically stable half-Heuslers. Our input data comprise a large set of descriptor variables, which are expected to correlate with $\kappa_\omega$ (Supplemental Material [40]) but is less expensive to obtain. Descriptors include the following:

 (i) *A priori* chemical information: atomic number and weight, position in the periodic table, atomic radius, Pauling electronegativity [41], and Pettifor's chemical scale $\chi$ (Ref. [42]).
 (ii) General compound information: lattice constant $a_{latt}$, band gap, formation enthalpy, effective masses of electrons and holes, Born effective charges, and dielectric tensor.
 (iii) Specific thermal conductivity information: specific heat $c_v$, spherically averaged speed of sound $c_s$, scaled nanograined-limit thermal conductivity $\tilde{\kappa}_{grain}$, and phase-space volume available for three-phonon scattering processes $P_3$.

After an exploratory phase, we conclude that a satisfactory fit can be safely achieved using only *a priori* data.

The random-forest method is performed in three steps. First, a large ensemble of decision trees is built by randomly selecting subsets of descriptors and observations. Second, the predictions of all trees are obtained for each data point. Third, the mode (for classification) or the mean (for regression) are taken as the result from the whole ensemble. The algorithm also provides an intrinsic metric to evaluate the importance of each descriptor. This is defined in relation to the effect of randomly permuting the values of that variable on the result [36] (the less resilient upon permutation, the more important).

The prediction of each tree in a random forest can only be a value from the training set, and thus the result of the regression is a weighted average. This average is bounded by the minimum and maximum values within the training data. A small set is unlikely to contain elements having extreme values. Hence, our random-forest regression is expected to have a marked centralizing effect, yielding values tightly grouped around their mean. The frequency densities of both $\kappa_{transf}$ and this new $\kappa_{forest}$ are displayed in Fig. 2b. The latter avoids extreme predictions with non-physical magnitudes, a result of the aforementioned centralizing effect.

TABLE I. Fully calculated thermal conductivities $\kappa_\omega$ for 32 compounds. These results are then used as the training set for the random-forest predictions. An estimate of the relative standard deviation of $\kappa_{\text{forest}}$ for each compound in the training set, as obtained using repeated fourfold cross validation, is also included. Compounds are always labeled with the element in position $X$ first.

| | $\kappa_\omega$ (Wm$^{-1}$ K$^{-1}$) | $\sigma_{\text{forest,CV}}$ (%) | | $\kappa_\omega$ (Wm$^{-1}$ K$^{-1}$) | $\sigma_{\text{forest,CV}}$ (%) | | $\kappa_\omega$ (Wm$^{-1}$ K$^{-1}$) | $\sigma_{\text{forest,CV}}$ (%) |
|---|---|---|---|---|---|---|---|---|
| AgKTe | 0.508 | 42 | GeCaZn | 2.75 | 9.6 | PtGaTa | 32.9 | 11 |
| BeNaP | 4.08 | 20 | GeNaY | 8.06 | 14 | PtGeTi | 16.9 | 9.0 |
| BiBaK | 2.19 | 11 | LiBaSr | 0.582 | 15 | PtLnNb | 16.5 | 8.2 |
| BiKSr | 1.96 | 6.4 | IrPTi | 27.4 | 7.8 | RhHfSb | 21.8 | 13 |
| BiLiSr | 3.04 | 10 | NiPbTi | 109 | 10 | RhNbSi | 15.3 | 11 |
| CoAsZr | 24.0 | 7.4 | NiSbSc | 19.5 | 11 | RuAsV | 23.5 | 13 |
| CoBiHf | 18.6 | 14 | NiSnTi | 17.9 | 9.3 | SbCaK | 2.70 | 9.3 |
| CoSbZr | 25.0 | 2.4 | NiSnZr | 19.6 | 11 | SiCdSr | 13.5 | 19 |
| CoScSe | 15.0 | 13 | OsSbTa | 29.6 | 12 | SnBaSr | 2.01 | 43 |
| CoSiTa | 37.8 | 7.7 | PdAsY | 5.48 | 9.5 | TeAgLi | 1.52 | 11 |
| FeNbP | 109 | 4.2 | PdSrTe | 1.16 | 19 | | | |

In this sense, machine-learning algorithms outperform crude extrapolations such as those behind $\kappa_{\text{transf}}$. Additionally, $\kappa_{\text{forest}}$ has the advantage that its predictions can be refined with controlled accuracy by changing the size of the training set. Even so, the Spearman rank correlation coefficient between $\kappa_{\text{transf}}$ and $\kappa_{\text{forest}}$ is still 0.66, corroborating the validity of the analysis based on $\kappa_{\text{forest}}$. Furthermore, we find that $\kappa_{\text{forest}}$ is strongly correlated with physical descriptors like $c_v$, $\tilde{\kappa}_{\text{grain}}$, and $P_3$. This confirms our earlier speculation about these methods.

An important concern when training a machine-learning model is whether the training set is diverse or representative enough to justify extrapolating the model to the remaining elements. The values of $\kappa_\omega$ needed for direct validation of the predicted $\kappa_{\text{forest}}$ are unavailable. Thus, we resort to a repeated fourfold cross validation among the data points in the training set to obtain an estimate of the out-of-sample error. More specifically, we evenly split our training set into four subsets. Then, we obtain a random-forest prediction for the HHs in each of the subsets by using only the remaining 75% of compounds as the new training set. We repeat the process 10 times for different divisions of the data and compute the standard deviation of these predictions. The results are included in Table. I. These estimates

TABLE II. Notation for thermal conductivities.

| Label | Definition |
|---|---|
| $\kappa_\omega$ | Lattice contribution to $\kappa$ from the "full calculation" |
| $\kappa_{\text{transf}}$ | Approximated $\kappa_\omega$ with anharmonic force constants from Mg$_2$Si |
| $\kappa_{\text{forest}}$ | $\kappa_\omega$ obtained random-forest regression |
| $\kappa_{\text{anh}}$ | $\kappa_\omega$ obtained with four exact anharmonic force constants and a linear model for the rest |
| $\kappa_e$ | Electronic contribution to $\kappa$ |
| $\tilde{\kappa}_{\text{grain}}$ | Scaled nanograined limit $\kappa_\omega$ |

support the notion that the model behind $\kappa_{\text{forest}}$ is reasonably insensitive to our choice of training sets. For each cross validation, we compute the Spearman rank correlation coefficient between the out-of-sample random-forest results for the 32 training compounds and their $\kappa_\omega$. The median value of these Spearman rank correlation coefficients is 0.74, corroborating $\kappa_{\text{forest}}$ as a reliable tool for predicting compound ordering.

The ordering predicted by descriptor $\kappa_{\text{forest}}$ is strongly correlated with that of $\kappa_\omega$. This allows us to pinpoint the main factors determining high or low thermal conductivities. The bimodal shape of the distribution in Fig. 2b suggests that two groups of half-Heuslers can be identified, with thermal conductivities spread around two different values. A robust version of the "$k$-means" algorithm [43] is employed to optimally place the medians of the low- and high-thermal-conductivity classes at 450 and 23.1 Wm$^{-1}$ K$^{-1}$, respectively. By analyzing the importance of variables in the classification, we identify a low Pettifor scale $\chi_X$ and a large average Pauling electronegativity $\bar{e}_{AB}$ as the most critical descriptors for low conductivity (Supplemental Material [40]).

Given the underlying correlations, many different choices can be used for the classification. A trend can even be suggested on the grounds of atomic radii by following a chain of correlations: If the two elements in equivalent positions are chosen so that their average radius is larger than 150 pm, then the probability of the compound being in the low-$\kappa_\omega$ class is 84%. Physically, this follows from the fact that $\kappa_\omega$ is highly correlated with the specific heat $c_v$ [Fig. 2a]. The latter is strongly negatively correlated with the lattice parameter $a_{\text{latt}}$: The larger $a_{\text{latt}}$ the lower $c_v$ [44].

In addition, $a_{\text{latt}}$ correlates well with the sum of the atomic radii of the three elements, quantities known *a priori*. The atomic radii of the species in positions $X$ concentrate around the average value. This leads to an accurate prediction of $a_{\text{latt}}$ by using only the average atomic

radius of atoms in positions $A$ and $B$, $\bar{r}_{AB}$. A large $\bar{r}_{AB}$ causes a large lattice constant, small specific heat, and finally, low thermal conductivity. Alternatively, the lattice parameter can be used as a good discriminant: Panel (c) in Fig. 2 is a "violin plot" illustrating the distribution of $a_{\text{latt}}$ in the classes of half-Heuslers with low and high thermal conductivities. Also, as it can be seen in Fig. 2a, our choice of easily computable descriptors such as $c_v$ and $P_3$ is supported by the result of this classification.

Our calculations are for the true bulk lattice thermal conductivity. They are unrelated to the minimum value proposed by other authors [45,46]. Nevertheless, some of the $\kappa_\omega$ obtained directly seem ultralow. They are even lower than $\sim 0.70$ Wm$^{-1}$ K$^{-1}$, as reported in the literature for AgSbTe$_2$ and AgBiSe$_2$ [47], and described as close to the achievable minimum. However, the minimum depends on the compound's structure. Even within the most stringent hypothesis of the shortest possible mean free path equal to interatomic spacing, the lowest found $\kappa_\omega$ is much higher than the theoretical minimum. Therefore, none of our predicted values violates the minimum lattice thermal conductivity. Note also that, once the goal of reducing the $\kappa_\omega$ under $\lesssim 1$ Wm$^{-1}$ K$^{-1}$ is achieved, its precise value loses relevancy as it is overtaken by the contribution of charge carriers, $\kappa_e$.

## III. SCREENING FOR THERMODYNAMICAL STABILITY

The ingredients of $\kappa_\omega$ for bulk ordered semiconductors depend only on a semilocal characterization of the potential energy surface around the equilibrium configuration. Hence, mechanical stability is sufficient to permit the calculation of the lattice thermal conductivity of a HH. For the analysis performed in the previous section, having the set of 450 mechanically stable HHs reduced and biased by external considerations such as thermodynamical stability would be detrimental to the performance of machine-learning techniques.

On the other hand, in order to propose particular candidates for experimentation, we must maximize the probability that they can be obtained in the laboratory. To this end, we obtain the ternary phase diagrams for each of the 450 mechanically stable HHs. This involves taking into account the formation enthalpies of a large number of possible competing phases. These include but are not limited to all relevant binary and ternary compounds in the ICSD [27]. More specifically, all the elemental compounds, 109,36 binary structures, and 4,363 ternary phases were considered. Many of these phases were already present in AFLOWLIB.org; others were computed specifically for this work. The total number of DFT calculations necessary to obtain the results presented here exceeds 300,000. Our thermodynamic analysis reveals that 77 of the 450 HHs are thermodynamically stable. Spin-polarized calculations reveal that two of the 77 have semimetallic
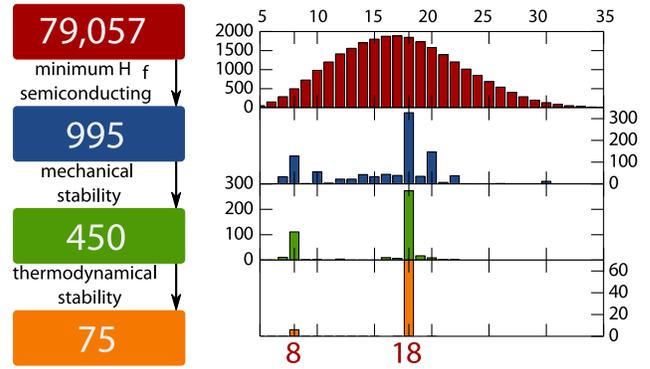


FIG. 3 (color online). Number of compounds during the screening (left panel) and evolution of the valence per unit cell distribution (right panel). All the final 75 compounds follow the 8/18 octet rule.

ground states. Then, only the remaining 75 compounds are further considered. The ternary phase diagrams of the final 75 systems are included in the Supplemental Material [40].

Interestingly, all of the 75 predicted stable compounds satisfy the octet or expanded octet rules by virtue of having 8 or 18 valence electrons per unit cell, respectively. We compare these numbers with the frequency distribution of valence electron counts in the initial 79,057-HH library. We conclude that the conditional probabilities of compounds having 8 or 18 valence electrons per unit cell being stable are 1.2% and 3.8%, respectively. While still small, the conditional likelihood of a compound satisfying one of these rules making it through all the filtering steps is much higher than the 0.1% a priori probability. Figure 3 shows the distribution of the valence during the reduction of the prototypes' list.

Even among the reduced list, $\kappa_{\text{forest}}$ still spans more than 1 order of magnitude, its extreme values being 2.33 and 40.3 Wm$^{-1}$ K, reinforcing our previous conclusions.

## IV. A DESCRIPTOR WITH QUANTITATIVE POWER

Neither of the two descriptors of $\kappa_\omega$ presented so far contains any information about the anharmonic interatomic force constants (IFCs) of each compound. On one hand, the last round of thermodynamical screening puts the number of surviving HHs within the limits of what can be realistically considered for anharmonic calculations. On the other, the qualitative success of $\kappa_{\text{transf}}$ shows that a detailed anharmonic description is not required. To enhance our estimates of the thermal conductivity of stable half-Heuslers, in this section we present a new machine-learning descriptor of $\kappa_\omega$ that integrates only the crucial pieces of the anharmonic properties of the solid. This aids in achieving quantitative accuracy with a much lower computational cost than the full calculation.
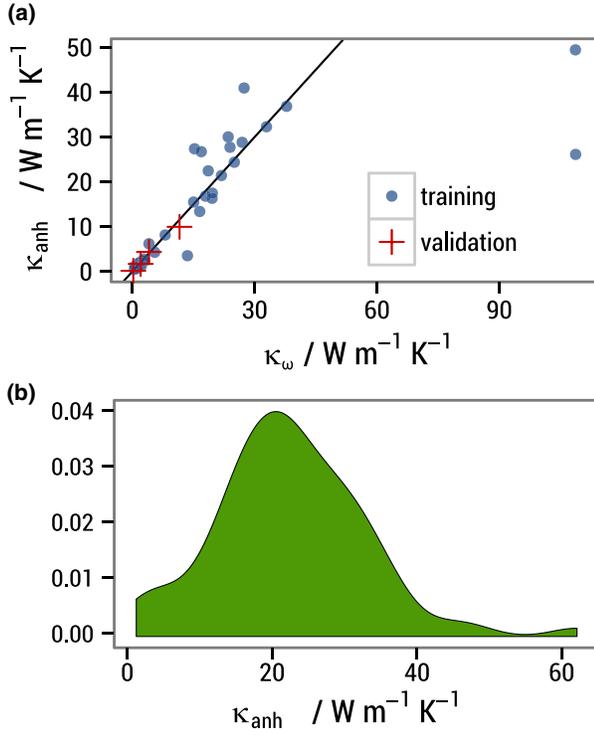
FIG. 4 (color online). (a) Comparison of $\kappa_{anh}$ with the exact $\kappa_{\omega}$ for the 32 compounds in the training set and the three compounds used for validation. (b) Distribution of $\kappa_{anh}$ over the 75 thermodynamically stable HHs.

Crystallographic symmetries and the equality of mixed partials impose linear constraints on the anharmonic IFCs. With the parameters described in the "Methods" section below and those constraints, we are left with 737

independent anharmonic IFCs per compound. However, many elements of this set are correlated among them, and others are too small to have a decisive role in the value of $\kappa_{\omega}$. To quantify these assertions, we perform a principal component analysis [48] on the third-order IFCs for the 32 compounds in Table I.

We find that the first four components account for $\sim 99\%$ of the variance in the set. From the results, we can extract an expression for each of the 737 IFCs as a linear combination of those components. Then we perform a multivariable multiple linear regression of the four components on four large and weakly correlated IFCs. By combining the two results, we arrive at a linear model for the whole set of anharmonic IFCs in terms of four parameters that can be obtained with 16 DFT calculations per compound. We use the term $\kappa_{\omega}$ to describe the third-order IFCs thus reconstructed, and $\kappa_{anh}$ for the second-order IFCs for each compound.

The blue circles in Fig. 4a show a comparison between $\kappa_{anh}$ and the exact $\kappa_{\omega}$ for the 32 compounds in the training set. With two exceptions (compounds with comparatively very high thermal conductivities), this new descriptor yields excellent quantitative estimates of $\kappa_{\omega}$. Moreover, fourfold cross validation shows that it is insensitive to the particular choice of training set. As a final test, we perform full thermal conductivity calculations for four compounds selected at random from those outside the training set: AgBaSb, AgNaTe, InCdY, and TlLaMg. The results are depicted as red crosses in Fig. 4a. This shows that the quality of the prediction is as good as for the 32 training compounds.

The distribution of $\kappa_{anh}$ over the 75 thermodynamically stable HHs [Fig. 4b] confirms the presence in the sample of

TABLE III. The values of $\kappa_{anh}$ for the 75 thermodynamically stable half-Heusler compounds.

| | $\kappa_{anh}$ (Wm$^{-1}$ K$^{-1}$) | | $\kappa_{anh}$ (Wm$^{-1}$ K$^{-1}$) | | $\kappa_{anh}$ (Wm$^{-1}$ K$^{-1}$) | | $\kappa_{anh}$ (Wm$^{-1}$ K$^{-1}$) |
|---|---|---|---|---|---|---|---|
| AuAlHf | 16.7 | FeAsNb | 47.6 | NiGaNb | 22.9 | RhAsZr | 27.1 |
| BLiSi | 62.1 | FeAsTa | 32.9 | NiGeHf | 19.6 | RhBiHf | 12.8 |
| BiBaK | 1.24 | FeGeW | 32.8 | NiGeTi | 25.3 | RhBiTi | 13.0 |
| CoAsHf | 20.0 | FeNbSb | 29.1 | NiGeZr | 21.1 | RhBiZr | 13.0 |
| CoAsTi | 37.1 | FeSbTa | 31.2 | NiHfSn | 19.5 | RhLaTe | 2.84 |
| CoAsZr | 27.7 | FeSbV | 24.1 | NiPbZr | 15.2 | RhNbSn | 15.7 |
| CoBiHf | 22.5 | FeTeTi | 26.2 | NiSnTi | 16.8 | RhSnTa | 20.3 |
| CoBiTi | 27.1 | GeAlLi | 16.5 | NiSnZr | 17.5 | RuAsNb | 43.7 |
| CoBiZr | 17.8 | IrAsTi | 30.1 | OsNbSb | 24.8 | RuAsTa | 33.4 |
| CoGeNb | 36.2 | IrAsZr | 17.4 | OsSbTa | 28.8 | RuNbSb | 22.7 |
| CoGeTa | 27.2 | IrBiZr | 12.8 | PCdNa | 6.05 | RuSbTa | 20.9 |
| CoGeV | 29.1 | IrGeNb | 33.0 | PdBiSc | 9.95 | RuTeZr | 21.3 |
| CoHfSb | 21.9 | IrGeTa | 37.2 | PdGeZr | 18.2 | SbNaSr | 3.49 |
| CoNbSi | 30.1 | IrGeV | 30.0 | PdHfSn | 15.1 | SiAlLi | 20.9 |
| CoNbSn | 20.7 | IrHfSb | 24.7 | PdPbZr | 10.3 | ZnLiSb | 6.44 |
| CoSbTi | 23.3 | IrNbSn | 19.8 | PtGaTa | 32.3 | | |
| CoSbZr | 24.4 | IrSnTa | 22.1 | PtGeTi | 26.7 | | |
| CoSiTa | 36.9 | NiAsSc | 17.5 | PtGeZr | 15.9 | | |
| CoSnTa | 22.7 | NiBiSc | 14.3 | PtLaSb | 1.72 | | |
| CoSnV | 19.8 | NiBiY | 10.6 | RhAsTi | 33.1 | | |

compounds with thermal conductivities much lower than 10–20 $Wm^{-1}K^{-1}$. This is characteristic of experimentally measured HHs. The values of $\kappa_{anh}$ for the 75 stable HHs are listed in Table III. Notably, the subset of 10 thermodynamically stable half-Heuslers for which $\kappa_\omega$ was directly computed already contains BiBaK, with $\kappa_\omega = 2.20$ $Wm^{-1}K^{-1}$. Outside of the training sample, the lowest $\kappa_{anh}$ values are 1.72, 2.84, and 3.49 for PtLaSb, RhLaTe, and SbNaSr, respectively.

## V. CONCLUSIONS

In this article, we have presented three computational methods for estimating the bulk $\kappa_\omega$ of a large library of half-Heusler compounds. We surmount the formidable task of full *ab initio* characterization. We find that $\kappa_\omega$ is spread over more than 2 orders of magnitude over mechanically stable half-Heuslers. This is a much broader range than that suggested by limited experimental available data. By using a set of descriptors and random-forest regression, we have built and tested an effective classification model. We found that compounds are most likely to have low thermal conductivity if the average atomic radius of the atoms in structural positions $A$ and $B$ is large. This also correlates with large lattice parameters and low specific heat.

Extensive thermodynamical calculations allow one to remove compounds with more stable competing phases from the list. We employ our third method, with better quantitative accuracy and higher computational cost, to perform a finer analysis of the distribution of $\kappa_\omega$ over the reduced library. We conclude that ordered half-Heusler compounds with a $\kappa_\omega \lesssim 3$ $Wm^{-1}$ $K^{-1}$ value (a factor of 3 below the best scenarios for ordered compounds and comparable to alloyed systems) very likely exist. The results corroborate the competitiveness of machine-learning methods in accelerated material design [1].

## VI. METHODS

*AFLOWLIB library of half-Heusler systems.*—The 79,057 half-Heusler systems are calculated with the high-throughput framework AFLOW [4,25,49,50] based on *ab initio* calculations of the energies by the VASP software [51] with projector augmented wave (PAW) pseudopotentials [52], and Perdew, Burke, and Ernzerhof exchange-correlation functionals [53]. The AFLOWLIB energies are calculated at zero temperature and pressure, with spin polarization and without zero-point motion or lattice vibrations. All crystal structures are fully relaxed (cell volume and shape and the basis atom coordinates inside the cell). Numerical convergence to about 1 meV atom$^{-1}$ is ensured by a high-energy cutoff (30% higher than the highest-energy cutoff for the pseudopotentials of the components) and by the dense 6,000 k-points per reciprocal atom Monkhorst-Pack meshes [54].

*Interatomic force constants.*—$3 \times 3 \times 3$ supercells are used in second-order IFC calculations. The Phonopy [55] package is used to generate a minimal set of atomic displacements by harnessing the point and translational symmetries of the crystal structure, and custom software was developed in order to do the same in anharmonic IFC calculations. For those calculations, $4 \times 4 \times 4$ supercells are generated and a cutoff radius of $0.85a_{latt}$ is imposed on the interactions. The $2 \times 2 \times 2$ and $3 \times 3 \times 3$ Monkhorst-Pack k-point grids are employed, and spin polarization is excluded to improve speed.

*Solution of the Boltzmann transport equation.*—Our self-consistent iterative approach is described in detail in Ref. [30]. Both three-phonon processes and the natural isotopic distribution of each element are taken into account as sources of scattering. A Gaussian smearing scheme with adaptive breadth [31] is chosen for integrations in the Brillouin zone. When using anharmonic IFCs from $Mg_2Si$ to approximate $\kappa_\omega$ for all materials, the solution to the Boltzmann transport equation failed to converge for five compounds, which are consequently excluded from the associated analysis.

*Regression and classification.*—The $R$ statistical computing environment [56] is chosen for all statistical analyses. Random-forest models are used as implemented in the "randomForest" package [57]. As a check, all regressions and classifications are repeated using a generalized boosted tree algorithm [58]; in all cases, the results are found to be in good agreement with those afforded by random forests.

## ACKNOWLEDGMENTS

[1] S. Curtarolo, G. L. W. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, and O. Levy, *The High-Throughput Highway to Computational Materials Design*, Nat. Mater. **12**, 191 (2013).

[2] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, and J. K. Nørskov, *Computational High-Throughput Screening of Electrocatalytic Materials for Hydrogen Evolution*, Nat. Mater. **5**, 909 (2006).

[3] G. Ceder, G. Hauthier, A. Jain, and S. P. Ong, *Recharging Lithium Battery Research with First-Principles Methods*, MRS Bull. **36**, 185 (2011).

[4] S. Wang, Z. Wang, W. Setyawan, N. Mingo, and S. Curtarolo, *Assessing the Thermoelectric Properties of*

*Sintered Compounds via High-Throughput Ab-Initio Calculations*, Phys. Rev. X **1,** 021012 (2011).

[5] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, *Computational Screening of Perovskite Metal Oxides for Optimal Solar Light Capture*, Energy Environ. Sci. **5,** 5814 (2012).

[6] L. Yu and A. Zunger, *Identification of Potential Photovoltaic Absorbers Based on First-Principles Spectroscopic Screening of Materials*, Phys. Rev. Lett. **108,** 068701 (2012).

[7] K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, and S. Curtarolo, *A Search Model for Topological Insulators with High-Throughput Robustness Descriptors*, Nat. Mater. **11,** 614 (2012).

[8] G. Ceder and K. Persson, *How Supercomputers Will Yield a Golden Age of Materials Science* (Scientific American, New York, 2013).

[9] G. L. W. Hart, S. Curtarolo, T. B. Massalski, and O. Levy, *Comprehensive Search for New Phases and Compounds in Binary Alloy Systems Based on Platinum-Group Metals, Using a Computational First-Principles Approach*, Phys. Rev. X **3,** 041035 (2013).

[10] C. Uher, J. Yang, S. Hu, D. T. Morelli, and G. P. Meisner, *Transport Properties of Pure and Doped* $MNiSn$ ($M = Zr$, $Hf$), Phys. Rev. B **59,** 8615 (1999).

[11] H. Hohl, A. P. Ramirez, C. Goldmann, G. Ernst, B. Wölfing, and E. Bucher, *Efficient Dopants for ZrNiSn-Based Thermoelectric Materials*, J. Phys. Condens. Matter **11,** 1697 (1999).

[12] G. S. Nolas, J. Poon, and M. Kanatzidis, *Recent Developments in Bulk Thermoelectric Materials*, MRS Bull. **31,** 199 (2006).

[13] C. Yu, T.-J. Zhu, R.-Z. Shi, Y. Zhang, X.-B. Zhao, and J. He, *High-Performance Half-Heusler Thermoelectric Materials* $Hf_{1-x}Zr_xNiSn_{1-y}Sb_y$ *Prepared by Levitation Melting and Spark Plasma Sintering*, Acta Mater. **57,** 2757 (2009).

[14] T. Sekimoto, K. Kurosaki, H. Muta, and S. Yamanaka, *Thermoelectric Properties of* $(Ti, Zr, Hf)CoSb$ *Type Half-Heusler Compounds*, Mater. Trans., JIM **46,** 1481 (2005).

[15] S. R. Culp, S. J. Poon, N. Hickman, T. M. Tritt, and J. Blumm, *Effect of Substitutions on the Thermoelectric Figure of Merit of Half-Heusler Phases at* $800°C$, Appl. Phys. Lett. **88,** 042106 (2006).

[16] G. Joshi, X. Yan, H. Wang, W. Liu, G. Chen, and Z. F. Ren, *Enhancement in Thermoelectric Figure of Merit of an n-Type Half-Heusler Compound by the Nanocomposite Approach*, Adv. Energy Mater. **1,** 643 (2011).

[17] X. Yan, G. Joshi, W. Liu, Y. Lan, H. Wang, S. Lee, J. W. Simonson, S. J. Poon, T. M. Tritt, G. Chen, and Z. F. Ren, *Enhanced Thermoelectric Figure of Merit of p-Type Half-Heuslers*, Nano Lett. **11,** 556 (2011).

[18] X. Yan, W. Liu, H. Wang, S. Chen, J. Shiomi, K. Esfarjani, H. Wang, D. Wang, G. Chen, and Z. F. Ren, *Stronger Phonon Scattering by Larger Differences in Atomic Mass and Size in p-Type Half-Heuslers* $Hf_{1-x}Ti_xCoSb_{0.8}Sn_{0.2}$, Energy Environ. Sci. **5,** 7543 (2012).

[19] J. Shiomi, K. Esfarjani, and G. Chen, *Thermal Conductivity of Half-Heusler Compounds from First-Principles Calculations*, Phys. Rev. B **84,** 104302 (2011).

[20] F. Casper, T. Graf, S. Chadov, B. Balke, and C. Felser, *Half-Heusler Compounds: Novel Materials for Energy and Spintronic Applications*, Semicond. Sci. Technol. **27,** 063001 (2012).

[21] W. Xie, A. Weidenkaff, X. Tang, Q. Zhang, J. S. J. Poon, and T. M. Tritt, *Recent Advances in Nanostructured Thermoelectric Half-Heusler Compounds*, Nanomaterials **2,** 379 (2012).

[22] S. Sakurada and N. Shutoh, *Effect of Ti Substitution on the Thermoelectric Properties of* $(Zr, Hf)NiSn$ *Half-Heusler Compounds*, Appl. Phys. Lett. **86,** 082105 (2005).

[23] J.-F. Li, W.-S. Liu, L.-D. Zhao, and M. Zhou, *High-Performance Nanostructured Thermoelectric Materials*, NPG Asia Mater. **2,** 152 (2010).

[24] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno Nardelli, N. Mingo, and O. Levy, *AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations*, Comput. Mater. Sci. **58,** 227 (2012).

[25] S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, *AFLOW: An Automatic Framework for High-Throughput Materials Discovery*, Comput. Mater. Sci. **58,** 218 (2012).

[26] X. Zhang, L. Yu, A. Zakutayev, and A. Zunger, *Sorting Stable versus Unstable Hypothetical Compounds: The Case of Multi-Functional ABX Half-Heusler Filled Tetrahedral Structures*, Adv. Funct. Mater. **22,** 1425 (2012).

[27] F. Karlsruhe, *Inorganic Crystal Structure Database*, http://icsd.fiz-karlsruhe.de/icsd.

[28] J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Oxford University, New York, 2001).

[29] A. Ward, D. A. Broido, D. A. Stewart, and G. Deinzer, *Ab Initio Theory of the Lattice Thermal Conductivity in Diamond*, Phys. Rev. B **80,** 125203 (2009).

[30] D. A. Broido, M. Malorny, G. Birner, N. Mingo, and D. A. Stewart, *Intrinsic Lattice Thermal Conductivity of Semiconductors from First Principles*, Appl. Phys. Lett. **91,** 231922 (2007).

[31] W. Li, N. Mingo, L. Lindsay, D. A. Broido, D. A. Stewart, and N. A. Katcho, *Thermal Conductivity of Diamond Nanowires from First Principles*, Phys. Rev. B **85,** 195436 (2012).

[32] Y. Xia, S. Bhattacharya, V. Ponnambalam, A. L. Pope, S. J. Poon, and T. M. Tritt, *Thermoelectric Properties of Semimetallic* $(Zr, Hf)CoSb$ *Half-Heusler Phases*, J. Appl. Phys. **88,** 1952 (2000).

[33] T. Sekimoto, K. Kurosaki, H. Muta, and S. Yamanaka, *High-Thermoelectric Figure of Merit Realized in p-Type Half-Heusler Compounds:* $ZrCoSn_xSb_{1-x}$, Jpn. J. Appl. Phys. **46,** L673 (2007).

[34] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni, *Ab Initio Calculation of Phonon Dispersions in Semiconductors*, Phys. Rev. B **43,** 7231 (1991).

[35] E. L. Lehmann, *Nonparametrics: Statistical Methods Based on Ranks* (Springer, New York, 2006).

[36] L. Breiman, *Random Forests*, Mach. Learn. **45,** 5 (2001).

[37] D. S. Palmer, N. M. O'Boyle, R. C. Glen, and J. B. O. Mitchell, *Random Forest Models to Predict Aqueous Solubility*, J. Chem. Inf. Model. **47,** 150 (2007).

[38] L. Auret and C. Aldrich, *Unsupervised Process Fault Detection with Random Forests*, Ind. Eng. Chem. Res. **49,** 9184 (2010).

[39] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, *Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling*, J. Chem. Inf. Comput. Sci. **43,** 1947 (2003).

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.4.011019 for a full table of descriptors for the 450 half-Heuslers library, a correlogram between variables and a set of ternary phase diagrams for the proposed 75 thermodynamically stable compounds.

[41] L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Cornell University Press, New York, 1960), 3rd ed.

[42] D. G. Pettifor, *A Chemical Scale for Crystal-Structure Maps*, Solid State Commun. **51,** 31 (1984).

[43] S. P. Lloyd, *Least Square Quantization in PCM*, IEEE Trans. Inf. Theory **28,** 129 (1982).

[44] The decreasing trend can be understood by considering that the specific heat per atom at high temperatures relates to the number of degrees of freedom, through the equipartition theorem. Hence, a $c_v \propto a_{latt}^3$ dependence should be expected. In fact, the observed trend is sharper because of the differences in Debye temperature among the compounds.

[45] G. A. Slack, *The Thermal Conductivity of Nonmetallic Crystals*, in Solid State Physics Vol. 34, edited by H. Ehrenreich, FX. Seitz, and D. Turnbull (Academic, New York, 1979), p. 1.

[46] D. G. Cahill, S. K. Watson, and R. O. Pohl, *Lower Limit to the Thermal Conductivity of Disordered Crystals*, Phys. Rev. B **46,** 6131 (1992).

[47] D. T. Morelli, V. Jovovic, and J. P. Heremans, *Intrinsically Minimal Thermal Conductivity in Cubic I-V-VI$_2$ Semiconductors*, Phys. Rev. Lett. **101,** 035901 (2008).

[48] I. T. Jolliffe, *Principal Component Analysis* (Springer, New York, 2002).

[49] W. Setyawan and S. Curtarolo, *High-Throughput Electronic Band Structure Calculations: Challenges and Tools*, Comput. Mater. Sci. **49,** 299 (2010).

[50] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, and S. Curtarolo, *High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials*, ACS Comb. Sci. **13,** 382 (2011).

[51] G. Kresse and J. Furthmüller, *Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set*, Phys. Rev. B **54,** 11 169 (1996).

[52] P. E. Blöchl, *Projector Augmented-Wave Method*, Phys. Rev. B **50,** 17 953 (1994).

[53] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77,** 3865 (1996).

[54] H. J. Monkhorst and J. D. Pack, *Special Points for Brillouin-Zone Integrations*, Phys. Rev. B **13,** 5188 (1976).

[55] A. Togo, F. Oba, and I. Tanaka, *First-Principles Calculations of the Ferroelastic Transition between Rutile-Type and CaCl$_2$-Type SiO$_2$ at High Pressures*, Phys. Rev. B **78,** 134106 (2008).

[56] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, (Vienna, Austria 2013), http://www.R-project.org/.

[57] A. Liaw and M. Wiener, *Classification and Regression by randomForest*, R News **2,** 18 (2002).

[58] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55,** 119 (1997).