ARTICLE   OPEN

Check for updates

Computational screening of materials with extreme gap deformation potentials

Pedro Borlido[1,2], Jonathan Schmidt[3], Hai-Chen Wang[3], Silvana Botti[1✉] and Miguel A. L. Marques[3✉]

In this work, we present a large-scale study of gap deformation potentials based on density-functional theory calculations for over 5000 semiconductors. As expected, in most cases the band gap decreases for increasing volume with deformation potentials that can reach values of almost −15 eV. We find, however, also a sizeable number of materials with *positive* deformation potentials. Notorious members of this group are halide perovskites, known for their applications in photovoltaics. We then focus on understanding the physical reasons for so different values of the deformation potentials by investigating the correlations between this property and a large number of other material and compositional properties. We also train explainable machine learning models as well as graph convolutional networks to predict deformation potentials and establish simple rules to understand predicted values. Finally, we analyze in more detail a series of materials that have record positive and negative deformation potentials.

INTRODUCTION

In the beginning of the 1970s it was empirically found that the band-gap of cubic semiconductors with the NaCl structure is inversely proportional to the square of the lattice constant, i.e., that the band-gap should be proportional to $V^{-2/3}$[1,2], where V is the volume of the unit cell. This variation of the band gap with respect to the volume is commonly measured by the deformation potential

$$\Xi_{\text{gap}} = V\frac{\mathrm{d}E_{\text{gap}}}{\mathrm{d}V} = \frac{\mathrm{d}E_{\text{gap}}}{\mathrm{d}\ln(V)}. \quad (1)$$

One can equally define a response function with respect to the pressure P as

$$a_P = \frac{\mathrm{d}E_{\text{gap}}}{\mathrm{d}P} = -\frac{\Xi_{\text{gap}}}{K}, \quad (2)$$

that is trivially related to Ξ_{gap} by the bulk modulus $K = -\mathrm{d}P/\mathrm{d}\ln(V)$. Later works studied deformation potentials of several other semiconductors, e.g., for wurtzite AlN, GaN, and InN[3,4]. A more complete list of deformation potentials can be found in ref. [5] (including a large number of zincblende and wurtzite semiconductors) and in ref. [6].

Theoretical models have been developed to explain the chemical trend of the values of deformation potentials in binary semiconductors with diamond and zincblende crystal structures[7]. A simple tight-binding model considers the minimum of the conduction band (CBM) consisting of anti-bonding s states and the maximum of the valence band (VBM) consisting of p bonding states. In that case, one expects a blue shift of the CBM with decreasing volume due to (i) stronger repulsion for the anti-bonding states with decreasing bond length and (ii) larger kinetic energy. The second effect should be identical for different systems, however one expects that the magnitude of (i) decreases for materials with anions with deeper s-levels, like halogens. For the VBM, the kinetic effect is similar, but the p bonding states shift

to lower energies for shorter bonds. Therefore, due to the cancellation of these two effects, one expects a smaller absolute value of the valence term Ξ_{VBM} with respect to that of the conduction term Ξ_{CBM}[7], with the following definition of the valence and conduction deformation potentials

$$\Xi_{\text{CBM/VBM}} = V\frac{\mathrm{d}E_{\text{CBM/VBM}}}{\mathrm{d}V} = \frac{\mathrm{d}E_{\text{CBM/VBM}}}{\mathrm{d}\ln(V)}, \quad (3)$$

where $E_{\text{CBM/VBM}}$ are the energies of the conduction band minimum and the valence band maximum, respectively. Following this model, and using $\Xi_{\text{gap}} = \Xi_{\text{CBM}} - \Xi_{\text{VBM}}$, one can explain the experimentally observed trend of increasing gap with decreasing volume mentioned above.

Of course, one should keep in mind that this model was developed for a particular family of materials, and cannot be easily generalized to arbitrary semiconductors. In fact, the existence of exceptions has been already known for a long time. For example, in lead chalcogenides the $V^{-2/3}$ scaling of the band gap is not verified[8,9]: in fact the value of the band gap decreases going from PbS, to PbTe, and then to PbSe[10,11], even though the lattice constant of PbTe is much larger than the one of PbSe.

The large amount of studies on deformation potentials is not surprising considering the importantance of the variation (or the stability) of the band gap in heterostructures inside modern opto-electronic devices, such as LEDs and transistors. One of the most basic sources of strain comes from lattice mismatch at interfaces or from different thermal expansion coefficients. In such cases large deformation potentials can be an undesirable property. In fact, in an epitaxial configuration the different lattice parameter of the materials needs to be accommodated, introducing strain in the junction[3]. Strain, besides being a possible source of structural defects, can modify some fundamental properties of the materials such as the position of the valence-band maximum and conduction-band minimum, and consequently the band gap. This effect must be carefully considered in situations where gap tuning

[1]Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena and European Theoretical Spectroscopy Facility, Max-Wien-Platz 1, 07743 Jena, Germany. [2]CFisUC, Department of Physics, University of Coimbra, Rua Larga, 3004-516 Coimbra, Portugal. [3]Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, D-06099 Halle, Germany. ✉email: silvana.botti@uni-jena.de; miguel.marques@physik.uni-halle.de

is imperative, as in InGaN/GaN photovoltaic devices and LEDs[12]. For some applications a controlled mismatch might actually be desired, giving origin to the so-called strain engineering[13].

Strain engineering is attracting growing attention, particularly for 2D materials[14] that exhibit large deformation potentials and better tolerance to mechanical strain than their bulk counterparts. For example, in tunneling field effect transistors made of MoS_2/BN heterojunctions[15] quantum band-to-band tunneling replaces thermal injection, with tunneling currents depending directly on the band-gap size. As a consequence, the on–off switching is considerably faster than in conventional field effect transistors. Because the size of the band gap can be tuned by the application of strain, strain engineering is a promising way of adjusting the properties of such devices. Controlled strain can also induce indirect–direct band-gap transitions in semiconductors, for example in SiGe alloys[16]. If the material has a large gap deformation potential this would imply enabling, e.g., light emission by strain and changing in a controlled way the color of the emitted light. Previous investigations also showed that the gap variation with respect to strain plays a key role in acoustic carrier scattering (see ref. [17] and references therein), with important consequences, e.g., for thermoelectrics.

To the best of our knowledge, there is no systematic investigation of gap deformation potentials for a large class of crystal structures and a variety of chemical compositions. As such, at the beginning of our investigation we did not know what makes a material 'normal', nor we knew what are the 'exceptional' materials that exhibit extreme values of deformation potentials. We decided therefore to answer this open question by performing calculations of gap deformation potentials for ~5300 semiconductors and insulators. The obtained dataset is then used for machine learning regression of deformation potentials and to understand which physical and chemical characteristics are correlated with extreme values of this property.

The remainder of this paper is organized as follows. In section "Results and discussions" we discuss the calculated deformation potentials. First we try to understand the data by analyzing the correlations between the deformation potentials and other properties of the systems. We then develop machine learning models capable of interpreting and predicting the data and we analyze in more detail some specific materials that we found to have exceptional values for the deformation potentials, followed by a short conclusion. In section "Methods" we present the details of the calculations involved, along with a description of the dataset used.

RESULTS AND DISCUSSIONS
Volume deformation potentials
The variation of the band gap with the volume depends on two factors: the variation of the cell volume and the change in the internal parameters (u_i, including both the position of the atoms in the unit cell, the cell angles and ratio between the cell parameters), i.e.,

$$\frac{dE_{gap}}{d \ln(V)} = \frac{\partial E_{gap}}{\partial \ln(V)} + \sum_i \frac{\partial E_{gap}}{\partial u_i} \frac{\partial u_i}{\partial \ln(V)}. \tag{4}$$

The first term is rather straightforward to calculate, and is therefore well suited for high-throughput studies. The second term requires geometry optimizations of both the cell parameters and positions of the ions at constant volume (or pressure) and is much more involved to calculate. We also know that, for group IV, III–V, and II–VI semiconductors, the effect of the second term is about 2% of the first term[7], meaning that it can (to some extent) be disregarded. However, this behavior is not general, and we can expect that for some systems the second term gives a more noticeable contribution to the deformation potentials. As it will be

discussed later, the size of the error is relevant for small to medium deformation potentials but it does not change the qualitative description of the largest deformation potentials in our set.

With this in mind, we decided to use a two step approach. First, we screen all systems using only the first term of Eq. (5) as an *estimator* of the deformation potential

$$\widetilde{\Xi}_{gap} = \frac{\partial E_{gap}}{\partial \ln(V)}. \tag{5}$$

We calculate this value using a two-point finite difference approximation to the derivative, obtained by straining the material isotropically and keeping the internal coordinates of the atoms unchanged. To strain the materials we compress lattice vector by 1%, i.e., after shrinking the new lattice vectors are simply $\mathbf{a}_i' = \mathbf{a}_i(1 - \gamma)$, with $\gamma = 0.01$. By construction this is equivalent to a reduction in the volume of approximately $3\gamma = 0.03$ while keeping the cell shape (angles and relative sizes of lattice vectors) constant. Of course this approach does not take into consideration, at least at this stage, the possible anisotropic response of the crystal, that can be quite relevant to its physics. Therefore, for all interesting compounds we validate the results by performing a full calculation of Ξ_{gap}. In the following, we will see that this procedure just leads to a couple of false positives in the list of 'exceptional' materials.

Distribution of values
In Figs. 1 and 2 we present distributions of estimators of the deformation potential (in gray, as given by Eq. (5)), and in Table 1 we show the statistical measures that characterize them. As we can see, the overall distribution is not centered at zero, but at -1.9 eV. Furthermore, the curve is highly skewed towards negative values of $\widetilde{\Xi}_{gap}$. Consequently, for the majority of materials the bandgap increases when they are compressed, which is the commonly expected behavior. We can also see that for negative deformation potentials, the curve has a fat tail, that extends up to -15 eV, while for positive values, the curve goes much faster to zero. From a visual inspection of the plot we can define 'normal' materials those crystalline compounds that have a deformation potential between around -5 and 3 eV, while we can define 'exceptional' materials those that exhibit values smaller than -10 eV or larger than 4 eV (this last approximately defined from the range of relative frequencies below 0.5%).
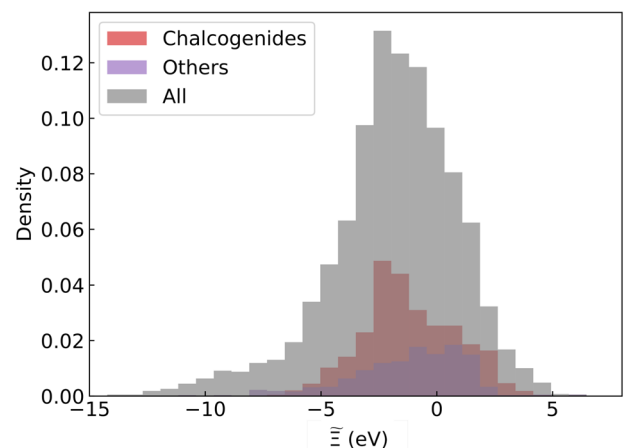

Fig. 1 Histogram of deformation potential estimator. Histograms of the estimators of the gap deformation potential ($\widetilde{\Xi}_{gap}$) for different chemical systems (All, Chalcogenides, and Others as defined in the main text). The values are calculated within the PBE approximation.
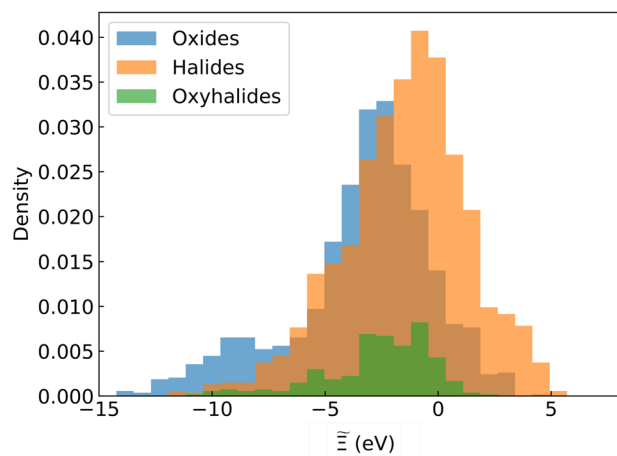
Fig. 2 Histogram of deformation potential estimator (cont.). Histograms of the estimators of the gap deformation potential ($\widetilde{\Xi}_{gap}$) for different chemical systems (Oxides, Halides and Oxyhalides as defined in the main text). The values are calculated within the PBE approximation.
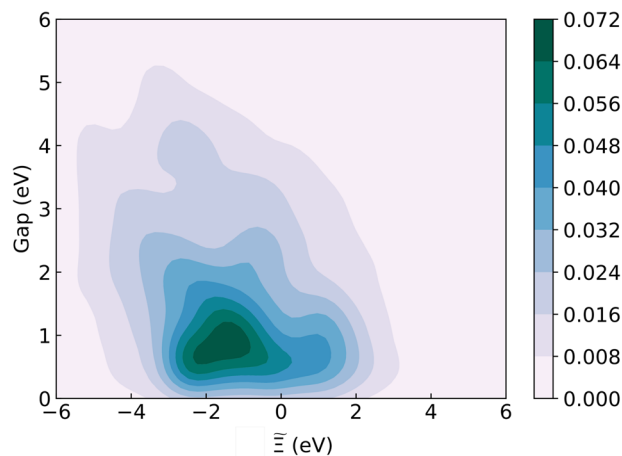

Fig. 3 $\widetilde{\Xi}_{gap}$ vs. gap heat map. Density plot of $\widetilde{\Xi}_{gap}$ with respect to the band gap. The values are calculated within the PBE approximation.

Table 1. Statistical analysis of distribution.

Set	All	Oxi.	Hal.	Oxyha.	Chalc.	Other
Size	5338	1277	1693	246	1452	670
Mean (eV)	−1.86	−3.36	−1.46	−2.68	−1.20	−1.17
σ (eV)	2.79	3.14	2.77	2.44	2.05	2.49
Skewness	−0.73	−0.77	−0.40	−1.04	−0.03	−0.76
Kurtosis	4.22	3.52	3.39	4.05	3.30	3.92
1st Q. (eV)	−3.26	−4.71	−3.00	−3.68	−2.49	−2.61
Median (eV)	−1.67	−2.84	−1.22	−2.17	−1.40	−0.75
3rd Q. (eV)	−0.05	−1.35	0.33	−0.83	0.30	0.74

Statistical measures of the estimator of the deformation potential for the full set (All), oxides (Oxi.), halides (Hal.), oxyhalides (Oxyhal.), chalcogens (Chalc.) and remaining (Other) materials of the set. 1st Q. and 3rd Q are the first and third quantiles of the distribution.

Table 2. Comparison to experimental data.

Material	ID	$\widetilde{\Xi}_{gap}$ (eV)	Exp. (eV)
GaP	mp-2490	−0.72 (−8.18)	– (−9.9)
GaAs	mp-2534	−7.30	−8.5
InP	mp-20351	−5.42	−6.4
ZnS	mp-10695	−4.37	−4.0
ZnSe	mp-1190	−3.98	−5.4
CdTe	mp-406	−2.97	−3.4
CdS	mp-2469	−2.23	−3.1
CdSe	mp-2691	−2.08	−3.0
AlN	mp-661	−8.98	−7.1/−9.5

Comparison between experimental values and estimated deformation potentials. All experimental values were obtained from refs. [5,6,75] and references therein. For GaP we show the deformation potential for the direct gap in parenthesis.

To make a preliminary analysis of how the chemistry influences the deformation potentials, we also plotted the distribution for subgroups of materials containing oxides, halides, oxyhalides, chalcogenides, and the remainder materials (see Figs. 1 and 2). We can see that the fat tail for negative values of $\widetilde{\Xi}_{gap}$ is mostly due to oxides, halides, and oxyhalides, with oxides yielding the largest contribution. On the other hand, materials with large positive values are almost exclusively halides. In any case, the maximum of the distribution seems to be rather insensitive to the chemistry, and we find in all subgroups materials with positive and negative values of the deformation potentials.

We plot in Fig. 3 a density plot of PBE band gaps as a function of the corresponding deformation potential. Most data points are concentrated in the region of band gaps smaller than around 2 eV with a slow decay (fat tail) for larger band gaps. From the plot, we can see that there exists only a weak correlation between the two quantities, so we do not expect the band gap to be a good indicator of the magnitude (or of the sign) of $\widetilde{\Xi}_{gap}$.

Finally, we would like to discuss the accuracy of our results in comparison to experimental values. Unfortunately, it is hard to find a large variety of experimental values for the hydrostatic deformation potentials in the literature. Most of the available values are for simple semiconductors and their alloys (which we do not consider here). The few materials for which we can

compare our estimator to reported experimental values are in Table 2. Since these compounds are mostly cubic with direct gaps, the estimator should match the real value of Ξ. We see that the PBE results are in very good agreement with experiment, with a tendency to overestimate the deformation potentials. We emphasize, however, that the set of materials in Table 2 is very small, and that this behavior may not hold for other more complicated compounds.

Correlations

We further investigate the correlation of the estimator of the gap deformation potential with other materials properties. This is important in view not only of the interpretation of the results, but also of the machine learning models that we develop in the next section. We considered possible correlations with the volume of the unit cell, the band gap, several elastic properties (bulk and shear moduli, Poisson ratios obtained from Materials Project database), the space group of the crystal structure, the net charges of the atoms, and the mean hole (m_h^*) and electron (m_e^*) effective masses. Concerning the calculations of charges, we applied both Bader[18,19] and DDEC6[20] analysis, while to get effective masses we interpolated the eigenvalues at regular k-points using Boltz-Trap2[21,22]. Furthermore, we considered a series of properties of the chemical elements that compose each compound using

Table 3. Correlation analysis.

Features	Spearman	Pearson
Effective m_h^*	−0.42	−0.01
Band gap	−0.32	−0.35
Maximum atomic number	−0.27	−0.25
Maximum DDEC6 charge	−0.25	−0.29
Shear modulus	−0.24	−0.20
Std DDEC6 charge	−0.21	−0.22
Std electronegativity	−0.20	−0.20
Mean DDEC6 charge	−0.20	−0.21
Bulk modulus	−0.19	−0.21
Mean row	0.41	0.41
Mean atomic mass	0.40	0.39
Volume of unit cell	0.34	0.33
Minimum melting point	0.30	0.26
Minimum row	0.28	0.30
Minimum atomic mass	0.26	0.27
Maximum atomic mass	0.26	0.25
Min thermal conductivity	0.23	0.10
Maximum row	0.22	0.21
Mean melting point	0.21	0.18

Spearman and Pearson correlation coefficients between deformation potentials $\tilde{\Xi}_{gap}$ and various materials properties. We show only features for which the absolute value of either coefficient is >0.2.

Table 4. Feature importance.

Feature	Importance
Mean row	538
Effective m_h^*	534
Volume of unit cell	346
Mean atomic mass	104
Maximum DDEC6 charge	52
Maximum Mendeleev number	28
Minimum melting point	21
Band gap	12
Minimum atomic mass	12
Minimum thermal conductivity	10
Minimum row	8
Std. DDEC6 charge	8

Mean of the feature importance of all features with a non-zero feature importance during at least one of the cross-validation runs.

Matminer[23]. We used for each property the maximum, minimum, average and standard deviation of the values of the constituent elements, as is common in machine learning applications[24].

In Table 3 we list the Spearman[25] and Pearson[25] correlation coefficients between the deformation potential $\tilde{\Xi}_{gap}$ and various materials properties. We only list properties for which at least one correlation coefficient is larger than 0.2. A full list can be found in the Supplementary Information.

The first thing to notice is that we do not find any strong correlation between the value of deformation potential and any other property. The largest value we find in Table 3 is the Spearman anticorrelation with the effective hole mass, yielding the value −0.42. This is most likely due to outliers reducing the Pearson correlation while leaving the Spearman correlation unchanged. Therefore, we can conclude that outliers, materials with exceptionally large (positive) deformation potentials have abnormally low hole effective masses. As we will see in the following, this happens because most of the materials with record positive deformation potentials are halide perovskites that have, indeed, untypically low hole masses[26]. We think that hole effective masses are important as their value depends strongly on the localization of bonding electrons, and therefore on the specific bond type and bond length, e.g. through changes of the hybridization of atomic states building the top valence. In this sense, hole effective masses are correlated with both bond lengths and band widths (and indirectly with band-gap sizes). In view of these considerations we can motivate that hole masses can be good features to predict deformation potentials.

Next we find correlations with the mean value of the row of the periodic table and of the atomic mass, that effectively describe the position in the periodic table of the constituent elements. These are generally known to be effective descriptors in composition-based machine learning models as they reveal a lot about the electronic structure of the elements. Only then do we arrive at the correlation with the unit cell volume and anti-correlation with the band gap. We also find that DDEC6 charges are anti-correlated

with the deformation potential, much more than Bader charges. This can be attributed to the fact that they attempt to represent accurately the electrostatic interaction in materials, while Bader charges do not achieve this objective[27]. The (Hill's averaged) bulk modulus and shear modulus also make an appearance in Table 3 but they show a very weak correlation. This is not unexpected as several mechanically similar materials with widely differing electronic properties exist in the dataset.

Machine learning

To better understand our results, we trained a simple explainable machine learning model to predict deformation potentials. We used model agnostic supervised local explanations (MAPLE)[28] in combination with random forests[29]. MAPLE models provide local explanations, feature selection, locally linear models and at least the accuracy of random forests[28]. As features we used all material and chemical properties that we considered in the previous section.

Using ten-fold cross-validation we achieved a mean absolute error of 0.83 ± 0.04 eV. This error should be compared to the mean absolute value and standard deviation of the cross-validation test sets of 2.54 and 2.74 eV. This means that the machine learning model has only a limited predictive capability for the deformation potential, which is perhaps expected due to the rather weak correlations found in Table 3. Also, the features selected (automatically) by MAPLE as the most relevant for the prediction (see Table 4) are not surprisingly found in Table 3. In view of these results, we conclude that simple chemical intuition, embodied in our interpretable machine learning models, is insufficient to satisfactorily explain the nature of the deformation potential of a material.

We also trained a machine learning model with better representation ability, to check if a more complex model is capable of reproducing the deformation potentials. We used a crystal graph convolutional network[30] with embedding size of 64 for the atomic representations before pooling, 128 after pooling, and three message passing steps. The network was trained for 1000 epochs with Adam[31], a learning rate of 0.01 and a batchsize of 64. As we tested larger and smaller network sizes with no improvement to the results, all other parameters were left at their default values in the CGCNN implementation by Xie et al. We used a split of 0.8/0.1/0.1 for training, validation, and testing, and the model with the best validation error was selected for testing. The mean absolute error for the training set at the end of the training was varying between 0.2 and 0.35 eV for different runs. This corresponds to roughly half the validation error, leading to an

Table 5. Full deformation potential for specific entries.

Material	MP-ID	Gap$_{PBE}$	Gap$_{mBJ}$	$\tilde{\tilde{\Xi}}_{gap}$ (eV)	Ξ_{PBE} (eV)	Ξ_{mBJ} (eV)
GeO_2	mp-470	1.19	3.54	−13.71	−10.21	−11.45
Sr_2AlSbO_6	mp-1208835	1.84	3.27	−12.92	−11.01	−11.39
$SrSnO_3$	mp-546973	0.94	2.95	−12.75	−12.34	−12.80
Sr_2GaSbO_6	mp-6065	0.68	2.60	−12.57	−11.20	−11.71
Al_2O_3	mp-1143	5.85	7.55	−12.50	−11.99	−12.47
$BaSr(SnO_3)_2$	mp-1227736	0.62	5.70	−12.34	−11.95	−12.22
$BaGe(PO_4)_2$	mp-1095485	4.09	8.98	−12.34	−6.12	−8.11
$TeWO_6$	mvc-5676	1.25	2.21	−12.25	−13.18	−14.27
$LiAlO_2$	mp-8001	6.17	8.05	−12.25	−9.59	−9.07
$SrGe(PO_4)_2$	mp-1101981	4.02	6.11	−12.24	−5.86	−3.35
PNO	mp-38975	5.14	6.42	−12.02	−7.78	−10.09
$LiAsF_6$	mp-9144	5.00	7.12	−11.91	−1.90	−1.72
$AlSbO_4$	mp-676861	1.52	3.04	−11.80	−8.42	−8.55
$GaBO_3$	mp-1078259	3.77	5.96	−11.76	−11.95	−12.63
$LiPF_6$	mp-9143	7.52	10.31	−11.70	−3.19	−2.03
Hg	mp-1184554	0.31	0.59	5.75	7.40	5.83
$CsSnCl_3$	mp-1070375	1.05	2.13	5.36	3.17	6.76
$CsSnBr_3$	mp-27214	0.79	1.36	5.14	2.38	2.72
Tl_3AsF_6	mp-1238924	3.16	4.75	5.07	1.19	−0.11
$RbSnBr_3$	mp-998157	0.97	1.59	4.91	5.16	7.96
$RbSnCl_3$	mp-998156	1.54	2.82	4.75	9.32	12.83
$TlGeBr_3$	mp-998153	1.51	2.21	4.65	4.73	6.25
$K_2InSbBr_6$	mp-1112183	0.24	0.48	4.57	6.80	6.64
Cs_2TlSbF_6	mp-1112563	2.50	4.95	4.54	3.55	6.63
SnS	mp-559676	0.39	0.72	4.54	2.05	3.08
$Rb_2InSbBr_6$	mp-1114390	0.31	0.55	4.53	6.77	6.87
$K_2InSbCl_6$	mp-1112017	0.69	1.30	4.51	6.92	10.11
Rb_3SbF_6	mp-1237359	3.69	5.08	4.50	−1.69	−2.54
LiAuS	mp-29829	1.57	2.32	4.50	4.81	5.30
$CsGeBr_3$	mp-570223	0.79	1.23	4.49	1.66	7.72

Values of the volume deformation potential for the materials with the highest absolute values of the estimator $\tilde{\tilde{\Xi}}_{gap}$. We show the materials project ID, the band gap calculated with both the PBE and the mBJ approximations, the estimator $\tilde{\tilde{\Xi}}_{gap}$, and the deformation potential calculated with PBE and mBJ.

acceptable level of overfitting. The final mean absolute error for the test set was 0.81 eV, not significantly better than the error obtained with MAPLE. Therefore, we assume that no sufficient data was available for the CGCNN model to improve its performance over the MAPEL model. Although the error is considerable, the machine is still able to efficiently identify the exceptional materials (according to the definition given before), based only on the structure. To demonstrate this we checked which materials from Table 5 were present in the test set and confirmed whether they were correctly predicted as compounds with a high absolute deformation potential by the CGCNN model. We repeated this process for four more random data set selections finding that, in the three cases when materials from Table 5 were present, 4 out of 6, 4 out of 4, and 3 out of 4 were correctly identified as 'extreme' materials. We also averaged the test errors of these five runs to arrive at an MAE of 0.82 ± 0.05 eV. The data and code to reproduce the machine learning models are available at https://github.com/hyllios/utils/tree/main/models/gap_deformation_potentials.

Specific materials

After having performed an overall analysis of the results, we select the materials with highest absolute $\tilde{\tilde{\Xi}}_{gap}$ for further analysis. We

start by computing the full deformation potential, Ξ_{gap}, i.e. including the effects of the change in internal parameters under pressure, for the 15 materials with the largest (positive) and the 15 materials with the lowest (most negative) deformation potentials. In practice this was done by straining isotropically the lattice vectors by ±1%, and by performing a geometry optimization at fixed volume. For each optimized structure, we obtained the respective band structure via interpolation, using the Boltz-Trap2[21,22] package. From the resulting values of the band gap at the three volumes (strained, stretched, and fully optimized) we computed Ξ_{gap}, using both the PBE and mBJ approximations for the exchange-correlation functional. The resulting values are shown in Table 5 together with the corresponding Materials Project ID for an unequivocal identification of the material.

By comparing the values of Ξ_{PBE} to the values of the $\tilde{\tilde{\Xi}}_{gap}$ we see that these are quite close for the vast majority of the materials, showing that $\tilde{\tilde{\Xi}}_{gap}$ is on average a good estimator of the deformation potential. There are some exceptions, however, and we find a few false positives in the table, the most striking of which are $LiAsF_6$, $LiPF_6$, Tl_3AsF_6, and Rb_3SbF_6. This is not unexpected, as there are materials where we can observe a strong variation of internal parameters with pressure[32,33].

We also see that in most cases the PBE and mBJ approximations lead to similar values for the deformation potential, with

discrepancies that often lie below 1 eV. However, it turns out that for some systems, and in particular few systems with extreme *positive* deformation potentials, these functionals yield different results. This happens for $CsSnCl_3$, $RbSnBr_3$, $RbSnCl_3$, Cs_2TlSbF_6, $K_2InSbCl6$, and $CsGeBr_3$. At the moment, it is not clear what is the source of this difference, nor which functional is best, due to the lack of experimental results for the deformation potentials of these specific systems.

In the following we analyze in more detail the compounds included in Table 5. Band structures and state-resolved densities of states for all these materials can be found in the Supporting Information.

a. *GeO_2* Rutile germanium oxide is a well-known material which, thanks to its optical properties[34] finds applications in lens and optic-fiber manufacturing[35]. It has been the focus of several studies in the literature, being sometimes described as an 'ultra-wide band gap' material[34,36], presenting a direct gap of 4.68 eV[37]. The sensitivity of the electronic properties of GeO_2 with respect to pressure is also known. In ref. [38] a PBE (G_0W_0) value of −13.05 eV (−17.30 eV) for Ξ_{gap} was found by studying the gap variation with volume (at fixed c/a ratio), which is in good agreement with our estimator of −13.71 eV. Note, however, that the assumption of constant c/a ratio is not representative of the actual response of the material under pressure[39]. In fact, relaxing this assumption decreases the PBE value by around 25% (see Table 5).

b. *$SrSnO_3$, $BaSr(SnO_3)_2$, Sr_2AlSbO_6, and Sr_2GaSbO_6* Gathering $SrSnO_3$, $BaSr(SnO_3)_2$, Sr_2AlSbO_6 and Sr_2GaSbO_6 we obtain a group of perovskites and double perovskites with large deformation potentials, some of which were already reported in the literature[40,41]. Besides $SrSnO_3$, we remark that also $BaSnO_3$ appears in our study with a considerable deformation potential ($\widetilde{\Xi}_{gap} = -11.69$ eV). In refs. [40,41] it was already recognized that $SrSnO_3$ and $BaSnO_3$ exhibit a very large variation of the band gap with volume, while the gap is not very sensitive to other types of strain. This behavior could be well modeled using a Fermi gas, and therefore the authors describe the strong variation of the gap as a signature of simple *s*-electron behavior in these complex oxides[41].

The mixing of Sr with Ba leads to the mixed perovskite $BaSr(SnO_3)_2$ that displays a $\widetilde{\Xi}_{gap}$ intermediate between the one of the two parent phases. A similar behavior can also be verified in other mixed perovskites, e.g. $KNbO_3$ (−1.94 eV), $NaNbO_3$ (−4.60 eV), and $KNa(NbO_3)2$ (−2.49 eV). This observation hints to the possibility of tuning deformation potentials of materials via cation mixing.

For other double perovskites such as Sr_2AlSbO_6 and Sr_2GaSbO_6, the parent ternary perovskites $SrSbO_3$ and $SrAlO_3$ are thermodynamically unstable (energy above hull larger than 50 meV/atom), while the structure of $SrGaO_3$ is not included in Materials Project. Therefore, their $\widetilde{\Xi}_{gap}$ are not calculated in the present work. Another related double perovskite, Ba_2InSbO_6, also displays a highly negative deformation potential of $\widetilde{\Xi}_{gap} = -11.17$ eV.

For these perovskites, the values of $\widetilde{\Xi}_{gap}$ differ slightly between antimonates and their stannate counterparts. However, there is a clear trend of less negative $\widetilde{\Xi}_{gap}$ in double perovskites going down group III from Al to In. Moreover, when substituting Sr with heavier Ba, the absolute value of $\widetilde{\Xi}_{gap}$ also decreases. This indicates that in (double) perovskites that share a common anion and have cations in the same elemental group, the heavier the cation the less negative is $\widetilde{\Xi}_{gap}$. Similar trends are found in ref. [42] for ABX_3 halide perovskites.

c. *Al_2O_3* α-Al_2O_3 is the alumina polymorph corresponding to the mineral corundum. As sapphire it finds applications as a

window material in high pressure experiments, while its doped counterpart ruby is used as a pressure indicator in diamond anvils. Previous calculations of gap versus pressure[43] already remarked its extraordinary behavior: a volume gap deformation potential of −12.51 eV was estimated in good agreement with the present results. The values of $\widetilde{\Xi}_{gap}$ for isomorphic α-Ga_2O_3 and α-In_2O_3 are respectively −10.45 and −6.83 eV, indicating that the deformation potential increases from Al to In, in agreement with the trend in perovskites discussed above.

d. *$BaGe(PO_4)_2$ and $SrGe(PO_4)_2$* $BaGe(PO_4)_2$, $SrGe(PO_4)_2$ are germanium double orthophosphates with crystal structures resembling that of the mineral yavapaiite ($KFe(SO_4)_2$, space group $C2/m$). They can be used as ceramic materials for microwave applications[44]. These two compounds turn out to have much lower absolute values of Ξ_{gap} than of $\widetilde{\Xi}_{gap}$, and therefore do not exhibit exceptional values of the deformation potential.

e. *$TeWO_6$* $TeWO_6$ crystallizes in the $Fm\overline{3}m$ space group and exhibits a record negative deformation potential (among the materials we studied). The $R\overline{3}$ phase of $TeWO_6$ (the same as $LiAsF_6$ and $LiPF_6$) has a less negative $\widetilde{\Xi}_{gap}$ of −7.57 eV. It is also interesting to remark that there are some tellurium double perovskites with a large negative $\widetilde{\Xi}_{gap}$, such as $CaTeBa_2O_6$ (−11.20 eV) and $MgTePb_2O_6$ (−11.18 eV). However the gaps of their tungsten counterparts are much less sensitive to compression, e.g. $CaWBa_2O_6$ (−3.36 eV) and $MgWPb_2O_6$ (−2.02 eV).

f. *$LiAlO_2$* Lithium aluminate (α-$LiAlO_2$) is one of the most efficient tritium breeding materials due to excellent performance under high radiation. The pressure-induced gap change of various phases of $LiAlO_2$ has been investigated using DFT, yielding an estimated Ξ_{gap} of −8.1 eV[45], in good agreement with our results.

g. *$LiAsF_6$ and $LiPF_6$* Lithium hexafluorophosphate and lithium hexafluoroarsenate are commonly used as the main conducting lithium salt in carbonate-based electrolytes for Li-ion batteries[46]. Similarly to $TeWO_6$, these two compounds crystallize in the $R\overline{3}$ space group. It turns out that there is a strong interplay between pressure and internal coordinates in these systems, leading to a strong decrease of the value of the deformation potential when these effects are taken into account. For example, in some classes of semiconductors it was already observed in literature that the band gap is more sensitive to changes of internal parameters (e.g., selective changes of bond lengths) than to volume changes. A typical example is $Cu(In,Ga)(S,Se)_2$ and more in general ABC_2 chalcopyrites[32,47], or Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$[33], whose gaps are known to be strongly dependent on small variations of internal parameters. In these systems it is well understood that changes of the internal parameters, e.g. changes of bond distances, have an effect on the band gap as the valence band is made of *p*–*d* hybridized states, where *p* electrons come from the anion and *d* electrons from the cation participating in the bond. A change of bond length has a strong impact on the hybridization and therefore on the valence band width, leading to significant band gap variations.

h. *PON* Phosphorus oxynitride, PON, is an isoelectronic analog of SiO_2. Like the latter, it has a complex $P-T$ phase diagram, and the selected structure is a high-pressure α-quartz-like phase[48]. The gap of α-quartz SiO_2, on the other hand, is much less sensitive to the volume change ($\widetilde{\Xi}_{gap} = -7.57$ eV). This can be explained again by the fact that the *p*–*p* interaction is stronger between P/N and O than that between Si and O.

i. *$AlSbO_4$* Aluminum antimonate can be seen as an analog to aluminum phosphate ($AlPO_4$) and arsenate ($AlAsO_4$) for which $\widetilde{\Xi}_{gap} = -7.12$ and −8.76 eV, respectively. Going from P to Sb, the *p*–*p* and *s*–*s* coupling with O is slightly weaker, but with a

stronger p–d interaction between Sb–O and As–O. As such, Ξ^V_{BM} can be much higher, leading to a more negative $\tilde{\bar{\Xi}}_{gap}$.

j. *GaBO₃* Gallium borate is isostructural with calcite $CaCO_3$. The other calcite-like borates, such as $AlBO_3$ and $InBO_3$ have smaller magnitude of $\tilde{\bar{\Xi}}_{gap}$ (−4.55 and −8.41 eV, respectively). This is not the same trend observed for the various group III compounds discussed above. The difference can be explained by the fact that in $AlBO_3$ the conduction band maximum is dominated by p–p anti-bonding states coming from the B–O interaction, which can considerably lower the deformation potential of the conduction band maximum in $AlBO_3$.

k. *Hg* Liquid mercury is metallic in normal conditions, and it even becomes superconducting at very low temperature. However, several other phases of mercury have been predicted theoretically. Taking the Materials Project database[49] as a reference, we can find several structures within a distance of 20 meV per atom from the convex-hull of stability. These range from metals to semiconductors with gaps of up to almost 1.5 eV. This is in our opinion an indication of the extreme sensitivity of mercury's electronic structure to its crystal structure. It has been noted in the literature[50,51] that the band structure of certain phases of mercury is very sensitive to pressure (and indeed to calculation parameters), and that a band gap opens up upon expansion. This is in agreement with the positive value of the deformation potential that we find here. This variation was connected to the change in character of the states close to the Fermi level that go from p-like to having an increased s-character for decreasing density[51]. The reason why mercury comes out as semiconducting in our calculations is very likely due to the use of PBE which tends to underbind and therefore leads to larger lattice parameters. For example, for the $R\bar{3}m$ (#166) phase, the trigonal lattice constant is 2.9863 Å at 5 K[52], while the PBE gives 3.179 Å. In conclusion, if we could experimentally access expanded phases of mercury, these would likely exhibit extremely large *positive* values of the gap deformation potential.

l. *Halide perovskites* The system we found with record positive deformation potential is the halide perovskite $RbSnCl_3$. Moreover, in the list of exceptional materials with positive deformation potentials we find a series of perovskites of composition ABX_3, such as {Rb,Cs}Sn{Cl,Br}₃ and $TlGeBr_3$, and their counterparts double perovskites {K,Rb,Cs}₂{In,Tl}Sb{F,Cl,Br}. Of course, the latter can be easily derived from the former by doubling the lattice (usually resulting in a face-centered cubic configuration), and performing the isoelectronic modification $2Sn \rightarrow InSb$ or $2Sn \rightarrow TlSb$. Halide perovskites are known to be exceptional materials, excelling as absorbers for photovoltaic devices, achieving efficiencies of more than 20%[53]. In recent reviews[54,55] it has already been discussed how the electronic structure and band gap of this class of materials can be tuned via changes in atomic structure. For example, iodide perovskites and related compounds show a very wide range of band gaps from ≈1.2 to ≈4.0 eV[55]. The same sensitivity of the band gap can also be observed in mixed-halide perovskites[56]. Trends in band gap are often dictated by structural subtleties involving geometric distortions and the ordering of anions[56]. Moreover, these materials have exceptionally small effective hole masses due to a strong hybridization between the s-states of the B-site element and the p-states of the halogen near the Fermi energy[26]. In ref. [54] it is also noted that a compression of the unit cell (for example, by chemical substitution, compressive biaxial strain, or high pressure) narrows the band gap, in agreement with the positive deformation potential we found here. This is explained by the fact that compression forces the B- and X-site ions closer together, enhancing the antibonding interactions at the top of the valence band, thereby raising

its energy and decreasing the band gap. It is also known that the variation of the band gap comes predominately from the variation of the valence band maximum, while the conduction band minimum has a positive, small variation[42]. That these materials exhibit high deformation potentials is even more extraordinary when we remark that this happens for a large range of band gaps (see Table 5). This points to the fact that the unique hybridization of the orbitals between the B- and X-atoms in these structures makes indeed halide perovskites materials with extreme properties.

m. *Tl₃AsF₆ and Rb₃SbF₆* These two structures are the pathological cases of Table 5, since their calculated deformation potential differs considerably from the screened value. Both materials share the same hypothetical structure, derived from $TlAsF_6$[57] and $KAsF_6$. Tl_3AsF_6 and Rb_3SbF_6 are, in any case, rather close to the convex hull of stability[58,59], lying respectively at 84 and 94 meV/atom above the hull with the PBE approximation, 98 and 94 meV/atom with PBE for solids[60], and 158 and 135 meV/atom with the SCAN functional[61]. The electronic band structure of Tl_3AsF_6 is rather peculiar, which might explain the behavior of Ξ_{gap}. The valence band is flat and isolated, with a gap of around 1 eV to the next occupied band. The valence band is comprised of s−As and p−F states, with the latter dominating. For the conduction band the situation is more complicated as PBE and mBJ lead to qualitatively different results. In PBE, the band gap is indirect, from near N to F. On the other hand, with mBJ the gap is direct, from Γ to Γ. The bands are mostly comprised of p orbitals of Tl, As and F (in decreasing order of DOS magnitude), consistently for both functionals. The situation for Rb_3SbF_6 is similar, as it also exhibits an isolated valence band with similar orbital decomposition (albeit with Sb taking the place of As). In addition, the conduction bands are quite differently described by both functionals.

n. *SnS* This is the high-temperature orthorhombic β-phase of tin sulfide (TlI-type, Cmcm, #63)[62,63]. The λ-type transition between the low-temperature GeS-type (Pnma, #62) α-phase and the high-temperature β-phase is essentially due to the continuous movement of the Sn and S atoms along the [100] direction[63]. The Pnma phase can also be seen as a small distortion of the more symmetric high-temperature Cmcm crystal structure[64]. For comparison, our estimated $\tilde{\bar{\Xi}}_{gap}$ for the α-phase (mp-2231) is 2.0 eV. The high value of our estimator for this material ($\tilde{\bar{\Xi}}_{gap}$ = 4.54 eV) is considerably reduced to Ξ_{PBE} = 2.05 eV allowing for the internal relaxation of the structure, a value much closer to the one we obtained for the α-phase. This is therefore a false positive.

o. *LiAuS* LiAuS synthesizes in an orthorhombic space (Fddd, #70)[65], in a structure that has been described as composed of 'chicken-wire-like' layers formed by interweaving $(AuS)^{n-}_n$ threads[66]. Also of note is that a cubic phase of LiAuS has been identified as the most interesting strong topological insulator (together with NaAuS) of all possible half-Heusler compounds[67]. It was also found that the band gap (and band inversion) in this compound changes rapidly, and with a non-monotonic behavior) with bi-axial strength[67]. Although we could not find electronic structure calculations for this compound in the literature, there are calculations for the related semiconducting compound NaAuS[68]. For this compound the top of the valence is composed of s- and d-Au states hybridized with the p-states of S, while the conduction band results from a mixture of states from all atoms. As can be seen in the Supplementary Information, the same is true for LiAuS.

In conclusion, we presented a large-scale dataset of gap deformation potentials for a wide variety of semiconductors. This dataset covers a large portion of the periodic table and is

representative of the possible crystal structures in which we find semiconductors. Analysis of the data shows that the distribution of deformation potentials is peaked at around -2 eV, and quite skewed towards negative values with a fat tail that extends to almost -15 eV. We find also a sizeable fraction of systems with positive deformation potentials, extending to around 5 eV.

Aiming at rationalizing this data, we searched for correlations between the deformation potentials and several other materials and compositional properties. We found an anti-correlation with the effective hole mass and the band gap, and a positive correlation with the mean row of the periodic table, the mean atomic mass and the volume of the unit cell. In any case, it turns out that all correlation coefficients are relatively small. We then constructed machine learning models designed to interpret and predict the values of the deformation potentials. The models exhibited a limited accuracy, showing the difficulty to describe deformation potentials in terms of simple chemical quantities for a large set of materials. In any case, the developed models can be successfully applied for the identification of materials with extreme values of this quantity.

Finally, we analyzed in more detail some of the exceptional materials we found. With negative values of the deformation potential we find exclusively oxides, the record being $TeWO_6$. On the other side of the spectrum appear mostly halide perovskites, the record belonging to $RbSnCl_3$. In contrast with the difficulties mentioned above to establish general correlations, the variation of the values of the deformation potential within specific families of materials can usually be explained based on the character of the valence and the conduction states.

METHODS

Calculations

The initial survey of the dataset was performed with density-functional theory at the level of the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation[69]. This is arguably the most used approximation for the ab initio study of solids, but leads to band gaps that are around a factor of two too small when compared to experiment[70,71]. To have a more reliable estimation of the band gaps, we also provide band structures obtained with the modified Becke–Johnson functional (mBJ)[72]. This approximation, when combined with the local-density approximation for the correlation term, was shown to yield accurate band-gaps[70,71]. Of course, this does not mean that deformation potentials are also more accurate in the mBJ than with the PBE, but we see that both functional turn out to yield similar results for most of the systems studied (see also Section I of the Supporting Information). This points to the fact that deformation potentials appear to be more robust properties than band gaps with respect to the theoretical approach used.

All calculations were performed within the projector-augmented wave formalism[73], using the Vienna ab-initio simulation package (VASP version 5.4.4)[74]. We use the set of pseudo-potentials recommended by the Materials Project database[49]. In the present work we opted not to include Hubbard corrections, as general purpose values of U might not transfer into quality improvements in the band structure and we are not in the capacity to extensively benchmark these parameters. The effect of this choice is also minimized thanks to the removal of compounds containing magnetic elements (see below). The band gaps were calculated as the difference of the Kohn-Sham eigenvalues from self-consistent calculations. Our energy cut-off was set to 520 eV and we used a Γ centered k-point grids with 8000 k-points per reciprocal atom. Inclusion of spin–orbit coupling leads to an average reduction of the bandgap of 0.1 eV[71]. Since this value is much smaller than the typical error coming from the choice of exchange-correlation functional, we opt to neglect this effect in our calculations. Moreover, as we consider *differences* of band gaps at different strains, we expect the error due to the neglect of spin–orbit coupling to cancel to a large extent.

Dataset preparation

To build our dataset, we selected semiconductors from the Materials Project[49] containing less than 13 atoms in the unit cell. Using these criteria we could gather 6552 systems. We then removed all systems containing magnetic elements (Fe, Ni, Cr, Mn, and lanthanides except La) and electronic band gap smaller than 0.01 eV, obtaining a dataset including 5338 systems. All ground-state geometries were taken directly from the Materials Project database[49] and were not further optimized. A table containing the Materials Project ID, together with our calculated gaps and deformation potentials can be found as Supplementary Information. The dataset presents a good sampling of the periodic table although the frequency of each chemical element is not homogeneous (see Supplementary Information). The chalcogens (particularly oxygen) are the most represented group, followed by the halogens, pnictogens, and alkali metals. Overall, this distribution is representative of the distribution of stable semiconductors present in the Materials Project[49].

DATA AVAILABILITY

Values of band gaps and deformation potentials used in the making of this work are presented as supplementary information. For other reasonable inquires, please contact the corresponding authors.

CODE AVAILABILITY

The dataset, the code and the machine learning models are freely available at https://github.com/hyllios/utils/tree/main/models/gap_deformation_potentials.

REFERENCES

1. Dalven, R. Possible inverted band structure in PbPo. *Phys. Rev. Lett.* **28**, 91–93 (1972).
2. Dalven, R. Empirical relation between energy gap and lattice constant in cubic semiconductors. *Phys. Rev. B* **8**, 6033–6034 (1973).
3. Yan, Q., Rinke, P., Janotti, A., Scheffler, M. & de Walle, C. G. V. Effects of strain on the band structure of group-III nitrides. *Phys. Rev. B* **90**, (2014) https://doi.org/10.1103/physrevb.90.125118.
4. Yan, Q., Rinke, P., Scheffler, M. & de Walle, C. G. V. Strain effects in group-III nitrides: deformation potentials for AlN, GaN, and InN. *Appl. Phys. Lett.* **95**, 121111 (2009).
5. Vurgaftman, I., Meyer, J. R. & Ram-Mohan, L. R. Band parameters for III–V compound semiconductors and their alloys. *J. Appl. Phys.* **89**, 5815–5875 (2001).
6. Vurgaftman, I. & Meyer, J. R. Band parameters for nitrogen-containing semiconductors. *J. Appl. Phys.* **94**, 3675–3696 (2003).
7. Wei, S.-H. & Zunger, A. Predicted band-gap pressure coefficients of all diamond and zinc-blende semiconductors: chemical trends. *Phys. Rev. B* **60**, 5404 (1999).
8. Rabii, S. Energy-gap anomaly in the pb chalcogenides and electronic structure of PbPo. *Phys. Rev. B* **18**, 1876–1879 (1978).
9. Nie, S., Xu, X. Y., Xu, G. & Fang, Z. Band gap anomaly and topological properties in lead chalcogenides. *Chin. Phys. B* **25**, 037311 (2016).
10. Long, D. Energy bands in semiconductors. *J. Appl. Phys.* **33**, 1682–1696 (1962).
11. Dalven, R. Energy-gap anomaly in the semiconductor sequence PbS, PbSe, and PbTe. *Phys. Rev. B* **3**, 3359–3367 (1971).
12. Wu, J. When group-III nitrides go infrared: new properties and perspectives. *J. Appl. Phys.* **106**, 011101 (2009).
13. Balaghi, L. et al. Widely tunable GaAs bandgap via strain engineering in core/shell nanowires with large lattice mismatch. *Nat. Commun.* **10**, https://doi.org/10.1038/s41467-019-10654-7 (2019) .
14. Roldán, R., Castellanos-Gomez, A., Cappelluti, E. & Guinea, F. Strain engineering in semiconducting two-dimensional crystals. *J. Phys.: Condens. Matter* **27**, 313201 (2015).
15. Britnell, L. et al. Field-effect tunneling transistor based on vertical graphene heterostructures. *Science* **335**, 947–950 (2012).
16. Suckert, J. R., Rödl, C., Furthmüller, J., Bechstedt, F. & Botti, S. Efficient strain-induced light emission in lonsdaleite germanium. *Phys. Rev. Mater.* **5**, 024602 (2021).
17. Pei, Y., Wang, H. & Snyder, G. J. Band engineering of thermoelectric materials. *Adv. Mater.* **24**, 6125–6135 (2012).
18. Bader, R. F. W. Atoms in molecules: a quantum theory. In *The International Series of Monographs on Chemistry* No. 22 (Clarendon Press; Oxford University Press, Oxford [England]: New York, 1994).
19. Tang, W., Sanville, E. & Henkelman, G. A grid-based bader analysis algorithm without lattice bias. *J. Phys. Condens. Matter* **21**, 084204 (2009).

20. Manz, T. A. & Limas, N. G. Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology. *RSC Adv.* **6**, 47771–47801 (2016).

21. Madsen, G. K. & Singh, D. J. Boltztrap. a code for calculating band-structure dependent quantities. *Comput. Phys. Commun.* **175**, 67–71 (2006).

22. Madsen, G. K., Carrete, J. & Verstraete, M. J. BoltzTraP2, a program for interpolating band structures and calculating semi-classical transport coefficients. *Comput. Phys. Commun.* **231**, 140–145 (2018).

23. Ward, L. et al. Matminer: an open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60 (2018).

24. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *NPJ Comput. Mater.* **5**, (2019).

25. Chok, N. S. *Pearson's versus Spearman's and Kendall's Correlation Coefficients for Continuous Data*. Ph.D. thesis, University of Pittsburgh (2010).

26. Wang, H.-C., Pistor, P., Marques, M. A. L. & Botti, S. Double perovskites as p-type conducting transparent semiconductors: a high-throughput search. *J. Mater. Chem. A* **7**, 14705–14711 (2019).

27. Choudhuri, I. & Truhlar, D. G. Calculating and characterizing the charge distributions in solids. *J. Chem. Theory Comput.* **16**, 5884–5892 (2020).

28. Plumb, G., Molitor, D. & Talwalkar, A. S. Model agnostic supervised local explanations. In Bengio, S., Wallach, H., Larochelle, H., et al (eds) *Advance in Neural Information Processing Systems* Vol. 31, 2515 (Curran Associates, Inc., 2018).

29. Breiman, L. Random forests. *Mach. Learn.* **45**, 5 (2001).

30. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).

31. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).

32. Jaffe, J. E. & Zunger, A. Theory of the band-gap anomaly in ABC_2 chalcopyrite semiconductors. *Phys. Rev. B* **29**, 1882–1906 (1984).

33. Botti, S., Kammerlander, D. & Marques, M. A. L. Band structures of Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$ from many-body methods. *Appl. Phys. Lett.* **98**, 241915 (2011).

34. Mengle, K. A., Chae, S. & Kioupakis, E. Quasiparticle band structure and optical properties of rutile GeO_2, an ultra-wide-band-gap semiconductor. *J. Appl. Phys.* **126**, 085703 (2019).

35. Sierra, J. H. et al. Pedestal waveguides based on GeO_2–Bi_2O_3, GeO_2–PbO, Ta_2O_5 and SiO_xN_y cores as platforms for optical amplifiers and nonlinear optics applications: review of recent advances. *J. Lumin.* **236**, 118113 (2021).

36. Chae, S., Lee, J., Mengle, K. A., Heron, J. T. & Kioupakis, E. Rutile GeO_2: an ultrawide-band-gap semiconductor with ambipolar doping. *Appl. Phys. Lett.* **114**, 102104 (2019).

37. Stapelbroek, M. & Evans, B. Exciton structure in the uv-absorption edge of tetragonal GeO_2. *Solid State Commun.* **25**, 959–962 (1978).

38. Samanta, A., Jain, M. & Singh, A. K. Ultra-sensitive pressure dependence of bandgap of rutile-GeO_2 revealed by many body perturbation theory. *J. Chem. Phys.* **143**, 064703 (2015).

39. Haines, J., Léger, J., Chateau, C. & Pereira, A. Structural evolution of rutile-type and $CaCl_2$-type germanium dioxide at high pressure. *Phys. Chem. Miner.* **27**, 575–582 (2000).

40. Fan, X., Zheng, W., Chen, X. & Singh, D. J. 2degs at perovskite interfaces between $KTaO_3$ or $KNbO_3$ and stannates. *PLoS ONE* **9**, e91423 (2014).

41. Singh, D. J., Xu, Q. & Ong, K. P. Strain effects on the band gap and optical properties of perovskite $SrSnO_3$ and $BaSnO_3$. *Appl. Phys. Lett.* **104**, 011910 (2014).

42. Wang, S., Huang, M., Wu, Y.-N. & Chen, S. Absolute volume deformation potentials of inorganic ABX_3 halide perovskites: the chemical trends. *Adv. Theory Simul.* **4**, 2100060 (2021).

43. Boettger, J. C. High-precision, all-electron, full-potential calculation of the equation of state and elastic constants of corundum. *Phys. Rev. B* **55**, 750–756 (1997).

44. Tudorache, F. et al. Dielectric investigation of MIIMIV(PO4)2 double orthophosphates (MII = Ca, Sr, Ba, Pb; MIV = Ti, Zr, Hf, Ge, Sn. J. Alloys Compd. 509, 9127–9132 (2011).

45. Singh, B., Gupta, M. K., Mittal, R. & Chaplot, S. L. Phonons, phase transitions and thermal expansion in $LiAlO_2$: an ab initio density functional study. *Phys. Chem. Chem. Phys.* **20**, 12248–12259 (2018).

46. Xu, G., Shangguan, X., Dong, S., Zhou, X. & Cui, G. Formulation of blended-lithium-salt electrolytes for lithium batteries. *Angew. Chem. Int. Ed.* **59**, 3400–3415 (2019).

47. Vidal, J., Botti, S., Olsson, P., Guillemoles, J.-Fmc & Reining, L. Strong interplay between structure and electronic properties in CuIn(**S**,S)$_2$: a first-principles study. *Phys. Rev. Lett.* **104**, 056401 (2010).

48. Léger, J.-M. et al. Crystal structure and high pressure behaviour of the quartz-type phase of phosphorus oxynitride PON. *J. Phys. Chem. Solids* **60**, 145–152 (1999).

49. Jain, A. et al. The materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

50. Fritzson, P. & Berggren, K.-F. A pseudopotential calculation of the density of states of expanded crystalline mercury. *Solid State Commun.* **19**, 385–387 (1976).

51. Devillers, M. A. C. & Ross, R. G. Structure independent properties of expanded crystalline mercury. *J. Phys. F: Met. Phys.* **5**, 73–80 (1975).

52. Barrett, C. S. The structure of mercury at low temperatures. *Acta Crystallogr.* **10**, 58–60 (1957).

53. Stranks, S. D. & Snaith, H. J. Metal-halide perovskites for photovoltaic and light-emitting devices. *Nat. Nanotechnol.* **10**, 391–402 (2015).

54. Berger, R. F. Design principles for the atomic and electronic structure of halide perovskite photovoltaic materials: insights from computation. *Chem. Eur. J.* **24**, 8708–8716 (2018).

55. Xiao, Z., Zhou, Y., Hosono, H., Kamiya, T. & Padture, N. P. Bandgap optimization of perovskite semiconductors for photovoltaic applications. *Chem. Eur. J.* **24**, 2305–2316 (2018).

56. Mosconi, E., Amat, A., Nazeeruddin, M. K., Grätzel, M. & Angelis, F. D. First-principles modeling of mixed halide organometal perovskites for photovoltaic applications. *J. Phys. Chem. C* **117**, 13902–13913 (2013).

57. Goreshnik, E. & Mazej, Z. Single crystal structures of $InPF_6$, $InAsF_6$, $TlPF_6$ and $TlAsF_6$. *Solid State Sci.* **10**, 303–306 (2008).

58. Schmidt, J., Wang, H.-C., Cerqueira, T. F. T., Botti, S. & Marques, M. A. L. A new dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals https://archive.materialscloud/record/2021.60 (2021).

59. Schmidt, J., Wang, H.-C., Cerqueira, T. F. T., Botti, S. & Marques, M. A. L. A dataset of 175k stable and metastable materials calculated with the PBEsol and SCAN functionals. *Sci. Data* **9**, https://doi.org/10.1038/s41597-022-01177-w (2022).

60. Perdew, J. P. et al. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).

61. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).

62. von Schnering, H. G. & Wiedemeier, H. The high temperature structure of β-SnS and β-SnSe and the B16-to-B33 type λ-transition path. *Z. Kristallogr.-Cryst. Mater.* **156**, 143–150 (1981).

63. Chattopadhyay, T., Pannetier, J. & Schnering, H. V. Neutron diffraction study of the structural phase transition in SnS and SnSe. *J. Phys. Chem. Solids* **47**, 879–885 (1986).

64. Skelton, J. M., Burton, L. A., Oba, F. & Walsh, A. Chemical and lattice stability of the tin sulfides. *J. Phys. Chem. C* **121**, 6446–6454 (2017).

65. Huang, F. Q., Yang, Y., Flaschenriem, C. & Ibers, J. A. Syntheses and structures of LiAuS and Li_3AuS_2. *Inorg. Chem.* **40**, 1397–1398 (2001).

66. Axtell, E. A., Liao, J.-H. & Kanatzidis, M. G. Flux synthesis of LiAuS and NaAuS: "chicken-wire-like" layer formation by interweaving of (AuS)nn-threads. comparison with α-HgS and AAuS (A = K, Rb). *Inorg. Chem.* **37**, 5583–5587 (1998).

67. Lin, S.-Y. et al. Theoretical search for half-heusler topological insulators. *Phys. Rev. B* **91**, https://doi.org/10.1103/physrevb.91.094107 (2015).

68. Reshak, A., Khan, S. A., Kamarudin, H. & Bila, J. NaAuS chicken-wire-like semiconductor: electronic structure and optical properties. *J. Alloy. Compd.* **582**, 6–11 (2014).

69. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

70. Borlido, P. et al. Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. *Npj Comput. Mater.* **6**, 1–17 (2020).

71. Borlido, P. et al. Large-scale benchmark of exchange–correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theory Comput.* **15**, 5069–5079 (2019).

72. Tran, F. & Blaha, P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.* **102**, 226401 (2009).

73. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

74. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

75. Blacha, A., Presting, H. & Cardona, M. Deformation potentials of k = 0 states of tetrahedral semiconductors. *Phys. Status Solidi B* **126**, 11–36 (1984).

AUTHOR CONTRIBUTIONS

P.B. performed the screening calculations. J.S. performed the machine learning calculations. H.W. and P.B. performed calculations for specific materials. M.A.L.M. and S.B. designed the research and directed the work. All authors contributed for the analysis and writing of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00811-w.

Correspondence and requests for materials should be addressed to Silvana Botti or Miguel A. L. Marques.

Reprints and permission information is available at http://www.nature.com/reprints