ARTICLE   OPEN

Parametrically constrained geometry relaxations for high-throughput materials science

Maja-Olivia Lenz [1], Thomas A. R. Purcell [1]*, David Hicks [2], Stefano Curtarolo [2], Matthias Scheffler[1] and Christian Carbogno [1]

Reducing parameter spaces via exploiting symmetries has greatly accelerated and increased the quality of electronic-structure calculations. Unfortunately, many of the traditional methods fail when the global crystal symmetry is broken, even when the distortion is only a slight perturbation (e.g., Jahn-Teller like distortions). Here we introduce a flexible and generalizable parametric relaxation scheme and implement it in the all-electron code FHI-aims. This approach utilizes parametric constraints to maintain symmetry at any level. After demonstrating the method's ability to relax metastable structures, we highlight its adaptability and performance over a test set of 359 materials, across 13 lattice prototypes. Finally we show how these constraints can reduce the number of steps needed to relax local lattice distortions by an order of magnitude. The flexibility of these constraints enables a significant acceleration of high-throughput searches for novel materials for numerous applications.

INTRODUCTION

Symmetry preservation and breaking is one of the most fundamental processes in physics and chemistry. Many properties and applications, such as piezoelectricity,[1–4] pyroelectricity,[5,6] ferroelectricity,[7–10] topological insulators,[11,12] and nonlinear optics,[13–15] require certain selection rules to be met, and therefore require certain crystallographic symmetries to be maintained. Furthermore, it is not only global symmetry, the space and point groups of a material, but also local symmetry breaking that matters. For example, defects can cause significant changes in a material's mechanical[16,17] and optical[18,19] properties, as well as in its electronic[20–22] and thermal transport[23–25] coefficients. These effects can be particularly important at thin-film interfaces where interactions between different layers can induce systematic distortions in the structure of the film. In turn, these distortions can lead to novel properties in the materials, such as artificial ferroelectricity in layered perovskite supercell structures.[26,27] Clearly, such problems cannot be addressed by enforcing a global symmetry constraint, e.g., space group conservation, for the whole system, but require selectively preserving and breaking a material's symmetries locally, e.g., around defects or in the individual perovskite layers. This is paramount in computational material science, especially in high-throughput studies, which often aim at calculating and exploring yet unknown properties of already known materials, e.g., band structures,[28] defect-formation energies,[29] elastic and thermal properties,[30] and topological constants.[31,32] Similarly, many high-throughput studies aim at discovering potentially stable or metastable materials by decorating complex, well-known crystal structures such as Heuslers[33] and perovskites[34] with different species, or by systematically exploring a given alloy system.[35] To streamline such calculations, it is essential to keep both global and local symmetries under control, especially when complex materials or material properties are targeted. In this work, we achieve this goal by proposing and implementing parametric geometric constraints that allow for the enforcing or breaking of symmetries, both globally and locally. Before applying these constraints, an understanding of how they

map onto the target systems (e.g., the number of atoms in the unit cell, space group, or the effects of local distortions on the structure) is necessary. To facilitate setting up such constraints, we rely on the AFLOW Library of Crystallographic Prototypes[36,37] to generate the initial mapping of real space onto a reduced parameter space that fully describes a system. One can then manually alter the initial mapping to add or lift constraints as needed. This allows for the efficient targeting of specific geometric configurations and avoids revisiting and recalculating already-investigated configurations.

Traditionally, crystallographic symmetries are incorporated in first-principles codes already at the electronic-structure level (e.g., by sampling **k**-space grids in the irreducible part of the Brillouin zone, as implemented in Vasp[38] or by sampling real space in symmetry defined "irreducible wedges", as done in PARSEC[39]), since it leads to significant savings in memory and computational workload for highly symmetric crystals. Also, by this means the obtained forces on the atoms and stresses on the lattice vectors fully reflect the crystallographic symmetries. Since geometry relaxation algorithms such as steepest descent, conjugate gradient, Newton–Raphson, quasi-Newton (e.g., BFGS[40]), and truncated-Newton methods[41] rely on the forces and stresses to update the atomic and lattice degrees of freedom, global symmetries are inherently preserved in such approaches. However, this does not allow for the partial, local symmetry breaking discussed above. To address such cases in first-principles calculations, it is typically necessary to lift all crystallographic symmetry constraints and treat the atomic and lattice degrees of freedom as a set of freely changing parameters. Besides the increased computational cost, such unconstrained structure optimization can lead to long and inefficient relaxation trajectories, resulting in structures far from the ideal geometry. While in some cases, this problem can be circumvented by fixing atomic, lattice, or internal[42,43] degrees of freedom (as done in Quantum Espresso[44] or VASP[38]), mapping local distortions onto these constraints requires cumbersome manual inspection and analysis, if it is even possible. A more direct approach targeting how the

[1]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4–6, D-14195 Berlin, Germany. [2]Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA. *email: purcell@fhi-berlin.mpg.de
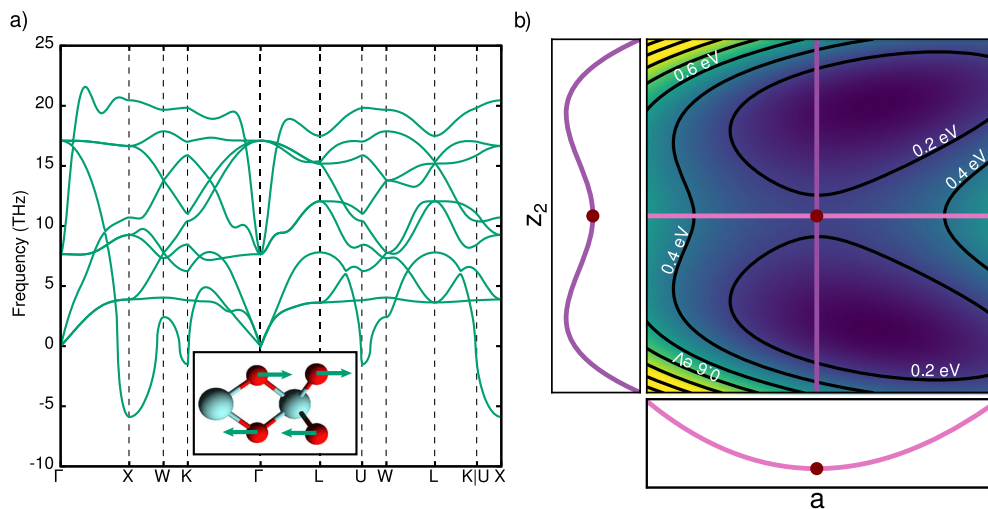
Fig. 1 Reduced Parameter Space for 12-Atom Zirconia Supercell. An illustration of the new relaxation scheme. **a** The phonon band structures of cubic ZrO_2. The inset illustrates the eigenvector of the imaginary mode that drives the system toward the tetragonal structure in a six-atom supercell. **b** A two-dimensional potential-energy surface for ZrO_2. The minimum energy structure is set to 0.0 eV, and the contour lines correspond to a 0.2 eV increase in energy. The maroon dot represents a structure corresponding to a high-temperature cubic phase of the material. The outsets show the one-dimensional potential-energy surface along each mode.

distortion changes the native crystal structure provides an easier and better way of treating these systems.

Here, we present a scheme to incorporate parametric constraints in structure optimizations that treats all levels of symmetry equally. The proposed approach employs a mapping of the relevant degrees of freedom onto a lower-dimensional representation of the structure; the respective forces and stresses are then automatically mapped in this reduced representation. With that, the implemented formalism does not require altering the employed relaxation algorithm, while still allowing the introduction of arbitrary constraints in a user-friendly manner. We first describe how the methodology works and the tools that can be used to quickly generate new structures. We demonstrate that these constraints allow for performing geometry optimizations on dynamically stabilized structures, which are not easily addressable otherwise. By analyzing the constrained and unconstrained relaxations of a test set of 359 materials, we then show that these constraints are also computationally beneficial for the relaxation of stable materials. Finally, we illustrate how the parameters can be used to selectively break symmetries and accelerate relaxations in supercells.

Transformation to reduced space

For a free relaxation, the optimizer acts on the full $3N + 9$ dimensional potential-energy surface $E(\mathbf{R}, \mathbf{L})$ of a material, which is encoded by the atomic, $\mathbf{R}$, and lattice, $\mathbf{L}$, degrees of freedom. The lattice degrees of freedom are stored as the three components of the three lattice vectors in the chosen unit cell, and the atomic degrees of freedom are the $3N$ components of the positions of the N atoms in a unit cell, represented by Cartesian or fractional coordinates. The forces, $\mathbf{F}$, acting on the atomic degrees of freedom are the derivatives of the energy with respect to $\mathbf{R}$

$$\mathbf{F} = -\frac{\mathrm{d}E}{\mathrm{d}\mathbf{R}}, \tag{1}$$

while the forces acting on the lattice vectors stem from the stresses, σ,

$$\sigma = \frac{1}{V}\frac{\mathrm{d}E}{\mathrm{d}\mathbf{L}}, \tag{2}$$

where V is the volume of the unit cell. In ab initio approaches, E is determined by solving the electronic-structure problem, and the

respective derivatives are obtained analytically via the Hellmann–Feynman Theorem. However, in practice this requires to account for additional terms, such as the Pulay terms and multipole corrections, as done in FHI-aims.[45,46]

Because the underlying potential-energy surfaces (PES) are complex, relaxing certain polymorphs of a material on these surfaces can be challenging or even impossible. As an example, zirconia (ZrO_2) can exist in its pure form in three different crystal phases: a high-temperature ($T > 2370\,°C$) cubic phase, an intermediate temperature ($1170\,°C \leq T \leq 2370\,°C$) tetragonal phase, and a low-temperature ($T < 1170\,°C$) monoclinic phase.[47] In particular, the cubic phase is a dynamically stabilized phase representing an average structure that is rarely, if ever actualized in pristine ZrO_2.[48] Accordingly, this phase constitutes a saddle point of the PES, and its phonon band structure exhibits an imaginary mode at the X point,[49] as illustrated in Fig. 1a. The associated eigenvector, illustrated in the respective inset, describes a pairwise, antiparallel distortion of the oxygen atoms that goes hand-in-hand with a stretching of the lattice and leads to the tetragonal structure.[47] To help illustrate this, in Fig. 1b we plot a two-dimensional PES for a 12-atom zirconia unit cell over a reduced parameter set describing the cubic lattice parameter a and the motion along the imaginary mode, z_2. The PES has two wells corresponding to the equivalent tetragonal structures, and a saddle point between them representing a high-symmetry configuration, i.e., the high-temperature cubic phase. On this PES, a free relaxation of the cubic phase would result in the material relaxing toward a local minima; however, by constraining the relaxation to act only on a, the cubic phase can be obtained as shown in the outsets in Fig. 1b.

To help demonstrate how parametric constraints can facilitate addressing these stable/unstable polymorphs, we provide the respective constraints in Tables 1 and 2. As discussed later, these symmetry blocks correspond to the input files that are required in the proposed formalism for the 12-atom zirconia unit cell. For the cubic polymorph, all atoms are at fixed fractional coordinates and only one parameter, i.e., the lattice constant, a, can change in a parametrically constrained relaxation. For the relaxation algorithm, this means that the optimizer can now act only on a, and all other degrees of freedom remain untouched. Clearly, this ensures that the initial space group is retained during the relaxation. As mentioned above, such constraints are necessary, since this polymorph corresponds to a saddle point of the PES. This means that an unconstrained or free relaxation would effectively allow

Table 1. Example of lattice vector parametric constraints.

	Cubic			Tetragonal		
	L^x	L^y	L^z	L^x	L^y	L^z
L_1	a	0	0	a	0	0
L_2	0	a	0	0	a	0
L_3	0	0	a	0	0	c

Parametric expressions for each component of the lattice vectors in the 12-atom cubic and tetragonal ZrO_2 supercell

Table 2. Example of atomic parametric constraints.

	Cubic			Tetragonal		
Atom	L_1	L_2	L_3	L_1	L_2	L_3
Zr	0.00	0.00	0.00	0.00	0.00	0.00
Zr	0.50	0.50	0.00	0.50	0.50	0.00
Zr	0.00	0.50	0.50	0.00	0.50	0.50
Zr	0.50	0.00	0.50	0.50	0.00	0.50
O	0.25	0.25	0.25	0.25	0.25	$0.25 - z_2$
O	0.25	0.75	0.25	0.25	0.75	$0.25 - z_2$
O	0.75	0.75	0.75	0.75	0.75	$0.75 + z_2$
O	0.25	0.25	0.75	0.25	0.25	$0.75 + z_2$
O	0.75	0.25	0.25	0.75	0.25	$0.25 - z_2$
O	0.25	0.75	0.75	0.25	0.75	$0.75 - z_2$
O	0.75	0.75	0.75	0.75	0.75	$0.75 + z_2$
O	0.25	0.25	0.25	0.25	0.25	$0.25 + z_2$

Parametric expressions for each component of the atomic positions in the twelve atom cubic and tetragonal ZrO_2 supercell

the system to break its symmetry and to descend into a local minimum. To explicitly explore these local minima, the constraints imposed on the cubic cell can be lifted in a stepwise manner, whereby the information contained in the imaginary phonon eigenvectors can be incorporated as parametric constraints. As shown in Tables 1 and 2, the pairwise distortion of the oxygen atoms for the imaginary mode at X (Fig. 1a) can be described by introducing one additional parameter for the oxygen distortion, z_2, and one for the tetragonality of the lattice c. These constraints ensure that the geometry optimization occurs along the imaginary phonon mode and leads to the tetragonal minimum. Conversely, a free relaxation can again lead to other local minima of the PES. In this textbook example, the same constraints could have been imposed by relaxing cubic and tetragonal zirconia in their primitive cells with six and three atoms, respectively. However, this is not always the case: some materials, e.g., bismuth oxide,[50,51] discussed later, have high-temperature polymorphs with the same number of sites as their stable structures, or more.

The previous input example illustrates the flexibility of these constraints, but knowledge of which reduced parameters to use and their relation to the full geometry, must be known before generating an input file. For crystals, these are determined by the space group and the Wyckoff positions, and can therefore be manually constructed. Luckily, these parameters are already a part of the definitions used in the AFLOW Library of Crystallographic Prototypes, allowing for an easy way to define these constraints for numerous materials via their utilities.[36,37] The library sorts materials by their space group, stoichiometry, and occupied Wyckoff sites, as calculated with AFLOW-SYM,[52] placing all materials that share those features into the same crystal

prototype.[36,37] A reduced parameter space can then be generated from a prototype definition, and used to describe that class of materials. For example, the tetragonal phase of zirconia can be described by only three parameters: the length of the lattice vectors in the **a** and **b** directions (a), the ratio of the lattice vectors ($\frac{c}{a}$), and the magnitude of the oxygen distortions (z_2). These parameters represent the same ones we defined earlier from the analysis of the phonon band structure, with an additional parameter allowing for the relaxation of the lattice upon the atomic distortions. For all prototypes defined in the library, the automatized generation of input geometries for VASP,[38] FHI-aims,[45] Quantum Espresso,[44] Abinit,[53] and more codes is supported by AFLOW. The AFLOW library contains 590 unique structure prototypes across all 230 space groups, and is thus extremely suitable as a starting point for high-throughput studies. As of version 3.1.204, the option `--add_equations` can be added to the AFLOW command to generate FHI-aims `geometry.in` files already containing the additional block required for the constrained relaxation. Because of this, we use the crystal prototypes defined by the AFLOW Library of Crystallographic Prototypes throughout this work. Due to the analytic representation of the parametric expressions, it is also straightforward to add additional parameters to allow for lower-symmetric structures or distortions as well as removing parameters to constrain specific components even further. In addition, because the AFLOW prototypes are partially based on the space group and occupied Wyckoff sites, it is also straightforward to adapt their techniques to include structure classes not currently in the library.

RESULTS

Relaxing metastable and unstable systems

In some cases, constraining a relaxation is necessary to keep the structure in its given polymorph. Similar to what was seen for zirconia, a material can have many phases that are metastable or unstable at zero-point conditions, but are stabilized by entropic contributions at higher temperatures or pressures. Here, we define a metastable phase to be one that is in a local minimum of the PES. While freely relaxing stable or metastable structures is possible by using an initial geometry near its corresponding (global or local) minimum on the PES, unstable systems will tend to relax towards lower energy and usually lower-symmetric structures, unless they are somehow constrained.

To demonstrate the ability of these constraints to optimize such structures, we relax the 12-atom cubic zirconia unit cell from Fig. 1. While most relaxations will be performed on the primitive cells of structures, we use this system as a simple, demonstrative example. The initial geometry for the cubic phase was taken from the AFLOW Library of Crystallographic Prototypes, while the one for the tetragonal phase is the same as the cubic phase with a minor perturbation along the imaginary mode seen in Fig. 1.

All calculations are done using the FHI-aims package, a full-potential, all-electron electronic-structure code. FHI-aims utilizes numeric atom-centered orbital basis functions, grouped into different tiers beyond the minimal set needed to describe free atoms. For these calculations, we use 'tier 1' (as defined in Blum et al.[45]) with 'light' basis settings, which were shown to calculate the lattice parameter and cohesive energy of face-centered cubic gold within 0.001 Å and 20 meV,[45] respectively. We use PBEsol as the exchange-correlation functional; SCF convergence criteria of 10^{-6} eV/Å and 5×10^{-4} eV/Å for the density and forces, respectively; and the structures are relaxed until the maximum forces on the degrees of freedom are below 0.005 eV/Å. All other inputs were taken to be the default values in FHI-aims. While a larger basis set and using a hybrid functional would increase the accuracy of the calculations, we do not expect it to affect the performance of the relaxation scheme.
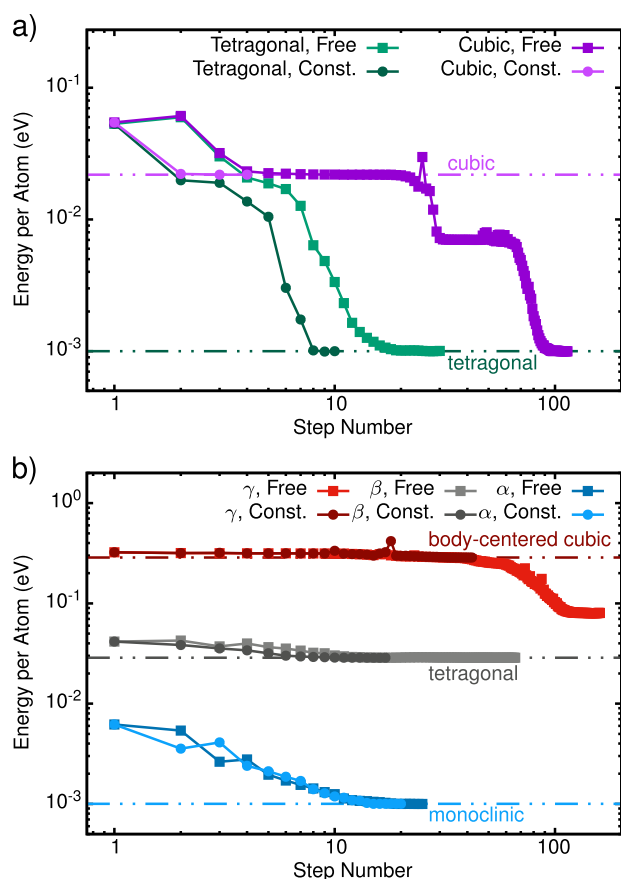
Fig. 2 Using parameteric constraints to converge to unstable and metastable phases. Convergence behavior of the free (squares) and constrained (circles) relaxations for **a** the tetragonal phase (green) and cubic (purple) phases of ZrO_2 and **b** the α-phase (blue), β-phase (gray), and γ-phase (red) of Bi_2O_3. For both the cubic phase of ZrO_2 and the γ-phase of Bi_2O_3, the free relaxation breaks the symmetry and finds an energetically lower structure which is the tetragonal phase for ZrO_2 and an unknown phase for Bi_2O_3. The energy scale on the y-axis is set to 1 meV below the minimum energy for each material.

Figure 2a shows that using the constraints both the cubic and tetragonal phase of ZrO_2 can be converged in four and ten steps, respectively, while only the tetragonal phase can be obtained in 30 steps with a free relaxation. The free relaxation of the cubic phase proceeds toward the tetragonal phase, but initially stalls at a non-physical simple cubic structure in 37 steps. If the relaxation convergence criteria is further reduced to 0.001 eV/Å, the structure reaches the tetragonal phase in 114 steps.

Another example of a material with many metastable phases is bismuth oxide. Bismuth oxide exists in several different polymorphs[50] including the low-temperature monoclinic phase, α-; the high-temperature, face-centered cubic phase, δ-; the metastable, body-centered cubic phase, γ-; and the metastable, tetragonal phase β-Bi_2O_3.[51] Upon heating α-Bi_2O_3 transforms into δ-Bi_2O_3 at around 730 °C, and remains stable until it melts at ~825 °C.[51] Depending on the cooling procedure, δ-Bi_2O_3 transitions to one of the two metastable phases, β- or γ-Bi_2O_3, at ~650 °C or ~640 °C, respectively.[51] Upon further cooling, β-Bi_2O_3 and γ-Bi_2O_3, respectively, return to the α-phase at ~300 °C and at a temperature dependent on the cooling rate.[51] Importantly, unlike ZrO_2, both the tetragonal and body-centered cubic phases have as many or more atoms in their primitive cells than the α-phase, so freely relaxing both structures in their primitive cells should not prevent them from exploring lower symmetry structures. For these

calculations, we use the same computational settings as those used for ZrO_2, but with the 'intermediate' settings for the basis set. The 'intermediate' settings and basis sets in FHI-aims increase the accuracy of the default numerical settings, but more importantly add an f-orbital to the 'tier 1' basis set for oxygen, which is necessary for describing monoclinic structures containing oxygen.

Figure 2b shows the relaxation of the α-, β-, and γ-bismuth oxide phases. Multiple experimental crystal structure refinements exist[51] for both the β- and γ-phases, so we take the relaxed structures from the Materials Project,[54] which were initialized from structures taken from the ICSD. In all cases, the constrained relaxation takes less time to converge with the γ-, β-, and α-phase, respectively, needing 42, 17, and 20 steps to converge. Freely relaxing both the α and β-phases simply increases the number of steps needed to converge the system to 25 and 66 steps, respectively, but removing the constraints for γ-Bi_2O_3 causes it to relax to an unknown, non-symmetric structure in 160 steps. While it is possible that the divergent relaxation is a result of an incorrect structure refinement, if that is the case, then it suggests that comparing the final energy and geometry of the constrained and free relaxation trajectories can be used to indicate if a structure is correct. Since the free relaxation of the γ-phase departs from the known phases of the material, a more in-depth study of it would be impossible without the constraints, as the fully relaxed structures no longer represent the same material.

Both of these cases demonstrate the need for constraining relaxations to their crystal prototype for high-throughput applications. For the high-symmetry phases of both zirconia and bismuth oxide, the free relaxation not only initially converged to a different phase but also to unknown and potentially physically unrealizable ones. While the relaxation of ZrO_2 does eventually reach one of its known phases, bismuth oxide remains in an incorrect structure. Any further calculations on those structures would be erroneous and could lead to both false positives and skewed property descriptors during both high-throughput and machine learning studies. While integrity checks could be made for some of the materials, the consistent breaking of the system's symmetry and the inconsistent degree of that symmetry breaking makes developing standardized checks impractical. This will be particularly useful for testing the predictions from crystal structure discovery methods, where exact knowledge of lattice type and decorations is necessary. Combining these constraints, with an automatically generated parametric representation, would provide an efficient means to optimize the newly predicted structures.

Bench-marking the algorithm

As the above examples show the implemented constraints not only ensure that symmetry is preserved but also accelerate the relaxation because the optimization of a reduced representation with less free parameters is by definition a less demanding task. To quantify this, the new relaxation algorithm is tested on a set of 359 materials across multiple lattice systems and AFLOW prototypes, as summarized in Table 3 and explicitly listed in the Supplementary Information (Supplementary Tables 1, 2). The chosen prototypes represent a sample of common materials with varied space groups and parametric relation complexity. To further understand how the method performs within a class, we try to include only prototypes with a significant amount of available structures. The initial geometry of each material is taken from either the AFLOW[55] or the Materials Project[54] database and converted into the right format using the Atomic Simulation Environment (ASE).[56] All relaxations are done in FHI-aims using the same settings as the zirconia calculations, using both the PBE and PBEsol functionals.

Table 4 and Fig. 3 illustrate the largest benefit of using the constraints: the preservation of the initial crystallographic prototypes. Approximately, 6% of all materials tested relax to a different

Table 3. Summary of materials test set.

AFLOW prototype	Space group	# of Materials	Atoms per unit cell	Free parameters	Full d.o.f./# free parameters
AB_oP8_62_c_c	62	8	8	7	4.71
A2B_oP12_62_2c_c	62	35	12	9	5.00
A2BC4_tI14_82_bc_a_g	82	35	7	5	6.00
A2BC4D_tI16_121_d_a_i_b	121	29	8	4	8.25
AB2_hP3_164_a_d	164	25	3	3	6.00
AB_hP4_186_b_b	186	37	4	2	5.25
AB_cF8_216_c_a	216	37	2	1	15.00
ABC_cF12_216_b_c_a	216	54	3	1	18.00
AB2_cF12_225_a_c	225	13	3	1	18.00
AB2C_cF16_225_a_c_b	225	14	4	1	21.00
AB_cF8_225_a_b	225	19	2	1	15.00
A_cF8_227_a	227	3	2	1	15.00
A2BC4_cF56_227_d_a_e	227	50	14	2	25.50

Summary of the materials used in the test data set

structure according to the AFLOW-XTAL-MATCH tool.[57] This tool measures the similarity between two materials using techniques similar to those of Burzlaff, et al.[58] and produces a misfit value, m,

$$m = 1 - (1 - \text{dev})(1 - \text{disp})(1 - \text{fail}), \qquad (3)$$

where dev, disp, and fail are normalized representations of deviations in the lattice vectors, atomic positions, and a failure indicator for when an atomic deviation is more than half the shortest distance in the coordination polyhedra. From m, we can determine if a structure is considered a match using the following mapping:

$$m \leq 0.1 : \text{structures are similar}$$
$$0.1 < m \leq 0.2 : \text{structures are within same family}$$
$$m > 0.2 : \text{structures are not compatible.}$$

All of the 26 materials with divergent relaxations in the data set follow the same pattern as cubic ZrO_2 and γ-Bi_2O_3. Therefore, constraining these material's relaxation is vital because without them all further calculations on these materials would no longer be physically relevant.

Constraining the relaxation has benefits even when the final structures are similar, as it significantly reduces the number of steps needed for the trajectories to converge. When the constrained and free relaxations proceed toward the same structures, the constraints reduce the number of relaxation steps by an average of 33.11% and 52.43% for calculations using the PBE and PBEsol functional, respectively. Including the divergent structures, increases the saving to 34.68% and 53.80%, but this is expected as the constrained and free relaxations are acting on qualitatively different PESs. Here we define the savings, S, as

$$S = \frac{N_{\text{free}} - N_{\text{constrained}}}{N_{\text{constrained}}} \times 100\% \qquad (4)$$

where N_{free} is the number of steps needed to converge the free relaxation and $N_{\text{constrained}}$ is the number of steps needed to converge the constrained relaxation. The seemingly better performance of the constraints when using PBEsol is likely because of larger differences between the initial and final structures when using this functional. On average, comparing the starting geometries with the freely relaxed ones gave an average m of 0.07654 (0.01630 when $m \neq 1$) and 0.1322 (0.02038 when $m \neq 1$) for the PBE and PBEsol calculations, respectively. Since increasing the distance to the final structure also increases

the likelihood of the free relaxation deviating from the constrained trajectory, larger savings using the PBEsol functional are expected. This is also supported by the larger differences between the final structures of constrained and free relaxations for the PBEsol functional.

Unfortunately, the savings are not consistent across the various prototypes. Figure 4 shows the total number of steps needed to relax the structures with and without constraints for the PBEsol calculations, sorted by the space group, and then the maximum number of relaxation steps needed. Similar to the average savings shown in Table 4, as the number of free parameters approaches the number of degrees of freedom in a material the savings from constraining the relaxation decrease. In some cases using, the constraints actually increases the number of steps needed for the relaxation to converge, but in all of these cases, for PBEsol calculations, the relaxation trajectories take unproductive steps shown in the inset of Fig. 4 for PtS_2. In these cases, the extra degrees of freedom allow the relaxation trajectories to pass the problematic regions faster and therefore converge to the final structures in fewer steps. For the PBE calculations, there are also some cases where the relaxation takes a few extra steps at the end of the trajectory where the total forces are near the convergence criteria. These results suggest that for lower symmetry structures, the potential savings from constraining the relaxation can be considerably smaller.

Constraining the relaxation can also decrease the computational time needed to perform further calculations after the relaxations. Despite their similarity to the final geometries of the constrained relaxation, freely relaxed structures always break symmetry to some degree. To measure how well the relaxation preserves symmetry, we compare the spglib calculated space groups of the initial and converged structures.[59] The package spglib calculates the space group of a material by iteratively searching for a given structure's primitive cell and symmetry operations and using those to generate the space group. The algorithm determines if a structure is preserved under a certain symmetry operation, by checking if the operator transforms all the atoms in the structure to sites occupied by the same type of atom within a given small euclidean distance, ε. By default, ε is set to 10^{-5} Å, and none of the tested materials remain in their initial space groups using this setting. To preserve the symmetry for the freely relaxed structures, a larger user-defined tolerance threshold is needed (0.01 Å is used in Table 4), consistent with findings reported by Hicks, et al.[52] However, when using the constraints, the symmetry is always perfectly preserved for all materials. This is

Table 4. Summary of constraint performance.

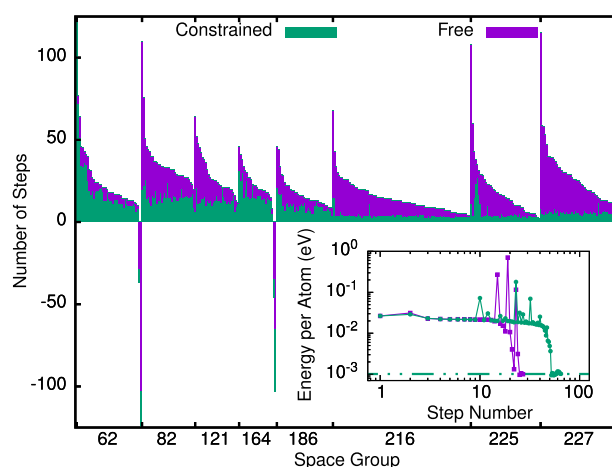| AFLOW prototype	Space group	# of materials	PBE			PBEsol		
			Average savings	# Preserved space group free	# XTAL match	Average savings	# Preserved space group free	# XTAL match
AB_oP8_62_c_c	62	8	10.23	3	8	24.61	4	8
A2B_oP12_62_2c_c	62	35	10.84	19	29	18.32	20	33
A2BC4_tI14_82_bc_a_g	82	35	40.32	29	32	59.98	28	34
A2BC4D_tI16_121_d_a_i_b	121	29	44.01	23	26	58.13	21	26
AB2_hP3_164_a_d	164	25	7.03	10	24	19.68	5	24
AB_hP4_186_b_b	186	37	30.34	23	36	41.68	19	36
AB_cF8_216_c_a	216	37	29.71	28	36	54.83	31	36
ABC_cF12_216_b_c_a	216	54	33.35	44	54	74.03	36	46
AB2_cF12_225_a_c	225	13	57.63	8	9	54.20	7	9
AB2C_cF16_225_a_c_b	225	14	19.77	12	14	80.02	7	12
AB_cF8_225_a_b	225	19	32.50	11	19	51.72	7	19
A_cF8_227_a	227	3	35.12	3	3	47.62	3	3
A2BC4_cF56_227_d_a_e	227	50	67.10	37	47	73.58	37	47
Full data set		359	34.68	69.64%	93.87%	53.80	62.67%	92.76%

Summary of the free and constrained relaxation performance by AFLOW prototype
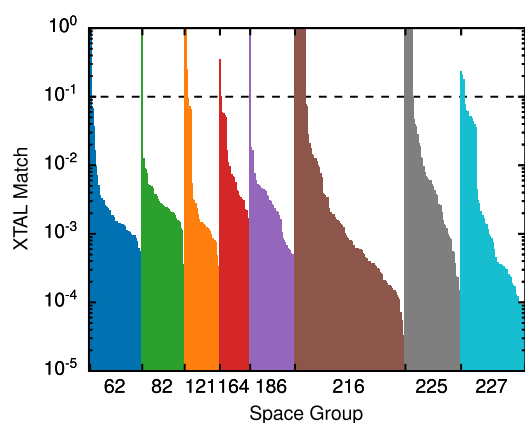
Fig. 3 Constrained and unconstrained relaxations go to the same structure. The misfit between the fully relaxed and constrained structures using AFLOW-XTAL-MATCH. All structures below the horizontal line are considered matching. The structures are ordered by space group and then by m.

Fig. 4 Constraints accelerate relaxation trajectories. The total number of steps needed for the constrained (green) and free (purple) relaxations. Negative step numbers represent the cases where the constrained relaxations take longer than the free relaxation. The bar for orthorhombic N_2O is not shown in the figure, since the free relaxation took 310 steps, and the constrained one required 296 steps to converge. The inset shows the constrained (green circles) and free relaxation (purple squares) trajectory of platinum (IV) sulfide, which is one of the materials where the constrained relaxation takes more steps than the free relaxation. The zero of the energy scale is set to 1 meV below the energy of the relaxed structure (horizontal dashed line).

advantageous as exploiting symmetry can greatly reduce the time needed to do further calculations for many applications. In particular, for phonon calculations using the finite difference approach, the symmetry of the system determines the number of atomic displacements, and therefore the number of force calculations, needed to calculate a complete set of force constants. The package *phonopy*, a very popular Python package for calculating phonon spectra with finite differences,[60] uses the default *spglib* settings to calculate the symmetry of a material, which would misrepresent all of the tested materials obtained via a free relaxation. Commensurate phonon calculations can be achieved by performing symmetry constrained relaxations and using more robust symmetry tools, such as AFLOW-SYM.[52]

Systems with local symmetries or distortions
One of the key advantages of defining constraints in this way is the ability to locally break symmetry to account for point defects. While the previously discussed benefits of lower relaxation times

and relaxing unstable structures could also be achieved by using symmetrized forces, that method would not be able to locally break a structure's symmetry. However, many of these defects exhibit some type of short range order, such as Jahn-Teller-type lattice distortions,[61–63] that can be parametrically added on top of the standard crystal structure. By including these distortions one can reduce the computational cost of relaxing supercells with defects, which is important when using a supercell approach to study point defects. This approach uses density-functional methods to calculate the total energy of a system in a series of

supercells of increasing size to study the effects of a defect on the system. If a scaling law is known for the system, the results of these calculations can be extrapolated to the experimentally relevant dilute limit. As the supercell size increases, the time needed to calculate each step also increases, therefore any reduction in the number of steps needed can save a significant amount of computational time.

To illustrate the ability of the new relaxation scheme to study point defects in materials, we study a polaronic distortion in MgO previously studied in our department at the Fritz Haber Institute.[64] Polarons are quasiparticles that couple point charges with lattice distortions in a material, reducing its charge-carrier mobility. Typically, the size of the polaron is controlled by the electron–phonon interaction, with a stronger interaction leading to a smaller polaron.[64,65] Because of the reduced charge-carrier mobility, understanding how polarons form and migrate through a material is vital for a number of applications ranging from catalysis[66–68] to thermoelectricity.[65] For rock-salt MgO, the lattice distortions for an electron hole polaron are Jahn-Teller like, with the oxygen and magnesium ions being, respectively, attracted and repelled from the hole. To model this system, we assume the electron hole is located on a fixed oxygen atom placed in the center of the supercell. We then allow the other ions to relax from their initial position along a line going through the center of the cell, with the magnitude and sign of the motion being controlled by a single free parameter for each atom. This constraint choice reduces the number of degrees of freedom to $N - 1$, where N is number of atoms in the supercell. While these constraints map all other interactions onto the main Coulombic distortion, we expect those interactions to have a minor effect on the final structure. Moreover, the restrictiveness of the constraints can be decreased further on subsequent calculations until the desired level of accuracy is achieved.

Figure 5 illustrates the uncorrected polaron energy for the free and constrained relaxation trajectories for these supercells.[64] Because charge localization is necessary when studying polarons, the HSE 06 functional is used with a screening parameter of 0.11 Bohr^{-1} with a fraction of exact exchange, α, of one. For the unrestricted spin relaxation, the SCF density, forces, total energy, and eigenvalues were all converged to 10^{-4} eV/Å, 10^{-4} eV/Å, 10^{-5} eV, and 10^{-2} eV, respectively. Five empty states per atom are used, and the structures were relaxed until the total forces on the free parameters were below 10^{-4} eV/Å. The constrained relaxation for the distorted geometry converges the 64-atom supercell in 11 steps, which is only one-eighth of the steps needed by the free relaxation. Significantly better performance is seen for the 216-atom supercell, which converges in 10 constrained relaxation steps, 96% less than in the unconstrained case needing 234 steps. This increase in efficiency is likely a result of the constraints focusing on the primary Coulombic interaction, reducing the amount of fine-tuning it needs to do to balance out the weaker perturbation in the other interionic interactions. Using the constraints does cause the relaxation to converge to a structure with slightly higher energy, but this is because the chosen constraints approximate an isolated polaron and thus do not account for the artificial interactions with periodic images due to the finite supercell size. In practice, this effect is rather negligible, since we find that the obtained polaron energy is 78.4 meV higher in energy than the free relaxation in the $2 \times 2 \times 2$ supercell and 69.1 meV higher in the $3 \times 3 \times 3$ supercell.

DISCUSSION

In this work, we presented a new scheme for parametrically relaxing structures in a general, symmetry-reduced space. After explaining the algorithm, we test it on 359 different materials across a broad range of material classes. In all cases, the new method was able to strictly preserve the symmetry of the materials, and on average reduced the number of steps needed to converge a material by 50%. We also demonstrated for the example of bismuth oxide how the constraints can be used to relax to metastable phases. Finally, we showcased the relaxation of structures with local symmetry breaking with known distortion patterns for polarons in MgO.

This new method will have a profound impact on computational materials discovery. Not only does the decreased cost of relaxing a material increase the velocity of high-throughput searches but it also allows for those searches to explore metastable and dynamically stabilized structures. The method also has the promise to improve the efficiency of supercell calculations and to study only physically relevant structures. Finally, by monitoring the difference between the full forces and symmetrized forces new stable phases can potentially be discovered from metastable or unstable polymorphs. Although we showed that the proposed algorithm is applicable to accelerate and improve standard solid-state physics calculations, its flexibility allows it to be applied to a much wider range of problems, e.g., transition-state searches or interface relaxations. Similarly, it is easily generalizable to other forms of coordinates and straightforwardly implementable in any electronic-structure theory code, and has already been implemented as constraints with ASE by the authors.

METHODS

Implementation details

In practice, the parametric constraints are implemented in the following fashion. Let us assume a (3×3)-dimensional lattice vector matrix, $\mathcal{L}$, and a ($N \times 3$)-dimensional matrix, $\mathcal{R_F}$, for the fractional atomic positions. Given the atomic forces, $\mathcal{F_R}$, on the Cartesian atomic positions and the stress tensor σ, we can calculate the derivatives of the energy with respect to the lattice components[69]

$$\frac{dE}{d\mathcal{L}} = \mathcal{L}^{T^{-1}} V \cdot \sigma, \qquad (5)$$

where V is the unit-cell volume and obtain the generalized forces on the
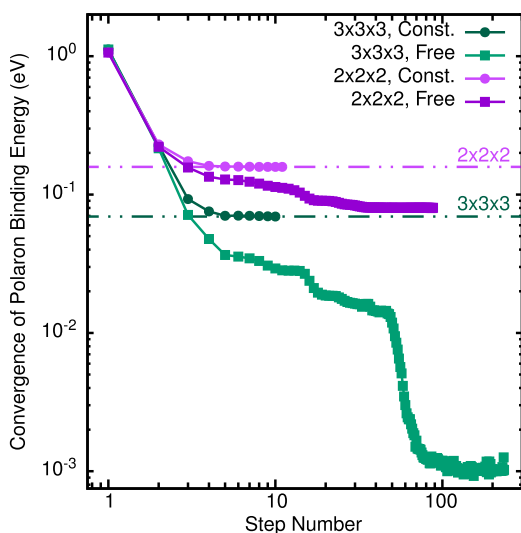
Fig. 5 Constraints quickly find distorted geometries. The convergence of the free (squares) and constrained (circles) relaxation of a polaronic distortion in a $2 \times 2 \times 2$ (purple) and $3 \times 3 \times 3$ (green) supercell of rock-salt MgO. The zero of the energy scale is set to be 1 meV below the energy of the minimal energy structure. The horizontal dashed lines correspond to the final energy of constrained relaxation.
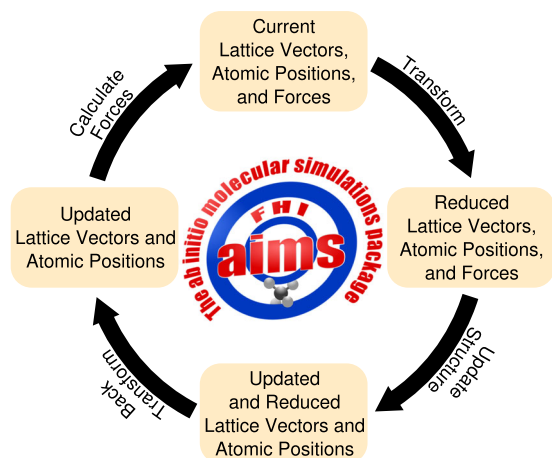
Fig. 6 The Parameteric constraints workflow. Workflow of the relaxation constrained to the "parameter reduced space".

lattice, $\mathcal{F}_{\mathcal{L}}$, after cleaning from the atomic contributions

$$\mathcal{F}_{\mathcal{L}} = -\frac{dE}{d\mathcal{L}} - \mathcal{R}_{\mathcal{F}}^T \, \mathcal{F}_{\mathcal{R}}. \tag{6}$$

Each of these matrices denoted by calligraphic letters, $\mathcal{R}_{\mathcal{F}}$, $\mathcal{F}_{\mathcal{R}}$, $\mathcal{L}$, and $\mathcal{F}_{\mathcal{L}}$, can be flattened to one-dimensional vectors that we will name R_F, F_R, L, and F_L, respectively. In the parameter representation, these quantities reduce to their small-letter counterparts, the M_R-dimensional r and F_r and the M_L-dimensional l and F_l, via

$$r = \mathcal{J}_{Rf}^{-1} \, (R_F - t_{Rf}), \tag{7a}$$

$$l = \mathcal{J}_L^{-1} \, (L - t_L), \tag{7b}$$

$$F_r = \mathcal{J}_R^T \, F_R, \tag{7c}$$

$$F_l = \mathcal{J}_L^T \, F_L, \tag{7d}$$

where M_R and M_L are the number of free parameters in the atomic and lattice degrees of freedom; $\mathcal{J}_R$, $\mathcal{J}_{Rf}$, and $\mathcal{J}_L$ are the Jacobian matrix for the transformations, and t_{Rf} and t_L are the translation vectors for the respective fractional atomic and lattice degrees of freedom. Here, $\mathcal{J}_R$ represents the transformation of the atomic coordinates from Cartesian space to the reduced space, which is calculated from $\mathcal{J}_{Rf}$ by

$$\mathcal{J}_R = \begin{pmatrix} \mathcal{L}^T & 0 & \dots & 0 \\ 0 & \mathcal{L}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathcal{L}^T \end{pmatrix} \mathcal{J}_{Rf}. \tag{8}$$

The translation vectors are used to include any constant shifts, which are not captured by the Jacobians. Because $\mathcal{J}_R$, $\mathcal{J}_{Rf}$, and $\mathcal{J}_L$ are not square and therefore not regularly invertible, we use the generalized left inverse[70] defined for a matrix $\mathcal{A}$ as

$$\mathcal{A}^{-1,L} = (\mathcal{A}^T \mathcal{A})^{-1} \, \mathcal{A}^T, \tag{9}$$

provided $\mathcal{A}$ has full column rank. The transformation back to real space can then be performed by inverting Eqs (7a)–(7d). The back-transformation of the forces to the full space is not necessary but can be helpful to obtain symmetrized Cartesian or Fractional forces, to check for the convergence of the relaxation.

To facilitate the construction of the Jacobian matrices, we assume a linear relationship between the full coordinates and the parameters. In principle, $\mathcal{J}_{Rf}$, and $\mathcal{J}_L$ can be constructed at each step by using analytical expressions to describe each real space degree of freedom as a function of the reduced parameter set; however, by assuming a linear relationship between the spaces they can be initialized at the start of the calculation and used at every step. For the atomic positions, this assumption is already fulfilled by using fractional instead of Cartesian coordinates. If we allow angles as unit-cell parameters, which is the case for the monoclinic and triclinic lattice systems, the relations become nonlinear containing for

example expressions like $c \cdot \cos(\beta)$. In these cases, the easiest solution is to substitute each nonzero lattice vector component with an independent parameter.

Before the relaxation, the Hessian, $\mathcal{H}$, is initialized in the full coordinate space, split into atomic and lattice blocks ($\mathcal{H}_R$ and $\mathcal{H}_L$, respectively), and individually transformed into reduced coordinate space, $\mathcal{H}_r$ and $\mathcal{H}_l$ via separate Jacobians, $\mathcal{J}_R$ and $\mathcal{J}_L$

$$\mathcal{H}_r = \mathcal{J}_R^T \mathcal{H}_R \mathcal{J}_R, \tag{10}$$

$$\mathcal{H}_l = \mathcal{J}_L^T \mathcal{H}_L \mathcal{J}_L. \tag{11}$$

Here, $\mathcal{J}_R$ is also divided by the average unit vector length, $V^{1/3}$, so $\mathcal{H}_r$ and $\mathcal{H}_l$ are on a similar scale. The total Hessian is then recombined resulting in

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_r & 0 \\ 0 & \mathcal{H}_l \end{pmatrix}. \tag{12}$$

Figure 6 illustrates the procedure for relaxing structures with these constraints. During the relaxation, a full SCF cycle is completed to obtain the forces and the stress tensor for the current geometry, at each step. If the convergence criterion is fulfilled, i.e., if the forces are below a given threshold, then the relaxation stops and returns the current geometry. Otherwise the lattice vectors as well as the atomic positions and their respective forces are mapped onto the reduced space using the transformation described in Eqs. (7a–d). The atomic coordinates and forces are, respectively, scaled by $V^{1/3}$ and $V^{-1/3}$, and then passed on to the optimizer. In FHI-aims this is usually a BFGS/modified BFGS optimizer. Once the optimized parameters are obtained, the full geometry is reconstructed from the parameters and a new relaxation step can begin.

DATA AVAILABILITY
The full input and output files for the calculations in this work are available in the NOMAD repository with the identifier [https://doi.org/10.17172/NOMAD/2019.10.19-1].[71]

CODE AVAILABILITY
The algorithms presented in this paper have been fully implemented in FHI-aims and is available in any version after 20191019. The constraints have also been implemented in ASE https://gitlab.com/ase/ase/merge_requests/1298.

REFERENCES
1. Panda, P. K. Review: environmental friendly lead-free piezoelectric materials. *J. Mater. Sci.* **44**, 5049–5062 (2009).
2. Wang, Z. L. et al. Progress in piezotronics and piezo-phototronics. *Adv. Mater.* **24**, 4632–4646 (2012).
3. Ok, K. M. Toward the rational design of novel noncentrosymmetric materials: factors influencing the framework structures. *Acc. Chem. Res.* **49**, 2774–2785 (2016).
4. Nguyen, T. D., Mao, S., Yeh, Y. W., Purohit, P. K. & McAlpine, M. C. Nanoscale flexoelectricity. *Adv. Mater.* **25**, 946–974 (2013).
5. Moure, C. & Peña, O. Recent advances in perovskites: processing and properties. *Prog. Solid State Chem.* **43**, 123–148 (2015).
6. Isaenko, L. I., Yelisseyev, A. P., Lobanov, S. I., Krinitsin, P. G. & Molokeev, M. S. Structure and optical properties of Li2Ga2GeS6 nonlinear crystal. *Optical Mater.* **47**, 413–419 (2015).
7. Damjanovic, D. Ferroelectric, dielectric and piezoelectric properties of ferroelectric thin films and ceramics. *Rep. Prog. Phys.* **61**, 1267 (1998).
8. Hang, T., Zhang, W., Ye, H. Y. & Xiong, R. G. Metal-organic complex ferroelectrics. *Chem. Soc. Rev.* **40**, 3577–3598 (2011).
9. Sun, E. & Cao, W. Relaxor-based ferroelectric single crystals: growth, domain engineering, characterization and applications. *Prog. Mater. Sci.* **65**, 124–210 (2014).
10. Shi, P. P. et al. Symmetry breaking in molecular ferroelectrics. *Chem. Soc. Rev.* **45**, 3811–3827 (2016).
11. Cava, R. J., Ji, H., Fuccillo, M. K., Gibson, Q. D. & Hor, Y. S. Crystal structure and chemistry of topological insulators. *J. Mater. Chem. C.* **1**, 3176–3189 (2013).
12. Fiete, G. A. et al. Topological insulators: crystalline protection. *Nat. Mater.* **11**, 1003–1004 (2012).
13. Evans, O. R. & Lin, W. Crystal engineering of NLO materials based on metal-organic coordination networks. *Acc. Chem. Res.* **35**, 511–522 (2002).

14. Wu, K. & Pan, S. A review on structure-performance relationship toward the optimal design of infrared nonlinear optical materials with balanced performances. *Coord. Chem. Rev.* **377**, 191–208 (2018).

15. Xu, J. et al. Halide perovskites for nonlinear optics. *Adv. Mate.* 1806736 https://doi.org/10.1002/adma.201806736 (2019).

16. Le, M. Q. & Nguyen, D. T. The role of defects in the tensile properties of silicene. *Appl. Phys. A: Mater. Sci. Process.* **118**, 1437–1445 (2014).

17. Sayle, T. X. et al. Mechanical properties of ceria nanorods and nanochains; the effect of dislocations, grain-boundaries and oriented attachment. *Nanoscale* **3**, 1823–1837 (2011).

18. Pan, X., Yang, M. Q., Fu, X., Zhang, N. & Xu, Y. J. Defective TiO2 with oxygen vacancies: synthesis, properties and photocatalytic applications. *Nanoscale* **5**, 3601–3614 (2013).

19. Wang, F. et al. Intrinsic "vacancy Point Defect" induced electro-chemiluminescence from coreless supertetrahedral chalcogenide nanocluster. *J. Am. Chem. Soc.* **138**, 7718–7724 (2016).

20. An, T. H. et al. Point defect-assisted doping mechanism and related thermoelectric transport properties in Pb-doped BiCuOTe. *J. Mater. Chem. A* **2**, 19759–19764 (2014).

21. Anno, H., Matsubara, K., Notohara, Y., Sakakibara, T. & Tashiro, H. Effects of doping on the transport properties of CoSb3. *J. Appl. Phys.* **86**, 3780–3786 (1999).

22. Senanayak, S. P. et al. Understanding charge transport in lead iodide perovskite thin-film field-effect transistors. *Sci. Adv.* **3**, e1601935 (2017).

23. Proshchenko, V. S., Dholabhai, P. P., Sterling, T. C. & Neogi, S. Heat and charge transport in bulk semiconductors with interstitial defects. *Phys. Rev. B* **99**, 014207 (2019).

24. Toberer, E. S., May, A. F. & Snyder, G. J. Zintl chemistry for designing high efficiency thermoelectric materials. *Chem. Mater.* **22**, 624–634 (2010).

25. Xie, H. et al. Beneficial contribution of alloy disorder to electron and phonon transport in half-heusler thermoelectric materials. *Adv. Funct. Mater.* **23**, 5123–5130 (2013).

26. Wang, H. et al. Stabilization of highly polar BiFeO3-like structure: a new interface design route for enhanced ferroelectricity in artificial perovskite superlattices. *Phys. Rev. X* **6**, 011027 (2016).

27. Bousquet, E. et al. Improper ferroelectricity in perovskite oxide artificial superlattices. *Nature* **452**, 732–736 (2008).

28. Ricci, F. et al. Data descriptor: an ab initio electronic transport database for inorganic materials. *Sci. Data* **4**, 170085 (2017).

29. Broberg, D. et al. PyCDT: a Python toolkit for modeling point defects in semiconductors and insulators. *Computer Phys. Commun.* **226**, 165–179 (2018).

30. Plata, J. J. et al. An efficient and accurate framework for calculating lattice thermal conductivity of solids: AFLOW-AAPL Automatic Anharmonic Phonon Library. *npj Comput. Mater.* **3**, 45 (2017).

31. Tang, F., Po, H. C., Vishwanath, A. & Wan, X. Comprehensive search for topological materials using symmetry indicators. *Nature* **566**, 486–489 (2019).

32. Maghirang, A. B. et al. Predicting two-dimensional topological phases in Janus materials by substitutional doping in transition metal dichalcogenide monolayers. *npj 2D Mater. Appl.* **3**, 35 (2019).

33. Roy, A., Bennett, J. W., Rabe, K. M. & Vanderbilt, D. Half-Heusler semiconductors as piezoelectrics. *Phys. Rev. Lett.* **109**, 037602 (2012).

34. Mazaheri, T. et al. Stochastic replica voting machine prediction of stable cubic and double perovskite materials and binary alloys. *Phys. Rev. Mater.* **3**, 063802 (2019).

35. Sutton, C. et al. Crowd-sourcing materials-science challenges with the NOMAD 2018 Kaggle competition. *npj Computational Materials*, **5**, 1–11 (2019)

36. Mehl, M. J. et al. The AFLOW library of crystallographic prototypes: part 1. *Comput. Mater. Sci.* **136**, S1–S828 (2017).

37. Hicks, D. et al. The AFLOW library of crystallographic prototypes: part 2. *Comput. Mater. Sci.* **161**, S1–S1011 (2019).

38. Hafner, J. et al. Ab-initio simulations of materials using VASP: density-functional theory and beyond. *J. Comput. Chem.* **29**, 2044–2078 (2008).

39. Kronik, L. et al. PARSEC—the pseudopotential algorithm for real-space electronic structure calculations: Recent advances and novel applications to nanostructures. *Phys. Status Solidi (B) Basic Res.* **243**, 1063–1079 (2006).

40. Fletcher, R. R. *Practical Methods of Optimization* (Wiley, 1987).

41. Nocedal, J. & Wright, S. J. *Numerical Optimization*, 2nd edn. (Springer, New York, 2006).

42. Freysoldt, C. et al. On-the-fly parameterization of internal coordinate force constants for quasi-Newton geometry optimization in atomistic calculations. *Comput. Mater. Sci.* **133**, 71–81 (2017).

43. Panosetti, C., Krautgasser, K., Palagin, D., Reuter, K. & Maurer, R. J. Global materials structure search with chemically motivated coordinates. *Nano Lett.* **15**, 8044–8048 (2015).

44. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **21**, 395502 (2009).

45. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Phys. Commun.* **180**, 2175–2196 (2009).

46. Knuth, F., Carbogno, C., Atalla, V., Blum, V. & Scheffler, M. All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals. *Computer Phys. Commun.* **190**, 33–50 (2015).

47. Bocanegra-Bernal, M. H. & de la Torre, S. D. et al. Phase transitions in zirconium dioxide and related materials for high performance engineering ceramics. *J. Mater. Sci.* **37**, 4947–4971 (2002).

48. Carbogno, C., Levi, C. G., Van De Walle, C. G. & Scheffler, M. Ferroelastic switching of doped zirconia: Modeling and understanding from first principles. *Phys. Rev. B - Condens. Matter Mater. Phys.* **90**, 144109 (2014).

49. Parlinski, K., Li, Z. Q. & Kawazoe, Y. First-principles determination of the soft mode in cubic ZrO2. *Phys. Rev. Lett.* **78**, 4063–4066 (1997).

50. Matsumoto, A., Koyama, Y. & Tanaka, I. Structures and energetics of Bi2O2 polymorphs in a defective fluorite family derived by systematic first-principles lattice dynamics calculations. *Phys. Rev. B* **81**, 94117 (2010).

51. Drache, M., Roussel, P. & Wignacourt, J.-P. Structures and oxide mobility in Bi-Ln-O materials: heritage of Bi2O2. *Chem. Rev.* **107**, 80–96 (2007).

52. Hicks, D. et al. AFLOW-SYM: platform for the complete, automatic and self-consistent symmetry analysis of crystals. *Acta Crystallogr. Sect. A Found. Adv.* **74**, 184–203 (2018).

53. Gonze, X. et al. First-principles computation of material properties: the ABINIT software project. *Computational Mater. Sci.* **25**, 478–492 (2002).

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

55. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).

56. Hjorth Larsen, A. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).

57. Toher, C. et al. The AFLOW Fleet for Materials Discovery. In: Andreoni, W., Yip, S. (eds), *Handbook of Materials Modeling.* 1–28 Springer, Cham (2018).

58. Burzlaff, H. & Malinovsky, Y. A procedure for the clasification of non-organic crystal structures. i. theoretical background. *Acta Crystallogr. Sect. A Found. Crystallogr.* **53**, 217–224 (1997).

59. Togo, A. & Tanaka, I. *spglib*: a software library for crystal symmetry search. Preprint at: http://arxiv.org/abs/1808.01590 (2018).

60. Togo, A. & Tanaka, I. et al. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).

61. Rodríguez, R. A. et al. Structural defects in LiMn2O4 induced by gamma radiation and its influence on the Jahn-Teller effect. *Solid State Ion.* **324**, 77–86 (2018).

62. Prentice, J. C., Monserrat, B. & Needs, R. J. First-principles study of the dynamic Jahn-Teller distortion of the neutral vacancy in diamond. *Phys. Rev. B* **95**, 14108 (2017).

63. Evarestov, R. et al. Jahn-Teller effect in the phonon properties of defective SrTiO 3 from first principles. *Phys. Rev. B - Condens. Matter Mater. Phys.* **85**, 174303 (2012).

64. Kokott, S., Levchenko, S. V., Rinke, P. & Scheffler, M. First-principles supercell calculations of small polarons with proper account for long-range polarization effects. *N. J. Phys.* **20**, 033023 (2018).

65. Nag, A. & Shubha, V. Oxide thermoelectric materials: a structure-property relationship. *J. Electron. Mater.* **43**, 962–977 (2014).

66. Dohnálek, Z., Lyubinetsky, I. & Rousseau, R. et al. Thermally-driven processes on rutile TiO2(1 1 0)-(1 × 1): a direct view at the atomic scale. *Prog. Surf. Sci.* **85**, 161–205 (2010).

67. Henderson, M. A. A surface science perspective on TiO2 photocatalysis. *Surf. Sci. Rep.* **66**, 185–297 (2011).

68. Rettie, A. J., Chemelewski, W. D., Emin, D. & Mullins, C. B. Unravelling small-polaron transport in metal oxide photoelectrodes. *J. Phys. Chem. Lett.* **7**, 471–479 (2016).

69. Doll, K. et al. Analytical stress tensor and pressure calculations with the CRYSTAL code. *Mol. Phys.* **108**, 223–227 (2010).

70. Rao, C. R. & Mitra, S. K. Generalized inverse of a matrix and its applications. In *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*. 601–620 (University of California Press, Berkeley, Calif, 1972) https://projecteuclid.org/euclid.bsmsp/1200514113.

71. Lenz, M.-O. et al. Constrained geometry relaxations: dataset. at https://doi.org/10.17172/NOMAD/2019.10.19-1 (2019).

AUTHOR CONTRIBUTIONS

M.O.L. and T.A.R.P. contributed equally on this work by implementing and bench-marking the discussed relaxation scheme. D.H. worked to incorporate the constraints into AFLOW. S.C., M.S. and C.C. directed the project. All authors analyzed the data and wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41524-019-0254-4.

Correspondence and requests for materials should be addressed to T.A.R.P.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.