Efficient generation of generalized Monkhorst-Pack grids through the use of informatics

Pandu Wisesa, Kyle A. McGill, and Tim Mueller

Phys. Rev. B **93**, 155109 — Published 6 April 2016

DOI: 10.1103/PhysRevB.93.155109

Efficient generation of generalized Monkhorst-Pack grids through the use of informatics.

Pandu Wisesa, Kyle A. McGill, and Tim Mueller[+]

Department of Materials Science and Engineering, Johns Hopkins University, Baltimore Maryland 21218, USA

Manuscript Submitted:

Manuscript Accepted:

Abstract: We present a method for rapidly generating efficient *k*-point grids for Brillouin zone integration by using a database of pre-calculated grids. Benchmark results on 102 randomly-selected materials indicate that for well-converged calculations, the grids generated by our method have less than half as many irreducible *k*-points as Monkhorst-Pack grids generated using a more conventional method, significantly accelerating the calculation of properties of crystalline materials.

I. INTRODUCTION

The calculation of many properties of crystalline materials requires the evaluation of integrals over the Brillouin zone in reciprocal space. These integrals are typically approximated using a discrete set of points, commonly known as *k*-points. Increasing the density of *k*-points in the Brillouin zone can increase the accuracy of the calculation, but the cost of approximating the integral typically scales linearly with the number of symmetrically irreducible *k*-points (i.e. the largest subset of *k*-points for which no two *k*-points in the subset are symmetrically equivalent). To minimize the cost of Brillouin

zone integration while maintaining sufficient accuracy, several methods have been developed for the selection of "special" points for Brillouin zone integration. This idea was introduced by Baldereschi [1], expanded by Chadi and Cohen [2], and further developed by Monkhorst and Pack [3], with the latter approach being the most widely used today.

In the method of Monkhorst and Pack, a regular grid of *k*-points is generated, and the Brillouin zone integral of a function is approximated by calculating the average value of the function over the *k*-points. The speed and accuracy of the approximation may be improved by shifting the grid so that no point falls on the high-symmetry Γ point at the center of the Brillouin zone. The axes of the grid align with the reciprocal lattice vectors, so the coordinates of a *k*-point on an $m_1 \times m_2 \times m_3$ grid are given by:

$$\mathbf{k} = \frac{n_1}{m_1}\mathbf{b}_1 + \frac{n_2}{m_2}\mathbf{b}_2 + \frac{n_3}{m_3}\mathbf{b}_3 + \mathbf{s} \tag{1}$$

where $\mathbf{s}$ is a vector representing the shift from the Γ point and n_1, n_2, and n_3 are integers that range from 1 to m_1, m_2, and m_3 respectively. The reciprocal lattice vectors, $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$, are defined by:

$$\begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3\end{pmatrix}^T = 2\pi\begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{pmatrix}^{-1} \tag{2}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are the lattice vectors for a primitive unit cell in real space (here we represent vectors as column vectors).

Monkhorst-Pack grids have a useful real-space interpretation. We can define an $m_1 \times m_2 \times m_3$ supercell of the primitive cell with lattice vectors $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$ given by

$$\begin{pmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3\end{pmatrix} = \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{pmatrix}\mathbf{M} \tag{3}$$

where $\mathbf{M}$ is a diagonal matrix in which $M_{11} = m_1$, $M_{22} = m_2$, and $M_{33} = m_3$. We will refer to the Bravais lattice with lattice vectors $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$ as the "superlattice". Bloch waves with wave vector $\mathbf{s}$ with respect to the supercell take the form:

$$\psi(\mathbf{r}) = u_{\text{super}}(\mathbf{r}) e^{i\mathbf{s}\cdot\mathbf{r}} \tag{4}$$

where $u_{\text{super}}(\mathbf{r})$ is a function with the periodicity of the real-space supercell. We can express $u_{\text{super}}(\mathbf{r})$ as a Bloch wave with respect to the *primitive* cell:

$$u_{\text{super}}(\mathbf{r}) = u_{\text{primitive}}(\mathbf{r}) e^{i\mathbf{g}\cdot\mathbf{r}} \tag{5}$$

where $u_{\text{primitive}}(\mathbf{r})$ is a function with the periodicity of the real-space primitive cell and $\mathbf{g}$ is a vector in reciprocal space. Because $u_{\text{super}}(\mathbf{r})$ has the periodicity of the real-space supercell, $\mathbf{g}$ must satisfy

$$\mathbf{g}^{\mathbf{T}} \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \end{pmatrix} = 2\pi \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix} \tag{6}$$

where n_1, n_2, and n_3 are integers. Combining equations (3) and (6), we get:

$$\mathbf{g}^{\mathbf{T}} \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{pmatrix} \mathbf{M} = 2\pi \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix}. \tag{7}$$

Solving for $\mathbf{g}$ and combining with equation (2) yields

$$\mathbf{g}^{\mathbf{T}} = \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix} \mathbf{M}^{-1} \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{pmatrix}^{\mathbf{T}}. \tag{8}$$

Combining equations (4), (5), and (8), we get an alternative expression for $\psi(\mathbf{r})$:

$$\psi(\mathbf{r}) = u_{\text{primitive}}(\mathbf{r}) e^{i\left(\frac{n_1}{m_1}\mathbf{b}_1 + \frac{n_2}{m_2}\mathbf{b}_2 + \frac{n_3}{m_3}\mathbf{b}_3 + \mathbf{s}\right)\cdot\mathbf{r}}. \tag{9}$$

Equations (1), (4), and (9) demonstrate that each Bloch wave on a Monkhorst-Pack grid with respect to the primitive cell is equivalent to a Bloch wave at wave vector $\mathbf{s}$ with respect to the supercell.

The real-space supercell interpretation outlined above is instructive because upon inspection, it becomes apparent that there is no reason the matrix $\mathbf{M}$ must be diagonal. Any non-singular integer

matrix $\mathbf{M}$ can be used to define the real-space supercell, and the coordinates of the generated *k*-points are then given by

$$\mathbf{k} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{pmatrix} \left(\mathbf{M}^{-1} \right)^{\mathrm{T}} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} + \mathbf{s} \tag{10}$$

where n_1, n_2, and n_3 are integers. We will refer to *k*-point grids described by equation (10) as generalized Monkhorst-Pack grids. In these grids, the total number of *k*-points in the Brillouin zone is equal to the number of primitive cells in the real-space supercell.

Examples of generalized Monkhorst-Pack grids on a two-dimensional rectangular lattice are shown in Fig. 1. The green 2×2 supercell in real space corresponds to a 2×2 *k*-point grid in reciprocal space. This grid is illustrative of the traditional approach for constructing *k*-point grids, in which the matrix $\mathbf{M}$ is diagonal. Removing the constraint that the matrix $\mathbf{M}$ must be diagonal enables the construction of the red and blue supercells and corresponding generalized *k*-point grids. The red and blue grids contain 3 and 2 *k*-points per reciprocal space unit cell respectively, with *k*-points that are more evenly spaced than would be achieved by 3×1 or 2×1 (yellow) traditional Monkhorst-Pack grids.

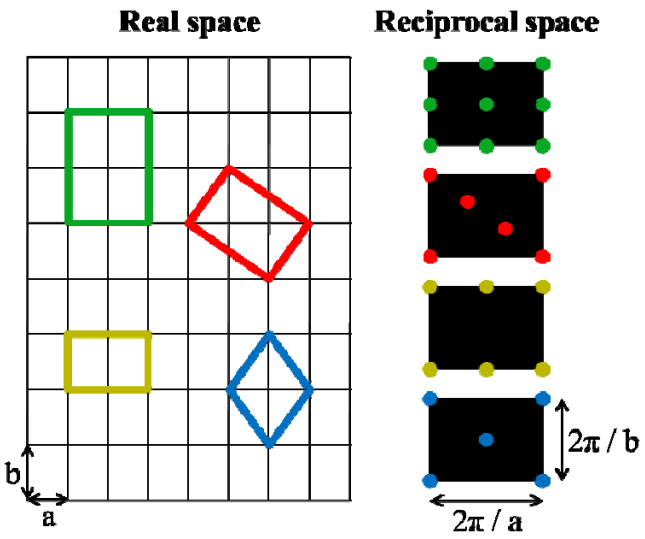

FIG. 1. A comparison of (left) real-space supercells on a rectangle lattice and (right) the corresponding *k*-point grids in a unit cell of the reciprocal lattice. To aid visual comparison, all *k*-point grids are shifted so that one *k*-point falls on the corner of the reciprocal space unit cell. The yellow grid is an example of a traditional 2x1 Monkhorst-Pack grid, and the blue grid is a generalized Monkhorst Pack grid with the same density that more evenly samples reciprocal space.

The benefit of using generalized grids was identified by Froyen [4], who suggested choosing the real-space supercell and shift vector in a way that minimizes the number of symmetrically irreducible *k*-points in the grid. Moreno and Soler demonstrated that generalized grids, as defined in equation (10), can always be expressed in terms of a diagonal matrix $\mathbf{M}$, provided the reciprocal lattice vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ are suitably chosen [5]. They expanded on Froyen's work by proposing the following approach for selecting the optimal *k*-point grid:

1) Select a minimum permissible distance between lattice points in the real-space superlattice. We will call this distance $r_{\min}$.

2) Of the possible superlattices in which all lattice points are separated by at least a distance of $r_{\min}$, find the one that corresponds to a *k*-point grid with the fewest symmetrically irreducible *k*-points.

There is intuitive appeal to the idea selecting *k*-point grids based on the minimum distance between points on the real-space superlattice. The density of the grid, and hence the speed and accuracy of the calculation, is determined by a single parameter, r_{min}. The grid will also naturally be chosen so that the *k*-points are evenly distributed in reciprocal space. Specifically, this approach favors fcc-like real-space superlattices, resulting in bcc-like *k*-point grids.

Despite the apparent advantages of the *k*-point grid generation approach advocated by Moreno and Soler, it has seen little use in practice. It has been partially implemented in the SIESTA software package [6], but the partial implementation does not identify the grid with the fewest irreducible *k*-points. The relatively poor adoption of this approach, compared to the much more popular Monkhorst-Pack approach, is likely due to the fact that the Moreno-Soler approach mandates a search through all possible superlattices in which lattice points are separated by a distance of at least r_{min} to identify the one that results in the fewest number of irreducible *k*-points. This is significantly more complicated and computationally expensive than the relatively simple task of generating an $m_1 \times m_2 \times m_3$ Monkhorst-Pack grid.

In this paper we demonstrate that the complexity and cost of the Moreno-Soler approach can be significantly reduced through the use of informatics. We have created a database of generalized *k*-point grids that can be rapidly searched to find the grid with the fewest irreducible *k*-points for which the points on the corresponding superlattice are separated by a distance of at least r_{min}. Based on benchmark calculations on 102 randomly-selected crystalline materials, we find that our approach has the potential to significantly accelerate quantum calculations on crystalline materials.

II. METHODOLOGY

A. Real-space interpretation of the integration error

We start by briefly providing an additional argument for the Moreno-Soler approach, in which *k*-point grids are selected according to $r_{\min}$. Consider a Hamiltonian operator H with eigenvalues $\varepsilon_n(\mathbf{k})$ corresponding to eigenfunctions $\psi_{n\mathbf{k}}(\mathbf{r})$, where n is the band index. The total energy at each *k*-point is given by

$$E(\mathbf{k}) = \sum_n \varepsilon_n(\mathbf{k}) f(\varepsilon) . \tag{12}$$

Where $f(\varepsilon)$ is an occupancy function. Let Δ represent the average integration error due to using a generalized Monkhorst-Pack *k*-point grid to approximate the integral of $E(\mathbf{k})$ across the Brillouin zone:

$$\Delta = \frac{1}{N_k} \sum_{\mathbf{k}} E(\mathbf{k}) - \frac{1}{\Omega} \int_{BZ} E(\mathbf{k}) d\mathbf{k} , \tag{13}$$

where Ω is the volume of the primitive cell Brillouin zone and the sum is over all N_k *k*-points in the Brillouin zone. Following Monkhorst and Pack [3],

$$\Delta = \sum_{m>1} f_m e^{i\mathbf{s}\cdot\mathbf{R}_m} , \tag{14}$$

where the sum is over all points $\mathbf{R}_m$ on the supercell lattice (excluding the lattice point $\mathbf{R}_1$ at the origin) and

$$f_m = \frac{1}{\Omega} \int_{BZ} E(\mathbf{k}) e^{-i\mathbf{R_m}\cdot\mathbf{k}} d\mathbf{k} . \tag{15}$$

A derivation of this result for generalized Monkhorst-Pack grids is provided in the supplemental material [7]. Thus the approximation error for a *k*-point grid that includes the Γ point (i.e. **s** has zero length) is:

$$\Delta = \sum_{m>1} f_m \,. \tag{16}$$

It is often desirable to shift the *k*-point grid by a half-multiple of one or more reciprocal lattice vectors of the superlattice. In this case, by equation (14), half of the terms in the sum in equation (16) are multiplied by -1.

To estimate the magnitude of f_m, we first consider the case of an insulator with time-reversal symmetry. Let $v_{j\mathbf{k}}(\mathbf{r})$ represent quasi-Bloch states generated by a unitary transformation of the occupied eigenstates $\psi_{n\mathbf{k}}(\mathbf{r})$. Because the transformation is unitary, we can write

$$E(\mathbf{k}) = \sum_j \varepsilon_j(\mathbf{k}) \,. \tag{17}$$

where

$$\varepsilon_j(\mathbf{k}) = \left\langle v_{j\mathbf{k}} \middle| H \middle| v_{j\mathbf{k}} \right\rangle \tag{18}$$

For an insulator with time-reversal symmetry, the quasi-Bloch states $v_{j\mathbf{k}}(\mathbf{r})$ can be chosen so that they are analytic with respect to $\mathbf{k}$ [8]. From these states, we can construct exponentially-localized Wannier functions, defined as

$$\left|\mathbf{R}_\mathbf{m} j\right\rangle = \frac{1}{\Omega}\int_{BZ} v_{j\mathbf{k}} e^{-i\mathbf{R}_\mathbf{m}\cdot\mathbf{k}} d\mathbf{k} \,. \tag{19}$$

Because we have defined $\mathbf{R}_1$ as the lattice point at the origin, equations (18) and (19) yield

$$\left\langle \mathbf{R}_\mathbf{1} j \middle| H \middle| \mathbf{R}_\mathbf{m} j \right\rangle = \frac{1}{\Omega}\int_{BZ} \varepsilon_j(\mathbf{k}) e^{-i\mathbf{R}_\mathbf{m}\cdot\mathbf{k}} d\mathbf{k} \,. \tag{20}$$

By equations (15), (17), and (20),

$$f_m = \sum_j \left\langle \mathbf{R}_\mathbf{1} j \middle| H \middle| \mathbf{R}_\mathbf{m} j \right\rangle \,. \tag{21}$$

Thus the error due to *k*-point sampling is the sum of the Hamiltonian matrix elements between Wannier functions at the origin and symmetrically equivalent Wannier functions at all other points on the supercell lattice. Because these Wannier functions can be constructed to be exponentially localized, it can be expected that the matrix elements will decay exponentially with $|\mathbf{R_m}|$ [9]. Under the simplifying *a-priori* assumption that the decay is isotropic, it follows that the *k*-point sampling error will strongly depend on the shortest distance between Wannier functions on the supercell lattice. Thus the shortest distance between points on the supercell lattice becomes a natural metric for estimating the error due to *k*-point sampling.

For insulators, the analyticity of $v_{j\mathbf{k}}(\mathbf{r})$ implies the analyticity of $\varepsilon_j(\mathbf{k})$ and hence $E(\mathbf{k})$. Because the values f_m are simply the coefficients of the Fourier transform of $E(\mathbf{k})$, they must decay exponentially with $|\mathbf{R}_m|$, the distance between $\mathbf{R}_m$ and the origin, when $E(\mathbf{k})$ is analytic [10]. This analysis is consistent with the above interpretation based on exponentially-localized Wannier functions. However for metals at 0 K, $E(\mathbf{k})$ is not analytic with respect to $\mathbf{k}$ due to the discontinuity in the occupancy function, and f_m cannot be expected to decay exponentially with $|\mathbf{R}_m|$. The integration error will be largely determined by how well a given *k*-point grid is able to identify the shape of the Fermi surface. Although the benefits of choosing bcc-like *k*-point grids (as opposed to, for example, fcc-like grids) are less clear in this case, we find that the evenly-spaced *k*-point grids returned by our method also work well for metals in practice. Because in many cases it is not known *a-priori* whether a material will have a calculated band gap, we use the Moreno-Soler approach, based on r_{min}, for all materials.

B. Database generation

For three-dimensional crystalline systems with time-reversal symmetry, the symmetry of possible *k*-point grids is given by one of 24 centrosymmetric symmorphic space groups [11]. For the 21 of these space groups that are neither triclinic nor monoclinic, the angles between lattice vectors are fixed, and only the lengths of the lattice vectors may change. To represent these space groups, we generated a set of 9094 sample structures with different lattice parameters, where the length of the longest conventional lattice vector could be up to 16 times as long as the length of the shortest vector. For each sample structure, we generated all possible real-space superlattices with up to 1728 (12×12×12) total primitive cells for orthorhombic, tetragonal, trigonal, and hexagonal space groups, and 5832 (18×18×18) total primitive cells for cubic space groups. Approximately 1% of the symmetrically distinct superlattices generated in this way had the same point group symmetry as the primitive lattice. When combined with the shift vectors described below (or no shift vector), these superlattices ensure that the set of *k*-points and all symmetrically equivalent points form a regular grid. They also make full use of symmetry to reduce the number of irreducible *k*-points. For these reasons, they were used to generate all *k*-point grids in the database.

For each symmetry-preserving superlattice, we calculated $r_{lattice}$, defined as the minimum spacing between points on the superlattice. For each sample structure, we identified the subset of superlattices that are on the Pareto frontier with respect to $r_{lattice}$ and the number of irreducible *k*-points in the corresponding *k*-point grid (Fig. 2); these are the superlattices for which there is no other superlattice that has greater (or equal) $r_{lattice}$ and for which the corresponding *k*-point grid has fewer irreducible *k*-points. For any superlattice that is not on the frontier, it is always possible to find one on the frontier that is superior with respect to $r_{lattice}$ and the number of irreducible *k*-points. For this reason, only the

superlattices on the Pareto frontiers were used to generate the *k*-point grid database. The resulting database contains 58,151 *k*-point grids.

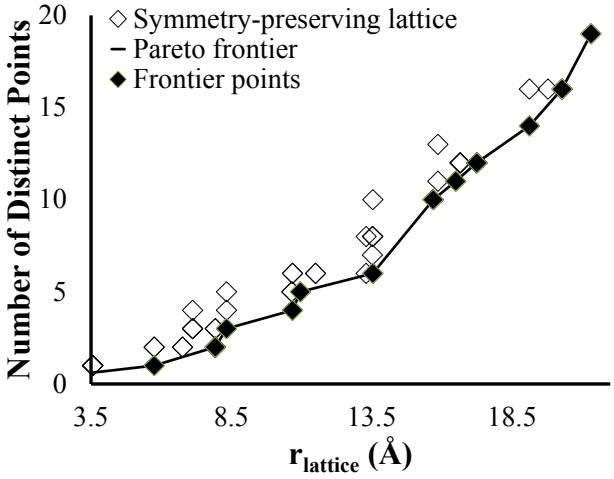

FIG. 2. An example of a Pareto frontier used to identify the best *k*-point grids for a given structure. All symmetry-preserving lattices are indicated by diamonds, and the lattices on the frontier are indicated by black diamonds. For any lattice not on the frontier, it is always possible to find a better lattice, in terms of the number of distinct *k*-points and $r_{lattice}$, on the frontier.

C. Dynamic grid generation

For the triclinic and monoclinic space groups, we do not use a database due to the large number of lattices that would need to be included in the database and because the benefit of pre-calculating Pareto frontiers is relatively small for systems with low symmetry. Instead we dynamically identify the smallest superlattices for which $r_{\text{lattice}} > r_{\text{min}}$ and the point group symmetry matches that of the primitive lattice. Of these, we use the lattice that results in the fewest irreducible k-points.

D. Large *k*-point grids

To maintain fast performance for dynamic grid generation we currently limit our dynamic search to superlattices with no more than 216 total primitive cells. If no superlattice with 216 or fewer primitive

cells is found, we identify the best superlattice for which $r_{\text{lattice}} > \frac{r_{\min}}{n}$, where the initial value of n is 2. If again no superlattice with 216 or fewer primitive cells is found, we iteratively increment n by 1 until we find a superlattice with no more than 216 primitive cells. We then return a *k*-point grid corresponding to a $n \times n \times n$ supercell of this superlattice. A similar method is used to generate lattices with more than 1728 *k*-points (for non-cubic space groups) and 5832 *k*-points (for cubic space-groups). For the results presented in this paper, convergence was always reached with $n = 1$.

III. RESULTS

A. Methods evaluated

We tested three different methods for generating *k*-point grids. The first, which we will refer to as the Generalized grid Database (GD), is the method described in the previous section. The second, which we will refer to as the Diagonal grid Database (DD), is the same as GD with the exception that the matrix $\mathbf{M}$ in equation (10) is constrained to be diagonal.

The two methods described above were compared to the automatic *k*-point generation scheme used in the Vienna Ab-initio Simulation Package (VASP) [12-16]. In this scheme, a Monkhorst-Pack grid is created where m_i, the number of grid points along the i^{th} reciprocal lattice vector $\mathbf{b_i}$, is determined by choosing a parameter *kspacing* and rounding $\frac{|\mathbf{b_i}|}{kspacing}$ up to the nearest integer. We will refer to this approach as a Simple Diagonal grid (SD). The SD method is representative of common approaches for generating Monkhorst-Pack grids.

In the DD and SD methods, the choice of the initial primitive cell is important, as the set of possible *k*-point grids is determined by the choice of lattice vectors for the primitive cell. We note that this is different from the GD method, for which identical *k*-point lattices are returned for any choice of the

primitive lattice vectors. For all methods, we used Minkowski-reduced lattice vectors [17,18] to represent all structures.

To generate shifted GD and DD grids (where the grid is shifted off the Γ point), the shift vector **s** was calculated as

$$\mathbf{s} = 0.5\mathbf{v}_1 + 0.5\mathbf{v}_2 + 0.5\mathbf{v}_3 . \tag{22}$$

where $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are the lattice vectors of a conventional unit cell for the *k*-point grid. For structures with trigonal and hexagonal symmetry, the shift vector was constrained to be parallel to the three-fold rotational axis to avoid breaking symmetry. For the SD method, we used the Monkhorst-Pack method implemented in VASP, for which

$$\mathbf{s} = 0.5d_1\mathbf{b}_1 + 0.5d_2\mathbf{b}_2 + 0.5d_3\mathbf{b}_3 . \tag{23}$$

where $d_i = 1$ if m_i (the number of grid points in the i^{th} direction) is even, and $d_i = 0$ if m_i is odd. We explored the common practice of using unshifted (a.k.a. Γ-centered) grids instead of shifted grids for trigonal and hexagonal lattices, but we found that after correcting for fatal errors (as described below) this made the SD results worse.

For fcc materials, shifted grids generated using the DD method present a special case. When using a Minkowski-reduced primitive cell for an fcc lattice, it is impossible to generate a symmetry-preserving *k*-point grid that does not include the Γ point. This problem can be addressed by careful selection of the vectors used to define the primitive cell (in a way that is not Minkowski-reduced). As this is not commonly done in practice and would amount to manually creating a GD grid, we instead chose to include the Γ point for all "shifted" *k*-point grids generated using the DD method for fcc materials.

B. Benchmark calculations

To test for the effectiveness of our approach, we used density functional theory (DFT) [19,20] as implemented in VASP to calculate the converged energies of 102 materials randomly selected from the Inorganic Crystal Structure Database (ICSD) [21]. For each material, the ionic positions were pre-relaxed, and then for benchmarking we performed a static run on the relaxed structure. Additional details of our calculations and the selected structures are provided in the supplemental material [7]. To determine the number of *k*-points required to calculate a converged energy value, we generated *k*-point grids for 33 different values of r_{min} for each material using each of the three different methods. These grids were generated by starting from $r_{\text{min}} = 100$ Å and incrementally reducing r_{min} by a factor of $2^{1/6}$ until we reached $r_{\text{min}} = 2.48$ Å. For the SD method, we used $kspacing = \frac{2\pi}{r_{\text{min}}}$. The lower limit of $r_{\text{min}} = 2.48$ Å was chosen to ensure that for all materials, the least-dense DD and GD grids contained exactly one irreducible *k*-point. For four materials, a lower value of r_{min} was required to generate a SD grid with exactly one irreducible *k*-point, but generating grids using $r_{\text{min}} < 2.48$ Å would not have changed the convergence results for any of those materials.

All calculations were done using the tetrahedron method with Blöchl corrections [22] except for when there were less than 4 irreducible points; in this case Gaussian smearing with a width of 0.05 eV was used. For each grid, we calculated the error, Δ_{grid}, as the absolute difference between the calculated energy per atom and that calculated using the densest grid. The calculation was considered to be converged at the least-dense grid for which Δ_{grid} for all grids with greater or equal density was less than the maximum acceptable error. For all grid-generation methods, anomalously high values of Δ_{grid} were observed for some calculations that had both BRMIX and DENTET warnings in VASP. All 496

calculations that had both BRMIX and DENTET warnings were re-run using Gaussian smearing, which removed the anomalous results.

For nearly 40% of the shifted grids generated using the SD method, VASP threw fatal errors. These errors are listed in the supplemental material [7]. In such cases, as if often done in practice, the shifted grid was replaced with a Γ-centered grid. For four calculations, using SD Γ-centered grids with $r_{\text{min}} = 100$ Å (the densest grids), VASP threw fatal errors or stalled. In these cases, energy was replaced by the average of the energies of the densest grids generated using other methods. Because there was little variation in the energies per atom for the densest grids, we do not believe this substitution significantly affected our results. For calculations using the DD and GD *k*-point grids, VASP did not throw fatal errors for any of the 13,332 calculations.

C. Benchmarking results

For the three different methods, the average number of irreducible *k*-points required to reach different levels of convergence for the calculated energy is shown in Fig. 3. For convergence of the energy within 1 meV / atom, the SD method required on average 2.25 times as many irreducible *k*-points as the GD method for Γ-centered grids and 2.69 times as many for shifted grids. Most of the gain appears to come from the uses of high-symmetry grids and the Pareto frontier, as the DD method also shows significant gains over the SD method, despite the fact that both methods result in "diagonal" grids aligned with the reciprocal lattice vectors. The relative advantage of the DD and GD methods over the SD method increases as the convergence criterion is tightened. This is primarily because the benefit of using highly symmetric grids, in terms of the reduction in the total number of irreducible *k*-points, is greater for grids that have a large number of total *k*-points.

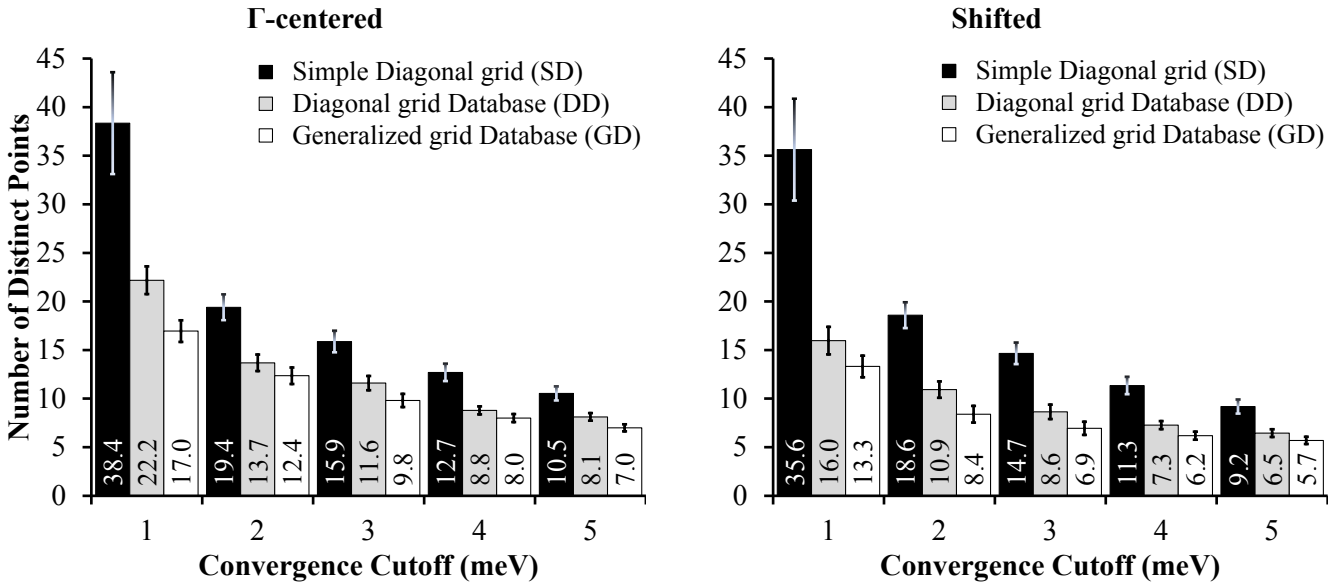

FIG. 3. The average number of distinct k-points required to reach convergence within different levels of accuracy for different grid-generation methods. Error bars represent the standard error of the mean.

Under the assumption that the computational time per *k*-point is constant for each material, we have calculated the speed of each method, relative to the SD method, for each material. The average of these speed-ups across all materials is shown in Fig. 4. The trends are similar to those for the average number of irreducible *k*-points. For Γ-centered grids, the GD method is about 50-100% faster than the SD method. For shifted grids, the difference is much greater. This is likely due to the facts that the SD shifted grids include a mix of calculations using shifted and Γ-centered grids, and the SD method determines the shifts differently than the GD or DD methods. The speed-up when going from DD grids to GD grids is roughly the same for both Γ-centered and shifted grids.

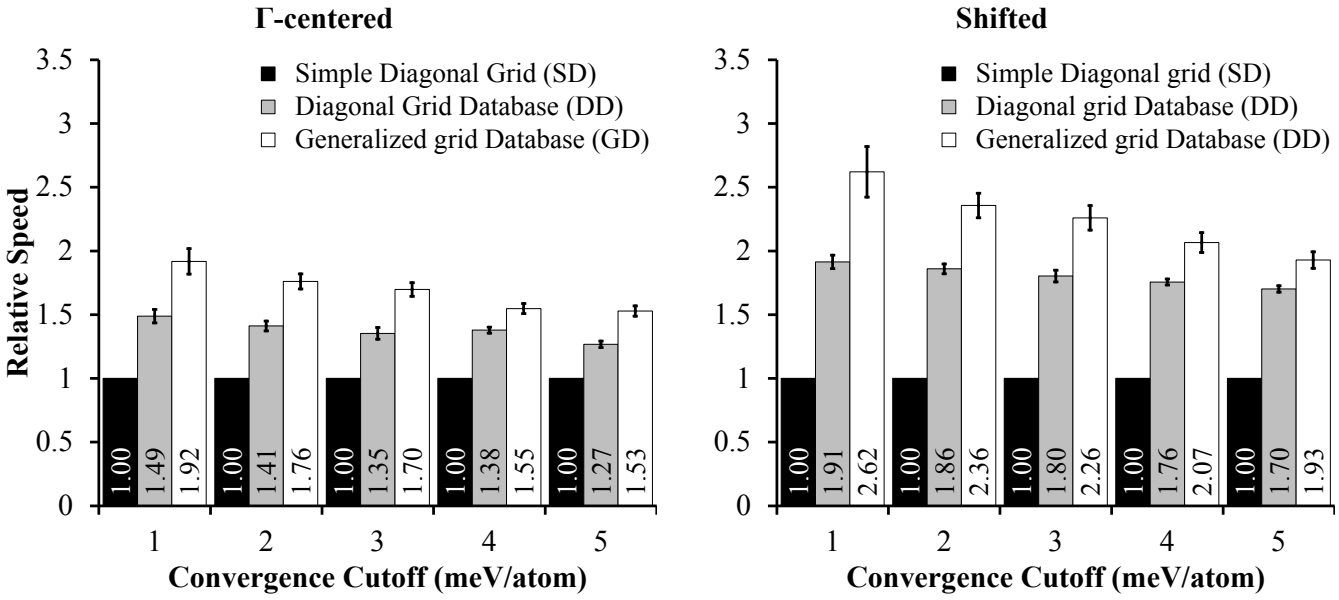

FIG. 4. The average speed-up per material at different levels of accuracy for different grid-generation methods under the assumption that the computational time per *k*-point is constant for each material. The SD method was used as the reference value for both Γ-centered and shifted runs. Error bars represent the standard error of the mean.

For high-throughput calculations, the total CPU time for the entire batch of materials (Fig. 5) is the metric that determines the throughput. To estimate the total computational cost, we assumed that the calculation time scales linearly with the number of *k*-points; analysis of the benchmark calculations supported this assumption. To minimize the noise in benchmark times due to e.g. running on different nodes at different times, for each material a single value for the average CPU time per *k*-point was used for all methods. This value was taken from the converged SD calculation using a Γ-centered grid. The benchmark time was then calculated as the number of irreducible *k*-points multiplied by the CPU time per *k*-point for the material, effectively weighting the number of irreducible *k*-points by a material-specific computational cost per *k*-point.

Shifted grids consistently resulted in lower CPU times, even for the SD method. To reach 1 meV / atom convergence, the GD shifted grids resulted in 2.78 times the throughput as SD Γ-centered grids,

and 1.95 times the throughput as SD shifted grids. The total performance gains shown in Fig. 5 are generally not as large as the average relative performance gains shown in Fig. 4. This is primarily because the total performance gains are effectively weighted by the CPU time for each material, and for some materials with large CPU times, such as those with large unit cells for which only one irreducible *k*-point is needed, there is relatively little difference among the three different methods.

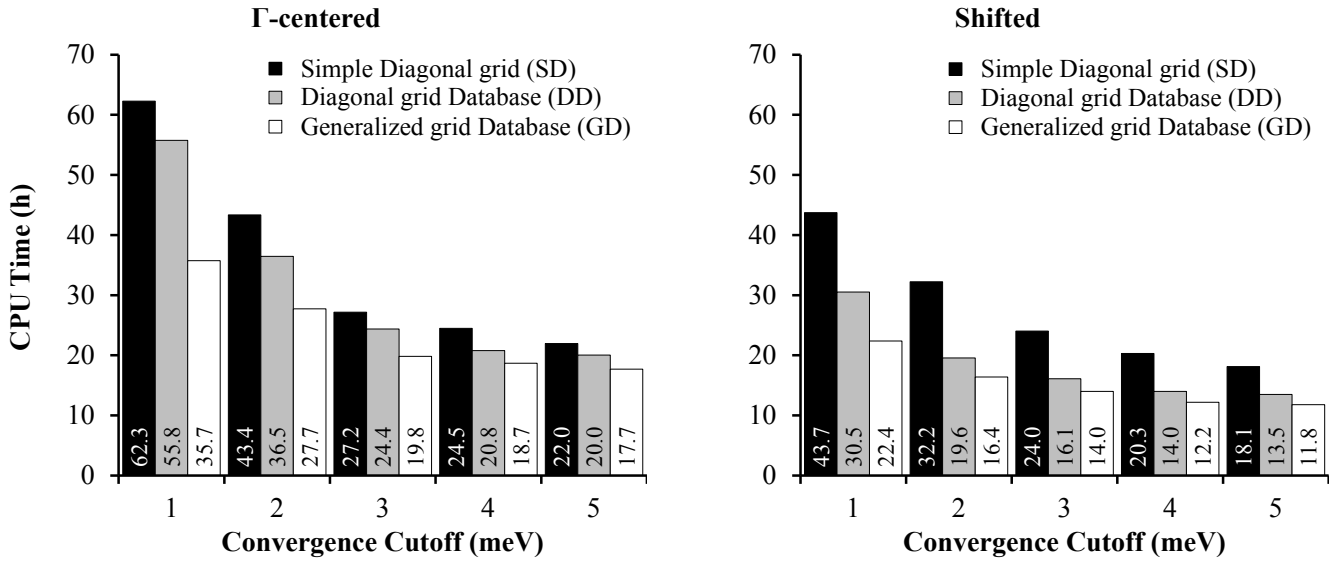

FIG. 5. The total computational time for all 102 materials in the benchmark set required to reach convergence within different levels of accuracy for different grid-generation methods. Error bars represent the standard error of the mean.

D. Determining r_{min}

In the Moreno-Soler approach, the speed and accuracy of the calculation are determined by the parameter r_{min}, the minimum acceptable distance between points on the real-space superlattice. Ideally, for a given material system, the user would test different values of r_{min} to identify the smallest value at which the calculation is sufficiently accurate. However in practice, it is helpful to use a heuristic that allows for rapid estimation of a reasonable value for r_{min}. This is especially true for high-throughput

calculations, in which the selection of the *k*-point grid is entirely automated and separately testing for *k*-point convergence for every material could be prohibitively expensive. Here we discuss several approaches for estimating the appropriate value for r_{min}.

1. General trends

We start by examining how the error in our set of 102 benchmark materials varies as a function of r_{min}. We consider two measures of the error: the average absolute error across all 102 materials, and the maximum absolute error across all 102 materials. We divide our analysis into two classes of materials: metals, defined as those materials for which there is no indirect band gap for a DFT calculation using the densest Γ-centered GD grid (i.e. $r_{\text{min}} = 100$ Å), and non-metals, defined as all other materials. There are 50 metals and 52 non-metals in the benchmark set. To calculate the errors used in this analysis, we use shifted GD grids, as we expect they will be the most widely used due to their superior efficiency. The average absolute error, maximum absolute error, and average CPU time as a function of r_{min} are presented in Fig. 6. Tabulated values are provided in the supplemental material [7].

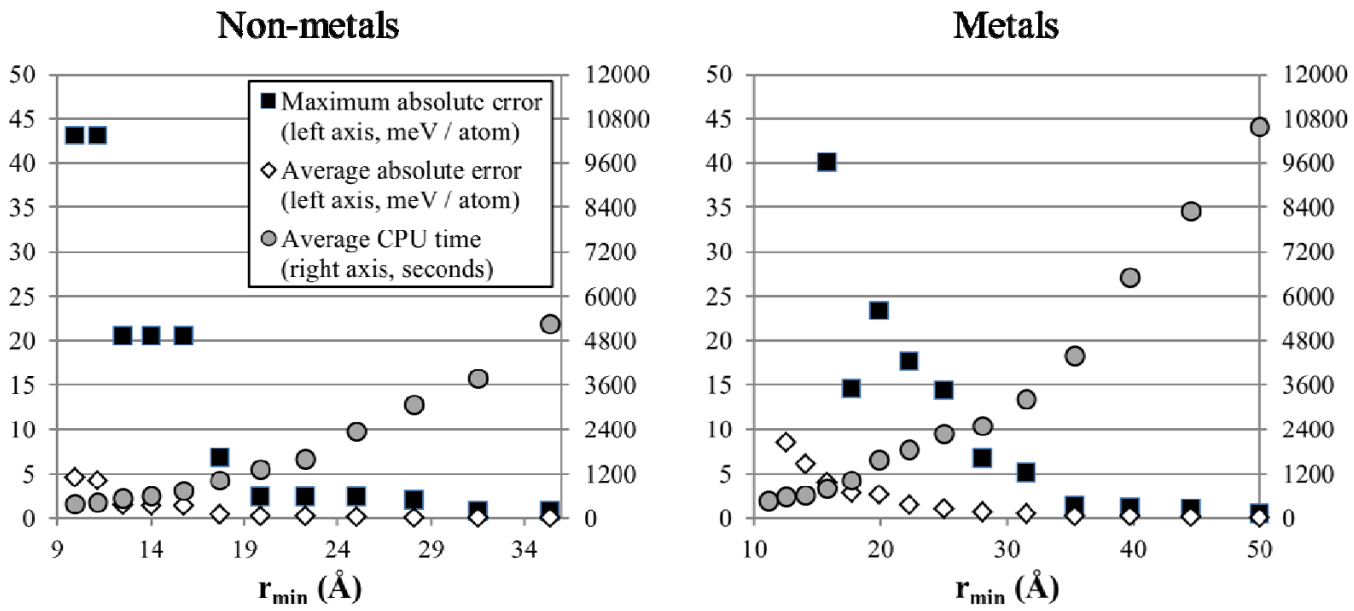

FIG. 6. The maximum absolute error (black squares, left axis), average absolute error (white diamonds, left axis), and average CPU time (grey circles, right axis) for metals and non-metals as a function of $r_{\min}$.

We have also calculated the fraction of the calculations that converge within a given level of accuracy as a function of $r_{\min}$ (Fig. 7). For non-metals, 100% of the calculations had converged within 3 meV / atom, and 92.3% had converged within 1 meV / atom, at $r_{\min} = 19.8$ Å. The results are only slightly worse at $r_{\min} = 17.7$ Å, where 98.1% have converged within 3 meV / atom and 100% have converged within 10 meV / atom. Under the assumption that computational cost scales linearly with the number of irreducible *k*-points, calculations at $r_{\min} = 19.8$ Å can be expected to take approximately 40% longer than calculations at $r_{\min} = 17.7$ Å, so the computational cost savings from reducing $r_{\min}$ might be enough to justify the small loss of accuracy.

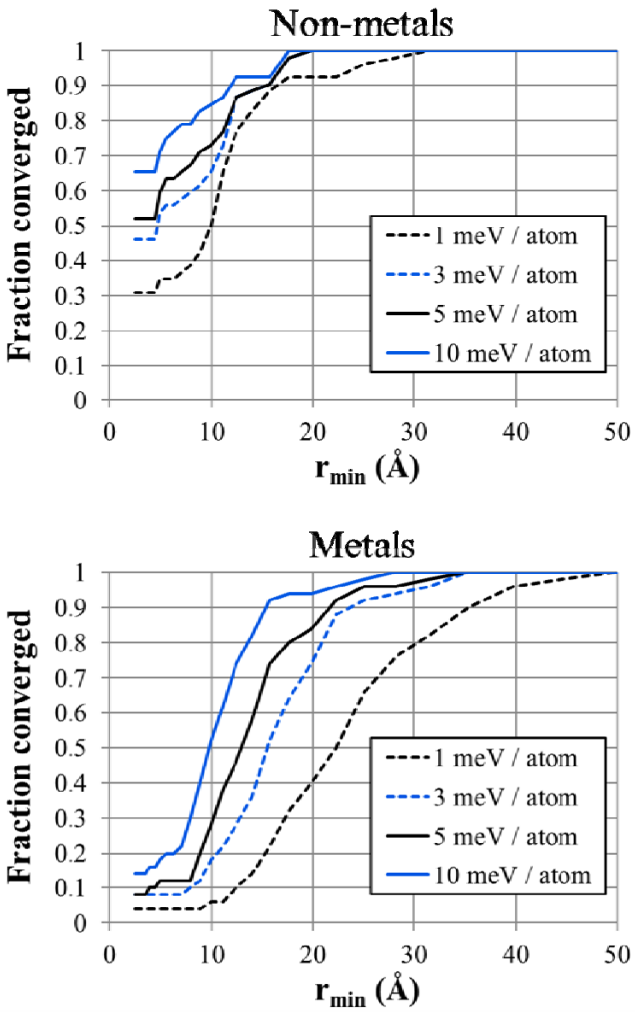

FIG. 7. The fraction of calculations for non-metals and metals that have reached convergence within different levels of accuracy as a function of r_{min}.

For metals, the first point at which 100% of the calculations converged within 1 meV / atom is at $r_{\text{min}} = 50$ Å. At $r_{\text{min}} = 28.1$ Å, all non-metals in our data set are converged within 3 meV / atom and all metals are converged within 7 meV / atom. For both metals and non-metals, the average absolute error is less than 1 meV / atom at $r_{\text{min}} = 28.1$ Å. Thus 28.1 Å might be a reasonable default value of r_{min} for calculations in which the existence of a band gap in the material is unknown and reasonably accurate energies are desired. However we note that calculations at $r_{\text{min}} = 28.1$ Å can be expected to cost, on

average, about 4 times as much as calculations at $r_{\min} = 17.7$ Å, demonstrating that prior knowledge of the existence of a band gap can significantly reduce the required computational time.

The above results were generated for our sample of 102 random materials, and the subsets of metals and non-metals only contain about 50 materials each. To reach any given level of accuracy, there are almost certainly materials that require larger values of $r_{\min}$ than any material in our data set. This is important to keep in mind for some applications, such as phase diagram calculations, in which a single material with a large error in its calculated energy can dramatically affect the outcome. If close to 100% convergence within a given level of accuracy is required, we advise using values of $r_{\min}$ that are larger than those indicated in Fig. 7.

2. As a function of the band gap for non-metals

For reasons discussed in section II.A, in the limit of large $r_{lattice}$, the error due to *k*-point sampling will decay as $e^{-\gamma r_{lattice}}$ for materials with a band gap. The rate of decay, γ, will be determined by the decay rate of the Hamiltonian matrix elements between Wannier functions, which can be expected to be the same as the rate of decay for the density matrix [23,24]. In the weak-binding limit, the exponential decay rate of the density matrix is expected to be proportional to E_{gap}, the direct band gap of the material [24,25]. Thus it can be expected that in the weak-binding limit, the value of $r_{\min}$ required to reach a given level of convergence will vary roughly inversely with the direct band gap. In the tight-binding limit (materials with a large band gap), the relationship between γ and E_{gap} has not been well established [24].

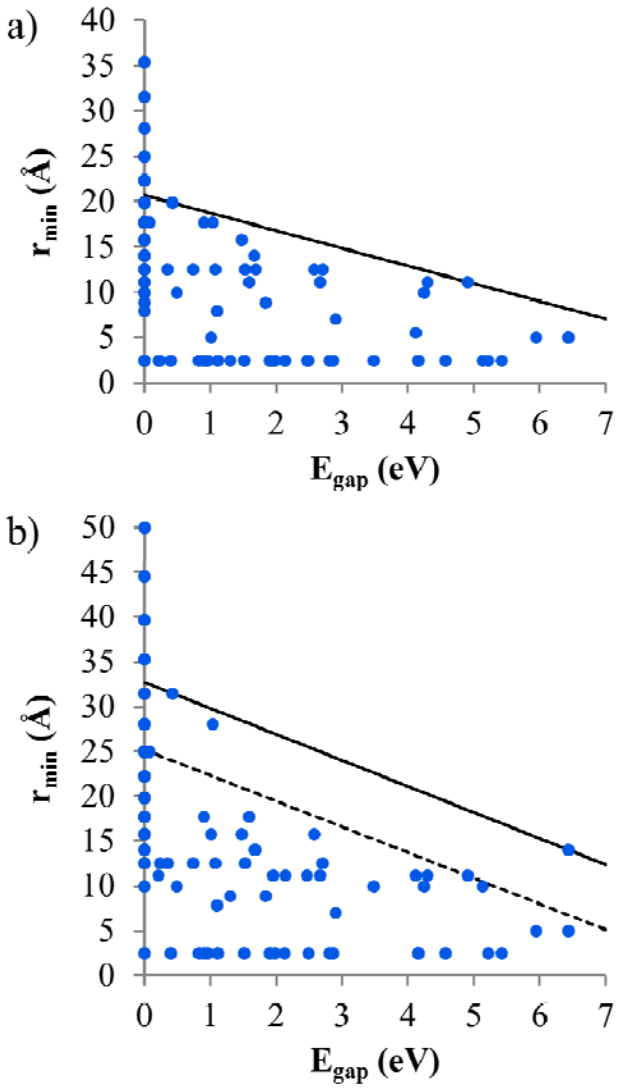

FIG. 8. The relationship between the direct band gap (E_{gap}) and the value of $r_{\min}$ at which convergence was reached within a) 3 meV / atom b) 1 meV / atom. The solid diagonal line in a) is a plot of equation (24). The solid diagonal line in b) is a plot of equation (25), and the dashed diagonal line in b) is a plot of equation (26).

For the 102 structures in our data set, we plot the relationship between E_{gap} (the direct band gap) and the value of $r_{\min}$ required to reach convergence within 3 meV / atom in Fig. 8a. For these plots, E_{gap} was calculated using DFT with a Γ-centered GD *k*-point grid at the maximum density ($r_{\min} = 100$ Å). The value of $r_{\min}$ at which convergence was reached was calculated using shifted GD grids. As

expected, average values for r_{min} decrease as E_{gap} increases. The upper edge of the scatterplot in Fig. 8 is roughly linear, and we find that all non-metals in our benchmark set would be converged within 3 meV / atom if r_{min} were at or above the value calculated using the following equation:

$$r_{\text{min}} = 20.69 - 1.95\left(E_{gap}\right) \tag{24}$$

where r_{min} is in Angstroms and E_{gap} is expressed in eV. Similar analysis for convergence levels of 4 meV / atom and 5 meV / atom yield the same equation.

To reach a convergence level of 1 meV / atom for all non-metals in our data set (Fig. 8b), the following equation could be used to set r_{min} :

$$r_{\text{min}} = 32.76 - 2.91\left(E_{gap}\right). \tag{25}$$

If equation (24) were used to set r_{min} instead of equation (25), 86.5% of the non-metals would still be converged within 1 meV / atom. As a compromise between the two approaches, the following equation would converge all non-metals within 3 meV / atom and 94.2% within 1 meV / atom:

$$r_{\text{min}} = 25.22 - 2.87\left(E_{gap}\right). \tag{26}$$

The lines represented by equations (24), (25), and (26) are all plotted in Fig. 8.

In practice, good estimates for E_{gap} are often not available before a calculation has been run, but there is often a sense of whether a material is a semiconductor or a large band-gap insulator. As an alternative to the linear bounds provided above, we note that for all materials in our data set with E_{gap} of more than 0 eV and less than 2 eV, we note that $r_{\text{min}} = 20$ Å would be sufficient to reach convergence within 3 meV / atom. For materials in our data set with E_{gap} greater than 2 eV, $r_{\text{min}} = 12.5$ Å would be sufficient.

3. Relationship with other methods

In some cases, particularly for high-throughput calculations, an automated method already exists for determining *k*-point grids. It is helpful to consider how such methods relate to the method described in this paper. For example, an alternative to using r_{min} to determine the grid density is to require that the grid contains at least N_{kpt} *k*-points in the Brillouin zone, where N_{kpt} is calculated separately for each material. It is straightforward to find a value for r_{min} that guarantees that this condition is met. For a given value of r_{min}, the number of total *k*-points will be minimized by generating an fcc real-space superlattice. Thus the following value of r_{min} will ensure that there are at least N_{kpt} total *k*-points in the Brillouin zone:

$$r_{\text{min}} = \left(N_{kpt} \times V_{prim} \times 2^{1/2} \right)^{1/3} \tag{27}$$

With r_{min} given by equation (27), in some cases the search along the Pareto frontier might discover a grid with significantly more than N_{kpt} total *k*-points but relatively few irreducible *k*-points. Such a grid can be expected to result in particularly low *k*-point approximation error at a low computational cost.

A second approach to choosing values for r_{min} in a way that is consistent with other methods is to match the percentage of calculations that converge within a given level of accuracy. For example, consider a hypothetical method for generating *k*-point grids which results in energy convergence within 5 meV / atom 90% of the time. We could define an equivalent value of r_{min} as that which would result in the same convergence rate.

We illustrate the second approach by first considering the relationship between $r_{lattice}$ and the number of *k*-points per reciprocal atom ($kAtom$). The Materials Project [26], Aflowlib [27], and Open Quantum Materials Database [28] all use $kAtom$ to set the density of *k*-point grids. For each of the 102 materials

in our database, we identified the shifted GD *k*-point grids at which the calculation had converged within 1 meV / atom. For each of these grids we calculated both $kAtom$ and $r_{lattice}$ (based on the corresponding real-space superlattice). We then generated two sorted lists of *k*-point grids: the first sorted according to $kAtom$, and the second sorted according to $r_{lattice}$. Each was sorted from smallest to largest.

Let k_n represent the value of $kAtom$ for the n^{th} item on the first list. If grids with $kAtom = k_n$ were generated for all 102 materials, calculations using these grids for the first $n-1$ materials on the first list would be expected to be converged within 1 meV / atom, and calculations using these grids for the remaining $102-n$ materials would not. Similarly, let r_n represent the value of $r_{lattice}$ for the n^{th} item on the second list. If grids with $r_{lattice} = r_n$ were generated for all 102 materials, calculations using these grids for the first $n-1$ materials on the second list would be expected to be converged, and calculations using these grids for the remaining $102-n$ materials would not. Thus using a lower bound of $kAtom_{\min} = k_n$ *k*-points per reciprocal atom can be expected to yield about the same percentage of converged calculations as setting $r_{\min} = r_n$.

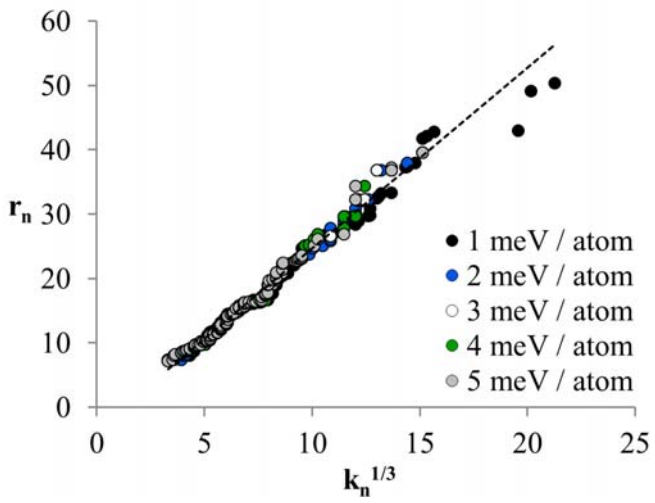

FIG. 9. A plot of $k_n^{1/3}$ vs. r_n for sorted lists generated at five different levels of convergence. For the n^{th} point on each list, there are $n-1$ materials that converged with a lower (or equal) value of r_n and $n-1$ materials that converged with a lower (or equal) value of k_n. The diagonal dashed line illustrates the best linear fit to all 510 points, given by equation (28).

In Fig. 9, we plot r_n vs $k_n^{1/3}$ for five different levels of convergence: 1, 2, 3, 4, and 5 meV / atom. The relationship is nearly linear and along the same line for all five levels of convergence. Linear regression yields the following estimate for r_n as a function of $k_n^{1/3}$, with $R^2 = 0.988$:

$$r_n \approx 2.8074\left(k_n^{1/3}\right) - 3.4008 \tag{28}$$

Equation (28) allows us to establish a relationship between $kAtom_{\min}$, the minimum allowed number of *k*-points per reciprocal atom, and $r_{\min}$, the minimum allowed value of $r_{lattice}$ (Fig. 10a). For example, a minimum of 1000 *k*-points per reciprocal atom corresponds to $r_{\min} = 24.7$ Å, and a minimum of 7000 *k*-points per reciprocal atom corresponds to $r_{\min} = 50.3$ Å.

We have done similar analysis for two other metrics for *k*-point grid density: the number of *k*-points per reciprocal cubic Angstrom ($kVol$) and the length of the longest vector in the Minkowski-reduced

representation of the *k*-point lattice in reciprocal space ($kDist$). We note that $kDist$ is similar to the $kspacing$ value used by VASP. The relationship between r_n and $kVol_n$ ($R^2 = 0.993$) is

$$r_n \approx 1.0688\left(kVol_n^{1/3}\right) - 2.5877\,, \tag{29}$$

and the relationship between r_n and $kDist_n$ ($R^2 = 0.994$) is

$$r_n \approx 1.0265\left(\frac{2\pi}{kDist_n}\right) + 1.0183\,. \tag{30}$$

Based on equations (29) and (30), the estimated equivalent values of r_{min} for different values of $kVol_{min}$ and $kDist_{max}$ are shown in Fig. 10b and Fig. 10c.

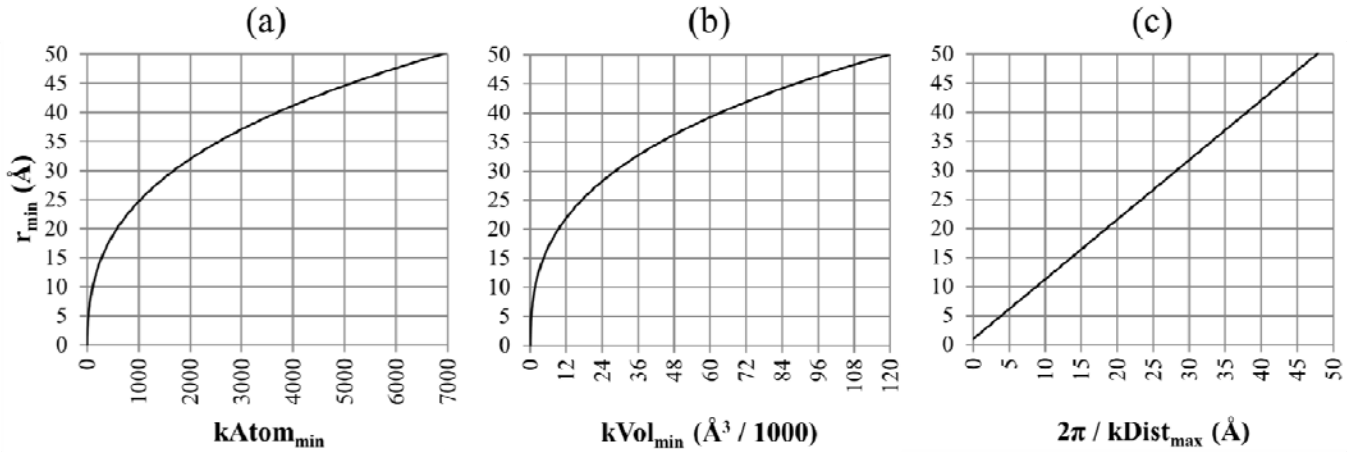

FIG. 10. The values of r_{min} that will yield approximately the same convergence rate as three different methods for setting the minimum *k*-point density.

E. Conclusion

We have presented a method for rapidly generating highly efficient *k*-point grids. There are several practical advantages to using the method presented in this paper. There is only one parameter, r_{min}, that needs to be set by the user, and we have provided guidance on how to select a good value for this

parameter. The generated *k*-point grids are always consistent with the symmetry of the material, which both preserves the symmetry of the system and maximizes the degree to which symmetry can be used to reduce the cost of the calculation. The grids are independent of the lattice vectors chosen to represent the real-space primitive cell. Thus the user can change the way in which the structure is represented in the input file without changing the results of the calculation (at least to the extent that those results depend on the choice of *k*-points). Perhaps most importantly, grids generated using our method result in a significant reduction in the number of irreducible *k*-points required to reach a given level of convergence, resulting in large savings in computational time. For Γ-centered grids, we observed average speed-up of about 50-100% compared to a more conventional approach (Fig. 4), and for shifted grids the speed-up is even greater.

To allow others to generate *k*-point grids using our method, we have constructed a free and publicly available *k*-point grid server that provides access to our database of generalized *k*-point grids. This server was used to generate all GD grids in this manuscript. On average, it took 0.3 seconds to generate each of the grids for calculations converged within 1 meV / atom. Instructions for the use of this server can be found in the supplemental material [7], and updates will be posted at our web site, http://muellergroup.jhu.edu.

There are a number of areas in which we are working to improve our approach. Currently, we only generate grids in a format suitable for use with VASP, but we will be building interfaces to other common software packages. Our tool is also currently limited by the assumptions that the systems have time-reversal symmetry and the decay in the Hamiltonian matrix elements is isotropic, but neither of these assumptions is an inherent limitation of our approach. The decay in the Hamiltonian matrix elements is unlikely to be isotropic for many systems [24,25,29], but prior knowledge about the nature of the anisotropy is usually not available. Thus for most materials, assuming isotropic decay is a

pragmatic approximation. However for some systems, such as slabs separated by vacuum, the nature of the anisotropy is clear, and we are updating our server to allow for the generation of suitably anisotropic *k*-point grids for such systems. In addition, we will improve the quality of our database by adding additional grids, including larger grids and grids representing non-centrosymmetric space groups, to it. Given the benchmark results presented in this paper, we anticipate that the use of this server will lead to significant acceleration of calculations on crystalline materials.

ACKNOWLEDGMENTS

We thank Kristin Persson, Anubhav Jain, Joseph Montoya, and Wei Chen from the Materials Project for beta testing. This work was funded by the National Science Foundation under award DMR-1352373.

[+]Electronic Address: tmueller@jhu.edu

[1] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973).
[2] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).
[3] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[4] S. Froyen, Phys. Rev. B **39**, 3168 (1989).
[5] J. Moreno and J. M. Soler, Phys. Rev. B **45**, 13891 (1992).
[6] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[7] See Supplemental Material at [URL will be inserted by publisher] for a derivation of equation (13), additional details about the structures used for benchmarking, addtional details about the DFT calculations, tabulated values for Fig. 6, and instructions for using the server.
[8] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Phys. Rev. Lett. **98**, 046402 (2007).
[9] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).
[10] A. Quarteroni, C. Canuto, M. Hussaini, and T. Zang, *Spectral methods: Fundamentals in single domains* (Springer, Berlin, 2006).
[11] M. I. Aroyo and H. Wondratschek, in *International Tables for Crystallography* (John Wiley & Sons, Ltd, 2006).
[12] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
[13] G. Kresse and J. Furthmüller, Comput. Mat. Sci. **6**, 15 (1996).
[14] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[15] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).

[16] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[17] P. Q. Nguyen and D. Stehl, ACM Trans. Algorithms **5**, 1 (2009).
[18] I. Semaev, in *Cryptography and Lattices*, edited by J. Silverman (Springer Berlin Heidelberg, 2001), pp. 181.
[19] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964).
[20] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[21] Inorganic Crystal Structure Database, (Fiz Karlsruhe) http://www.fiz-karlsruhe.de/icsd.html.
[22] P. E. Blöchl, O. Jepsen, and O. K. Andersen, Phys. Rev. B **49**, 16223 (1994).
[23] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).
[24] S. Ismail-Beigi and T. A. Arias, Phys. Rev. Lett. **82**, 2127 (1999).
[25] S. N. Taraskin, D. A. Drabold, and S. R. Elliott, Physical Review Letters **88**, 196405 (2002).
[26] A. Jain *et al.*, APL Mater. **1** (2013).
[27] S. Curtarolo *et al.*, Comput. Mater. Sci. **58**, 227 (2012).
[28] J. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM **65**, 1501 (2013).
[29] U. Stephan and D. A. Drabold, Phys. Rev. B **57**, 6391 (1998).