arXiv:1901.03449v1 [cond-mat.mtrl-sci] 11 Jan 2019

Efficient *Ab Initio* Calculations of Electron-Defect Scattering and Defect-Limited Carrier Mobility

I-Te Lu, Jin-Jian Zhou, and Marco Bernardi*
Department of Applied Physics and Materials Science, California Institute of Technology, Pasadena, California 91125
(Dated: January 14, 2019)

Electron-defect (e-d) interactions govern charge carrier dynamics at low temperature, where they limit the carrier mobility and give rise to phenomena of broad relevance in condensed matter physics. *Ab initio* calculations of e-d interactions are still in their infancy, mainly because they require large supercells and computationally expensive workflows. Here we develop an efficient *ab initio* approach for computing elastic e-d interactions, their associated e-d relaxation times (RTs), and the low-temperature defect-limited carrier mobility. The method is applied to silicon with simple neutral defects, such as vacancies and interstitials. Contrary to conventional wisdom, the computed e-d RTs depend strongly on carrier energy and defect type, and the defect-limited mobility is temperature dependent. These results highlight the shortcomings of widely employed heuristic models of e-d interactions in materials. Our method opens new avenues for studying e-d scattering and low-temperature charge transport from first principles.

I. INTRODUCTION

Defects in materials can scatter or capture charge carriers. They control carrier dynamics at low temperature, where phonons are frozen out, or even at room temperature in the presence of strong disorder. Impurity defects are key to controlling the carrier mobility in semiconductors [1–6], and defects are widely employed to engineer advanced materials and devices [7–10]. Weak localization [11–13] and conductance fluctuations [14, 15] are but two examples of subtle quantum effects that can be induced by defects. For all these reasons, understanding how defects interact with charge carriers is important in both applied and fundamental condensed matter physics.

Calculations of electron-defect (e-d) interactions have so far relied almost exclusively on empirical models [16], in which the atomic details of the defects are suppressed and simple approximations are employed for the electronic screening and band structure. For example, the perturbation induced by charged defects is typically modeled with a simple Yukawa potential, and models of e-d scattering due to neutral (non-charged) defects are scarce and mainly limited to spherical-well-potential approaches [17–19]. Yet, the perturbation induced by defects in real materials can be anisotropic and complex, and its effects on electronic screening intricate. How the band structure and defect states impact e-d scattering is difficult to predict.

Due to their severe approximations, empirical models are not reliable to compute e-d interactions, and their predictions can be qualitatively and quantitatively incorrect. First-principles approaches, such as those based on density functional theory (DFT) and related methods [20], provide quantitative accuracy in many areas of materials physics, and are therefore desirable to treat e-d interactions. Such *ab initio* e-d calculations can take advantage of existing tools developed for electron-phonon (e-ph) interactions, for which accurate calculations of relaxation times (RTs) [21–24], matrix elements and their interpolation [25], and phonon-limited carrier mobility [26–30] have been recently developed. These and other workflows developed for e-ph calculations can be generalized to treat e-d interactions.

However, there are open challenges specific to e-d calculations that currently prevent their broad applicability. First-principles e-d calculations need large supercells to obtain the electron wavefunctions and defect perturbation potentials; they additionally involve computationally costly e-d matrix elements, and require systematic convergence of the RTs with respect to supercell size and Brillouin zone (BZ) grids. Early work on first-principles e-d interactions [31–33] proposed an approach, called here the *all-supercell method*, in which the cost of computing e-d matrix element is prohibitively large due to the use of electron wavefunctions from large supercells. This work aims to address these open challenges and make *ab initio* e-d calculations more affordable and broadly applicable.

Here we formulate an efficient scheme to compute and converge the e-d matrix elements and the associated RTs and defect-limited carrier mobility. Our approach does not require the wavefunctions of large supercells, which dramatically speeds up the calculations. We apply our method to study e-d interactions in silicon in two cases, neutral vacancy and interstitial defects, for which we compute and converge the e-d RTs as a function of energy and the carrier mobility as a function of temperature below 150 K. The results show that, contrary to conventional wisdom, the RTs depend strongly on carrier energy and defect type, and the defect-limited mobility for neutral defects depends on temperature. Our results provide new microscopic insight into e-d scattering, and our approach, together with its future extensions, can uncover new defect physics in materials and devices for electronics, energy and quantum technologies.

The paper is organized as follows: Section II describes the new formulas and workflow for computing e-d matrix elements and RTs, and compares our approach with

the all-supercell method. In Section III we compute and converge e-d RTs for electrons and holes in silicon and the associated defect-limited mobility at low temperature. Section IV discusses technical points and outlines future research directions.

II. METHODOLOGY

In the following, we work under the assumption that the defects are neutral (non-charged) and that the e-d scattering events are independent, uncorrelated and elastic. The e-d scattering rate $\Gamma_{n\boldsymbol{k}}$ (and its inverse, the RT, $\tau_{n\boldsymbol{k}} = \Gamma_{n\boldsymbol{k}}^{-1}$) for a Bloch state $|n\boldsymbol{k}\rangle$, where n is the band index and $\boldsymbol{k}$ the crystal momentum, is computed using lowest-order perturbation theory (see Appendix A):

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar}\frac{n_{\rm at}C_{\rm d}}{N_{\boldsymbol{k}'}}\sum_{n'\boldsymbol{k}'}|M_{n'\boldsymbol{k}',n\boldsymbol{k}}|^2\,\delta\left(\varepsilon_{n'\boldsymbol{k}'}-\varepsilon_{n\boldsymbol{k}}\right), \quad (1)$$

where $n_{\rm at}$ is the number of atoms in a primitive cell, $C_{\rm d}$ the (dimensionless) defect atomic concentration, $N_{\boldsymbol{k}'}$ the number of BZ $\boldsymbol{k}'$-points used in the sum, and $\varepsilon_{n\boldsymbol{k}}$ the unperturbed energy of the state $|n\boldsymbol{k}\rangle$. The e-d matrix elements $M_{n'\boldsymbol{k}',n\boldsymbol{k}}$ are the central quantities computed in this work; they encode the probability amplitude for scattering from the unperturbed state $|n\boldsymbol{k}\rangle$ to $|n'\boldsymbol{k}'\rangle$ due to the perturbation potential $\Delta V_{\rm e-d}$ from a defect:

$$M_{n'\boldsymbol{k}',n\boldsymbol{k}} = \langle n'\boldsymbol{k}'\,|\,\Delta V_{\rm e-d}\,|\,n\boldsymbol{k}\rangle\,. \quad (2)$$

Within DFT [34], the e-d perturbation can be computed as the difference between the Kohn-Sham (KS) potentials $V_{\rm KS}$ of a defect-containing supercell and a pristine supercell with no defect, namely, $\Delta V_{\rm e-d} = V_{\rm KS}^{\rm (d)} - V_{\rm KS}^{\rm (p)}$. Here and below, we use superscripts (d) and (p) to denote the defect-containing and pristine systems, respectively. When using, as we do here, norm-conserving pseudopotentials in the Kleinman-Bylander (KB) form [35], the Kohn-Sham potential can be written as a sum of local and nonlocal parts [20]:

$$V_{\rm KS} = V_{\rm L}(\boldsymbol{r}) + \hat{V}_{\rm NL}. \quad (3)$$

The local potential $V_{\rm L}(\boldsymbol{r})$ comprises the Hartree and exchange-correlation potentials plus the local part of the pseudopotentials,

$$V_{\rm L}(\boldsymbol{r}) = V_{\rm H}(\boldsymbol{r}) + V_{\rm XC}(\boldsymbol{r}) + V_{\rm pp}(\boldsymbol{r}). \quad (4)$$

The nonlocal potential $\hat{V}_{\rm NL}$, which is due to the pseudopotentials, is as a sum over all atoms in the supercell of KB projectors $|\beta_i^{(s)}\rangle$, each localized in the core region of atom s:

$$\hat{V}_{\rm NL} = \sum_{s=1}\sum_{ij} D_{ij}^{(s)}\,|\beta_i^{(s)}\rangle\,\langle\beta_j^{(s)}|\,, \quad (5)$$

where i and j are orbital angular momentum quantum numbers, and $D_{ij}^{(s)}$ are KB coefficients [20].

Accordingly, we separate the matrix elements into a local and a nonlocal part,

$$M_{n'\boldsymbol{k}',n\boldsymbol{k}} = M^{\rm L}_{n'\boldsymbol{k}',n\boldsymbol{k}} + M^{\rm NL}_{n'\boldsymbol{k}',n\boldsymbol{k}}\,, \quad (6)$$

each due to the respective defect perturbation, $\Delta V_{\rm L}(\boldsymbol{r}) = V_L^{\rm (d)}(\boldsymbol{r}) - V_L^{\rm (p)}(\boldsymbol{r})$ for the local and $\Delta\hat{V}_{\rm NL} = \hat{V}_{\rm NL}^{\rm (d)} - \hat{V}_{\rm NL}^{\rm (p)}$ for the nonlocal part:

$$\begin{aligned} M^{\rm L}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \langle n'\boldsymbol{k}'\,|\,\Delta V_{\rm L}(\boldsymbol{r})\,|\,n\boldsymbol{k}\rangle \\ M^{\rm NL}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \langle n'\boldsymbol{k}'\,|\,\Delta\hat{V}_{\rm NL}\,|\,n\boldsymbol{k}\rangle\,. \end{aligned} \quad (7)$$

A. Electron-defect matrix element computation

We develop a new approach to reduce the computational cost of the local and nonlocal matrix elements. We Fourier transform the local perturbation potential $\Delta V_{\rm L}(\boldsymbol{r})$ and compute the local matrix elements as (see Appendix B)

$$M^{\rm L}_{n'\boldsymbol{k}',n\boldsymbol{k}} = \sum_{\boldsymbol{G}}\Delta\widetilde{V}_{\rm L}(\boldsymbol{k}'-\boldsymbol{k}+\boldsymbol{G})\,\langle u_{n'\boldsymbol{k}'}|\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}}\,|u_{n\boldsymbol{k}}\rangle_{\rm uc}\,, \quad (8)$$

where $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is the periodic part of the Bloch wavefunction (normalized in a primitive cell with volume $\Omega_{\rm uc}$), $\boldsymbol{G}$ are reciprocal lattice vectors of the primitive cell, and $\Delta\widetilde{V}_{\rm L}$ are Fourier coefficients of the local perturbation potential (computed in a supercell with volume $\Omega_{\rm sup}$):

$$\Delta\widetilde{V}_{\rm L}(\boldsymbol{q}) = \frac{1}{\Omega_{\rm uc}}\int_{\Omega_{\rm sup}} d\boldsymbol{r}\;\Delta V_{\rm L}\left(\boldsymbol{r}\right)e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}, \quad (9)$$

where $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k} + \boldsymbol{G}$ is the transferred momentum in the $|n\boldsymbol{k}\rangle \to |n'\boldsymbol{k}'\rangle$ scattering process.

Through Eq. (8), we effectively separate the local matrix element calculation into two independent parts, the plane-wave matrix elements $\langle u_{n'\boldsymbol{k}'}|\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}}\,|u_{n\boldsymbol{k}}\rangle_{\rm uc}$, which are easily computed by integrating over the *primitive* cell, and the Fourier coefficients $\Delta\widetilde{V}_{\rm L}(\boldsymbol{q})$, which are computed in the supercell. In addition, since the local perturbation potential is smooth and decays in real space over a few angstroms, its Fourier coefficients $\Delta\widetilde{V}_{\rm L}(\boldsymbol{q})$ decay rapidly in reciprocal space, and the summation over $\boldsymbol{G}$ in Eq. (8) can be truncated to a small cutoff − typically, just a few reciprocal lattice vectors. For the same reason, the Fourier coefficients $\Delta\widetilde{V}_{\rm L}(\boldsymbol{q})$ can be efficiently interpolated (in our case, with a B-spline) at arbitrary $\boldsymbol{q}$ starting from their calculation at a few thousand $\boldsymbol{q}$-points in a cubic box centered at $\boldsymbol{q} = 0$. A great advantage of this formulation is that one can compute the local matrix elements using only the wavefunctions of the primitive cell (as opposed to those of the supercell), and effectively interpolate the perturbation potential to fine BZ grids.

The nonlocal matrix elements $M^{\rm NL}_{n'\boldsymbol{k}',n\boldsymbol{k}}$ are computed as the difference between the nonlocal potentials of a defect-containing and a pristine supercell:

$$\begin{aligned} M^{\rm NL}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \langle n'\boldsymbol{k}'|\hat{V}_{\rm NL}^{\rm (d)} - \hat{V}_{\rm NL}^{\rm (p)}|n\boldsymbol{k}\rangle \\ &= \langle n'\boldsymbol{k}'|\hat{V}_{\rm NL}^{\rm (d)}|n\boldsymbol{k}\rangle - \langle n'\boldsymbol{k}'|\hat{V}_{\rm NL}^{\rm (p)}|n\boldsymbol{k}\rangle\,, \end{aligned} \quad (10)$$

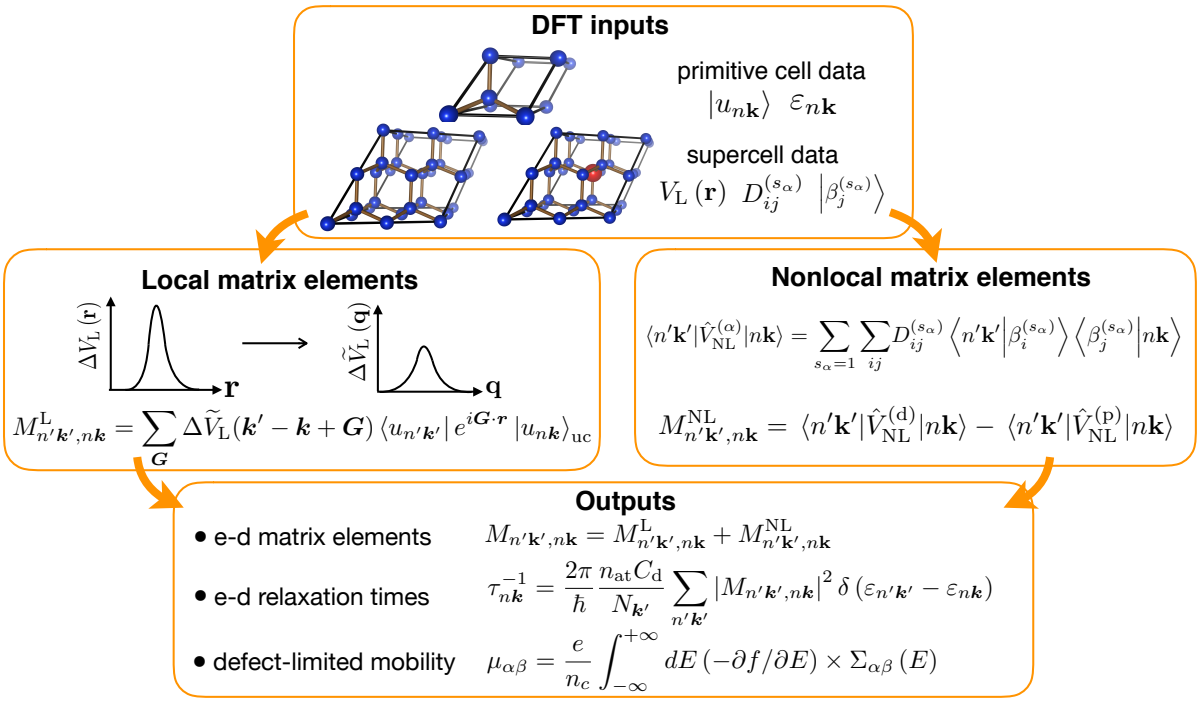

FIG. 1. Workflow for computing the e-d matrix elements and relaxation times, and the defect-limited mobility. In the first step, several inputs are computed with DFT, including the KS wavefunctions and eigenvalues of a primitive cell, and the local and nonlocal parts of the KS potential, separately in a pristine and a defect-containing supercell. The local matrix elements are then computed by Fourier transforming and interpolating the local perturbation potential, $\Delta V_{\mathrm{L}}(\boldsymbol{r})$, and combining it with the plane-wave matrix elements of the primitive cell. The nonlocal matrix elements are similarly computed by splitting the calculation into a primitive cell and a supercell part, using only the KS wavefunctions of the primitive cell. The total e-d matrix elements are then formed by adding the local and nonlocal parts, following which they are employed to compute as output the properties of interest, including the band and momentum-resolved e-d relaxation times and the defect-limited mobility.

where the matrix elements of $\hat{V}_{\mathrm{NL}}$ for each of the two supercells (labelled by $\alpha = \mathrm{d}, \mathrm{p}$), using Eq. (5), read:

$$\langle n'\boldsymbol{k}'|\hat{V}^{(\alpha)}_{\mathrm{NL}}|n\boldsymbol{k}\rangle = \sum_{s_\alpha=1}\sum_{ij} D_{ij}^{(s_\alpha)} \langle n'\boldsymbol{k}'|\beta_i^{(s_\alpha)}\rangle \langle \beta_j^{(s_\alpha)}|n\boldsymbol{k}\rangle\,. \tag{11}$$

Similar to the local matrix elements, the computation in Eq. (11) is split into a primitive cell and a supercell calculation, by expressing the scalar products $\langle\beta_j^{(s)}|n\boldsymbol{k}\rangle$ as (see Appendix C)

$$\langle\beta_j^{(s)}|n\boldsymbol{k}\rangle = \frac{1}{\sqrt{\Omega_{\mathrm{uc}}}}\sum_{\boldsymbol{G}} B^{(s)*}_{j\boldsymbol{k}}(\boldsymbol{G})\,\langle e^{i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_{\mathrm{uc}}\,, \tag{12}$$

where $B^{(s)}_{j\boldsymbol{k}}(\boldsymbol{G})$ is the Fourier coefficient of the KB projector $\beta_j^{(s)}$ (multiplied by the phase factor $e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$) at the primitive-cell reciprocal lattice vector $\boldsymbol{G}$:

$$B^{(s)}_{j\boldsymbol{k}}(\boldsymbol{G}) = \frac{1}{\sqrt{\Omega_{\mathrm{uc}}}}\int_{\Omega_{\mathrm{sup}}} d\boldsymbol{r}\,\left(\beta_j^{(s)}(\boldsymbol{r})e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\right)e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}. \tag{13}$$

Note that the nonlocal matrix elements, which are computed as the difference in Eq. (10), are nonzero both because the atomic positions change upon relaxing the structure in the defect-containing supercell and because the number and type of atoms, in general, differ in the two supercells, as is the case when considering a vacancy or impurity.

Figure 1 shows our workflow for computing properties related to e-d interactions from first principles. To obtain the relevant DFT inputs, the KS equations are solved in a primitive cell and separately in pristine and defect-containing supercells. The local and nonlocal matrix elements are then computed by splitting the calculation into a primitive cell and a supercell part, an approach that dramatically reduces the computational cost. Computing local matrix elements is the most expensive step, while the nonlocal matrix elements only involve relatively inexpensive reciprocal space sums. Importantly, only the KS wavefunctions, band structure and $\boldsymbol{k}$-points of the primitive cell are used, while the supercells are employed only to obtain the perturbation potential due to the defect. Once computed, the e-d matrix elements are employed to calculate the e-d RTs and the defect-limited mobility, among other quantities of interest. This approach allows us to systematically converge the RTs and other properties related to e-d interactions with respect to supercell size and BZ grids.

B. Comparison with the all-supercell method

In the all-supercell method [32], one uses the pristine and defect-containing supercells to provide all the necessary quantities for computing the e-d matrix elements, including the wavefunctions, band structure, perturbation potentials, and BZ grids. However, using supercell wavefunctions makes it challenging to compute and converge the e-d matrix elements and RTs, and ultimately to carry out accurate e-d calculations, since unconverged e-d RTs and transport properties can differ widely from the converged results.

All e-d calculations need to be converged with respect to supercell size; as we discuss below, converging the RTs for a neutral defect typically requires very large supercells with hundreds of atoms. In our approach, this convergence does not constitute a challenge since the same (primitive cell) wavefunctions are employed, regardless of supercell size. Conversely, in the all-supercell method, one uses wavefunctions from the pristine supercell, and the computational cost to compute and store the wavefunctions and obtain the matrix elements increases dramatically with supercell size, making accurate convergence tests too computationally demanding.

Let us analyze the cost of the most computationally intensive step, namely obtaining the local e-d matrix elements, $M^{\mathrm{L}}_{n'\boldsymbol{k}',n\boldsymbol{k}}$. Using a uniform BZ grid with $N_{\boldsymbol{k}}$ points, one obtains $\mathcal{O}(N_{\boldsymbol{k}}^2)$ matrix elements, each for a distinct $|n\boldsymbol{k}\rangle \to |n'\boldsymbol{k}'\rangle$ e-d scattering process. In a typical calculation, a uniform grid with at least $N_{\boldsymbol{k}} \approx 10^6$ points is needed to converge the RTs in the entire BZ. In a mobility calculation, one typically selects a small energy window of $\sim$100 meV near the band edges (in a semiconductor, or near the Fermi energy in a metal), which reduces the required number of $\boldsymbol{k}$-points to $N_{\boldsymbol{k}} \approx 10^4$.

In the all-supercell method, the local matrix elements are computed as:

$$M^{\mathrm{L}}_{n'\boldsymbol{k}',n\boldsymbol{k}} = \langle n'\boldsymbol{k}'|\Delta V_{\mathrm{L}}(\boldsymbol{r})|n\boldsymbol{k}\rangle_{\mathrm{sup}} \tag{14}$$

where the subscript (sup) denotes that both the local perturbation potential $\Delta V_{\mathrm{L}}(\boldsymbol{r})$ and the wavefunctions are obtained from a DFT calculation on a supercell. Since the cost of the DFT calculations scales with system size as roughly N_{at}^3, where N_{at} is the number of atoms in the supercell, computing $N_{\boldsymbol{k}}$ supercell wavefunctions, from which the local matrix elements are computed on the uniform grid, costs $N_{\boldsymbol{k}}N_{\mathrm{at}}^3$ in the all-supercell method.

By contrast, in our method only the primitive cell wavefunctions are used, and thus the computational cost of the matrix elements does not depend on N_{at} through the wavefunctions. To obtain the local matrix elements on the uniform grid with our method [see Eq. (8)], the only supercell data one needs are the Fourier coefficients $\Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{q})$ of the local perturbation potential. Obtaining these coefficients at a few thousand $\boldsymbol{q}$ points − from which an interpolation table can be constructed − has a cost that scales as N_{at}^3, but this step is required only once for a given supercell size. Therefore, computing the local matrix elements on a uniform BZ grid with $N_{\boldsymbol{k}}$ points has a cost of order N_{at}^3 in our method, versus a cost of $N_{\boldsymbol{k}}N_{\mathrm{at}}^3$ in the all-supercell method. For the typical mobility calculation mentioned above, this represents a speed-up by a factor of $N_{\boldsymbol{k}} \approx 10,000$ over the all-supercell method. Note that our mobility calculations are already expensive (tens of thousands of core-hours), so approaches that are thousands of times more expensive are clearly impractical. An additional benefit is that in our approach the large supercell wavefunctions are never stored or loaded into memory, so the speed up is significant even for computing a single e-d RT.

Finally, one would like to map the e-d scattering processes onto the band structure of the primitive cell, as is done for e-ph scattering processes. This is possible in our approach due to our use of primitive cell band structures and $\boldsymbol{k}$-point grids, but impractical in the all-supercell method, where one uses the supercell band structures and $\boldsymbol{k}$-point grids, which depend on the choice of a supercell and differ from those of the primitive cell due to nontrivial BZ folding effects. Due to its computational efficiency and convenience, we thus believe that our approach solves key technical challenges that have so far prevented efficient and accurate *ab initio* calculations of e-d interactions.

Figure 2 validates our approach, by comparing the local matrix elements computed with our method [using Eq. (8)] with those obtained with the all-supercell method using Eq. (14), for vacancy defects in silicon (see below for the computational details). It is seen that for a test case of a primitive cell and a $2\times2\times2$ supercell the results obtained with the two methods are in perfect agreement. This is but one of many benchmark tests we have performed.

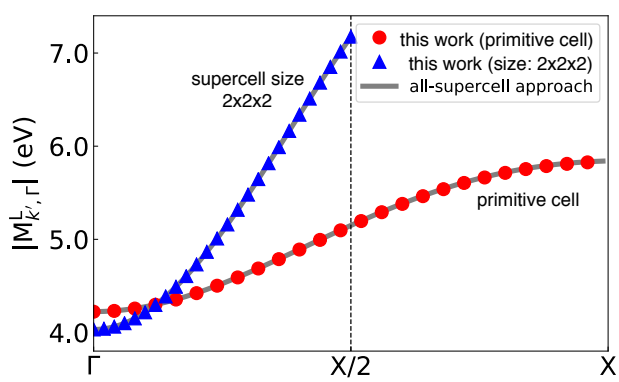

FIG. 2. Absolute value of the local e-d matrix elements, obtained from our approach and, for comparison, with the all-supercell method, for two cell sizes, a primitive cell and a $2\times2\times2$ supercell. The initial state is in the lowest valence band at Γ, and the final states are in the same band with crystal momenta $\boldsymbol{k}'$ along the $\Gamma-X$ high-symmetry line.

III. RESULTS

A. Computational details

We apply our approach to compute the e-d RTs and defect-limited mobility in silicon, separately for vacancy and (tetrahedral) interstitial defects. A defect concentration of 1 ppm (one defect in 10^6 atoms) is assumed in both cases. The ground states of the primitive cell and of supercells with size $N \times N \times N$ (where N is the number of primitive cells along each lattice vector) are computed using DFT within the local density approximation [36], using a plane wave basis and norm-conserving pseudopotentials [35] with the QUANTUM ESPRESSO package [37]. Briefly, for the primitive cell we use a lattice constant of 5.43 Å, a 40 Ry kinetic energy cutoff and a $12 \times 12 \times 12$ Monkhorst-Pack $\boldsymbol{k}$-point grid [38], converging the total energy to within 10 meV/atom; a consistent lattice constant and total energy convergence criterion is employed for the supercells. In the defect-containing supercells, the atomic forces are relaxed to within 25 meV/Å to account for the structural changes induced by the defect, and the resulting KS potentials are used to compute the e-d matrix elements. Due to the different reference potentials in the pristine and defect-containing supercells, we employ the core-average potential alignment method [39] to align the local potentials of the two supercells when computing the local perturbation potential; the reference potential is taken as the average of the local potential at the atom that is farthest from the defect site. In the e-d RT and mobility calculations, we select the electronic states of relevance in a small (∼100 meV) energy window near the band edges, and interpolate the band structure using maximally localized Wannier functions [40] with the WANNIER90 code [41, 42]. All e-d calculations have been implemented in our PERTURBO code [43] following the workflow in Fig. 1.

B. Relaxation times and their convergence

We first analyze the e-d RTs for vacancy defects in silicon, and study their convergence. Figure 3(a) gives the converged RTs (and their inverse, the scattering rates) of electrons and holes due to scattering with vacancy defects with a 1 ppm concentration. As the electron and hole energy increases away from the respective band edges, the scattering rates increase (and the RTs decrease) due to the increased phase space for scattering − that is, the larger number of final states that can be accessed in the scattering process. The latter is quantified by the density of states (DOS), which indeed shows a trend similar to the scattering rates. Overall, the e-d RTs are in the ps − ns range, and thus much longer than the typical e-ph RTs near room temperature, which are in the fs − ps range [44]. This result is consistent with the fact that e-ph scattering dominates near room temperature, while e-d scattering becomes important at low temperatures, where phonons are frozen out.

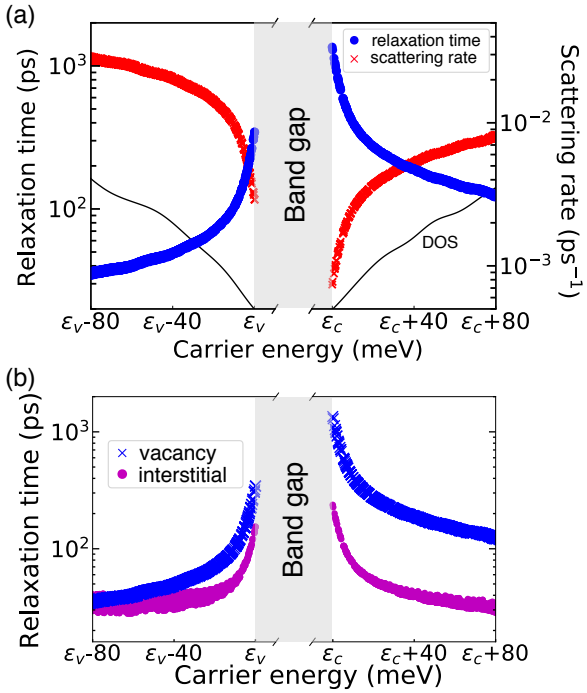

FIG. 3. (a) The relaxation times and their inverse, the scattering rates, for electrons and holes due to e-d scattering with vacancy defects in silicon. The valence band maximum is labeled ε_v and the conduction band minimum ε_c. The density of states (DOS) is also plotted, in arbitrary units. (b) Comparison between the relaxation times due to e-d scattering with vacancy and interstitial defects in silicon. The calculations use a 200^3 BZ grid with a 5 meV broadening, and a supercell size of $6\times6\times6$ (432 atoms) for vacancy and $8\times8\times8$ (1024 atoms) for interstitial defects.

We find that the RTs depend on carrier energy, type of carrier, and type of neutral defect, at odds with simplified empirical models employed for decades to model e-d scattering [17]. Figure 3(b) shows the RTs as a function of carrier energy in silicon for vacancy and interstitial defects, both with a 1 ppm concentration. For the vacancy defects, the electron RT is roughly 1 ns at the conduction band minimum and the hole RT is roughly 300 ps at the valence band maximum; both these RTs decrease by an order of magnitude ∼100 meV away from the band edges. Overall, the holes relax significantly faster than the electrons, both near the band edge and at higher energies, a result we attribute to the higher density of states near the three-fold degenerate valence band maximum [45]. A similar asymmetry in the electron and hole RTs has been predicted recently in e-ph scattering in GaN [45], where it was also attributed to the valence band degeneracy.

For interstitials, in contrast, we find that the RTs are similar for electrons and holes. They are 150 ps for electrons and 100 ps for holes at the respective band edges, and for both carriers the RTs approach a value of 30 ps roughly 100 meV away from the band edges. The unexpected energy, carrier type, and defect type dependence

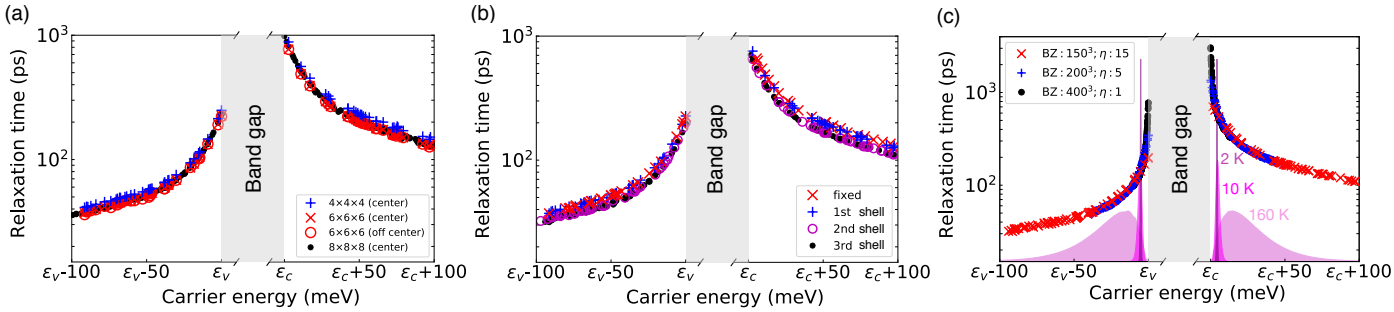

FIG. 4. Convergence of the e-d RTs, shown here for vacancy defects in silicon. We consider the effect of (a) supercell size, (b) structural relaxation, and (c) BZ grid and energy broadening in Eq. (1), where for each broadening value we use a converged BZ grid. Also shown in (c) is the function $(-\partial f/\partial E) \times \Sigma(E)$ in Eq. (16), at several temperatures in arbitrary units.

of the RTs cannot be explained by the widely used Erginsoy formula [17], which predicts an energy-independent RT for e-d scattering due to neutral defects. Different from the Erginsoy model, our *ab initio* calculations take the atomic and electronic structure into account, providing accurate results that are material and defect specific.

In *ab initio* e-d calculations, there are significant challenges with converging the RTs, which have so far not been examined in detail. This convergence is crucial since many transport properties and physical observables associated with e-d interactions depend sensitively on the RTs. Figure 4 shows how to systematically converge the e-d RTs with respect to three key factors – the supercell size, structural relaxation, and BZ grid used in the sum over final states in Eq. (1). This convergence study is discussed here for the vacancy case, although we find similar results for interstitial defects.

Figure 4(a) shows the convergence of the RTs with respect to supercell size. Results are given for supercell sizes ranging from $4 \times 4 \times 4$ to $8 \times 8 \times 8$, in each case containing one vacancy at the center of the supercell. To isolate the role of supercell size, the atomic structure is not relaxed in these calculations. The RTs in the $4 \times 4 \times 4$ and $6 \times 6 \times 6$ supercells are within 20% and 5%, respectively, of the $8 \times 8 \times 8$ supercell results, which can be considered fully converged. To verify that the RTs do not depend on defect position, we compute the RTs for an off-center vacancy that is placed away from the center of a $6 \times 6 \times 6$ supercell. For the same supercell size, the RTs of the off-center and centered vacancy match exactly, as they should.

Structural relaxation can be extensive around a defect and is expected to play an important role in accurately computing e-d interactions. Since it is costly to relax the structure in large supercells, an approximate scheme that retains accurate RTs is desirable. To this end, we compute the RTs in a supercell of a fixed size (here, $6 \times 6 \times 6$) in which only the atoms up to i-th nearest neighbor shell of the vacancy defect are relaxed, and those that are farther away are kept fixed. Figure 4(b) shows the RTs for structural relaxation up to the 1st, 2nd and 3rd nearest neighbor shell. We find that the RTs are almost converged for structural relaxation within the 2nd shell, and nearly identical to those for relaxation up to the 3rd shell, which can be considered converged since the atomic forces are negligible outside the 3rd nearest neighbor shell. The conclusion is that one needs to relax only a small portion of the atoms around the defect to accurately compute the e-d RTs.

Most critical when computing the e-d RTs is converging the $\boldsymbol{k}'$-point grid in the sum over final states in Eq. (1), which is equivalent to converging the grid of transferred momenta, $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k} + \boldsymbol{G}$. There is a cross-convergence effect between this grid and the energy broadening η employed to represent the delta function in Eq. (1), which is implemented as a normalized Gaussian with broadening, $\delta_\eta(x) = \frac{1}{\sqrt{2\pi}\eta} e^{-x^2/2\eta^2}$. The situation is fully analogous to converging the e-ph scattering rates [28]. Briefly, the broadening has to be small enough to not alter the final result, but the smaller the broadening the denser the $\boldsymbol{k}'$-point BZ grid needed to converge the sum in Eq. (1). Systematic convergence is achieved by starting with a small broadening (say, $\eta \approx 10$ meV) and converging the $\boldsymbol{k}'$-point BZ grid, and then decreasing the broadening to a smaller value and converging the BZ sum again. At convergence, the RTs do not change upon decreasing the broadening and converging the BZ sum. Note that the $\boldsymbol{k}'$-point grid can in principle be distinct from the $\boldsymbol{k}$-point grid at which the RTs are computed, but this is feasible in practice only if one has a mechanism to effectively interpolate the matrix elements. When this is possible, using random or importance sampling $\boldsymbol{k}'$-point grids can significantly speed up the calculations [44]. Here, in each calculation, we use the same uniform BZ grid for $\boldsymbol{k}$- and $\boldsymbol{k}'$-points, and refer to it below as the BZ grid (a uniform $M \times M \times M$ grid will be denoted as an M^3 grid).

Figure 4(c) shows the RTs for several values of the energy broadening η and gives the corresponding BZ grid at convergence. The BZ grid required to converge the RTs are denser for smaller values of the broadening; for η values of 1, 5 and 15 meV, uniform BZ grids with 400^3, 200^3 and 150^3 points are needed, respectively. For electron energies higher than 25 meV above the conduction

band minimum, a 15 meV broadening and a 150^3 BZ grid are sufficient to converge the RTs. For electron energies within 25 meV of the band edge, a 5 meV broadening with a 200^3 BZ grid gives the same RTs as a smaller 1 meV broadening with a 400^3 BZ grid [46]. The broadening and BZ grid values at convergence are similar for electrons and holes, and for vacancies and interstitials.

Importantly, the RTs computed with unconverged grids can differ widely from the converged values, especially at energies near the band edges [see Fig. 4(c)], which critically contribute to charge transport. It is therefore essential to have an efficient method for computing and converging the e-d RTs on fine BZ grids to accurately compute charge transport at low temperature.

C. Defect-limited carrier mobility

At room temperature, where e-ph interactions typically dominate, charge transport can be accurately predicted from first principles in several families of materials [26–30]. However, many devices and experiments operate at low temperature, where charge transport is governed by e-d interactions. It is thus critically important to develop *ab initio* calculations that can predict carrier dynamics at low temperature in the presence of e-d scattering. We compute the defect-limited carrier mobility μ at temperature T within the RT approximation of the Boltzmann transport equation [44]:

$$\mu_{\alpha\beta}(T) = \frac{e}{n_c}\int_{-\infty}^{+\infty} dE\,\left[-\partial f(T,E)/\partial E\right]\times\Sigma_{\alpha\beta}(E)\,, \tag{15}$$

where e is the electron charge, n_c the carrier concentration, $f(T,E)$ the Fermi-Dirac distribution, and $\Sigma(E)$ the transport distribution function (TDF) at energy E,

$$\Sigma_{\alpha\beta}(E) = \frac{2}{\Omega_{\rm uc}}\sum_{n\boldsymbol{k}}\tau_{n\boldsymbol{k}}\mathbf{v}^{\alpha}_{n\boldsymbol{k}}\mathbf{v}^{\beta}_{n\boldsymbol{k}}\delta\left(E-\varepsilon_{n\boldsymbol{k}}\right), \tag{16}$$

where α and β are Cartesian directions. The TDF is computed with a tetrahedron integration method [28], using converged e-d RTs and Wannier-interpolated band velocities $\mathbf{v}_{n\boldsymbol{k}}$ [41, 42]. To estimate the carrier energy range contributing significantly to the mobility, we plot the integrand of the mobility formula in Eq. (15), the function $(-\partial f/\partial E)\times\Sigma(E)$, in Fig. 4(c) for temperatures of 2, 10 and 160 K. As the temperature increases, the peak of the function broadens and moves away in energy from the band edges, indicating that the energy region contributing to the mobility shifts to higher carrier energies. The most stringent conditions for computing the RTs are below 10 K, where the contribution to the mobility peaks 5 meV away from the band edge; in this regime, BZ grids as dense as 200^3 $\boldsymbol{k}'$-points and a broadening of 5 meV are needed to accurately compute the mobility.

We compute the mobility of electrons and holes in silicon, considering separately vacancy and interstitial defects. Figure 5 shows the computed mobility curves at temperatures below 150 K. The electron and hole mobilities for vacancy defects are higher than the corresponding mobilities for interstitial defects due to the longer e-d RTs for vacancies. We find that, in all cases, the defect-limited mobility is roughly constant below 10 K, and decreases at higher temperatures. Note that e-ph interactions are not included here, so these trends are due solely to the e-d interactions. Above 50 K, the temperature dependence of the mobility is well approximated by a power law, $\mu \propto T^{-\alpha}$, with coefficients α of order 0.5. For vacancy defects, the best-fit values of α are 0.53 for electrons and 0.55 for holes, and for interstitials 0.46 for electrons and 0.38 for holes.

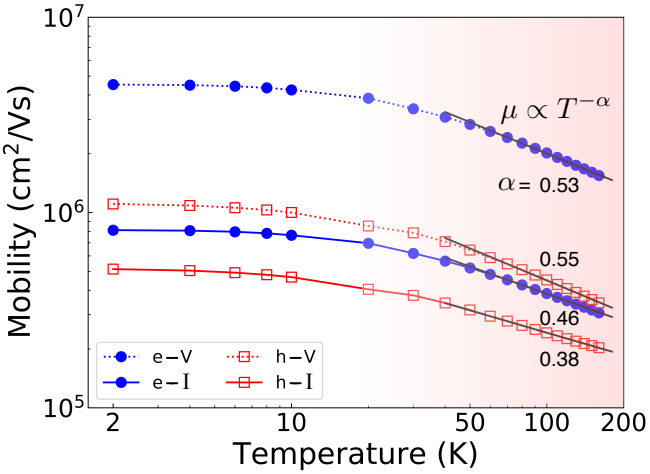

FIG. 5. Defect-limited mobilities in silicon, as a function of temperature below 150 K. Shown are the results for electron-vacancy (e−V), hole-vacancy (h−V), electron-interstitial (e−I), and hole-interstitial (h−I) interactions. The electron (solid circles) and hole (empty squares) mobilities are given for vacancy (dotted line) and interstitial (solid line) defects. The temperature dependence of the mobility above 50 K follows approximately a power law, $\mu \propto T^{-\alpha}$, with coefficients α given in figure. A defect concentration of 1 defect in 10^6 atoms is assumed.

An interesting interpretation, which is particularly apt for vacancies, is that the defects can be regarded as a substance added to the pure crystal to make an "alloy". One thus expects that the temperature dependence of the carrier mobility for defect scattering is similar to that of alloy scattering, for which a power law with $\alpha = 0.5$ is expected based on existing models [47, 48]. Our results above 50 K are consistent with this interpretation, but we additionally find that the value of the exponent α depends on carrier and defect type. By contrast, the Erginsoy formula [17] predicts a temperature independent mobility for neutral defect scattering, which is clearly inconsistent with our results, and also with experiment.

An early experiment [49] on n-type doped silicon obtained the mobility due to neutral impurity scattering by subtracting the lattice and ionized impurity contributions. The resulting mobility, which is limited by neutral

defect scattering alone, decreases above 50 K, in agreement with our results. While this trend has been attributed by the authors to inelastic e-d scattering [49], our results show that it can be explained by accurately computing elastic e-d scattering. The mobility decrease is due to the energy dependence of the RTs − as the temperature increases, so does the average energy of the electrons contributing to the mobility [see Fig. 4(c)], and their average RTs decrease as we have shown, causing the mobility decrease with temperature.

The mobility computed here using e-d interactions with only neutral point defects can be seen as an upper bound corresponding to an ideally pure material. In silicon, the mobility in the purest crystals (with impurity concentrations of $\sim 10^{12}$ cm^{-3}) is roughly 10^6 cm^2/Vs at 10 K, a value that can be explained by ionized impurity scattering [50]. In these pure silicon samples, the concentration of neutral defects can also be as low as 10^{12} cm^{-3} ($C_{\mathrm{d}} \approx 10^{-10}$ in dimensionless units) [50]. We extend our results to this lower neutral defect concentration, and estimate a mobility limit in silicon of $\sim 10^{10}$ cm^2/Vs at 10 K for an ideal scenario in which ionized impurity scattering is absent and only neutral point defects scatter the carriers. This mobility limit is higher than the value measured in samples with ionized impurity scattering, since the latter is much stronger than neutral defect scattering due to its long-range character. More extensive quantitative comparisons between computed and measured low-temperature charge transport data will be the subject of future work. We conclude that our approach is a powerful tool to compute charge transport at low temperature and estimate mobility limit values in materials.

IV. DISCUSSION

We close by discussing technical remarks and future research directions. Several improvements are possible to our calculations of e-d RTs. Similar to e-ph calculations, interpolating the e-d matrix elements, for example using Wannier functions or atomic orbitals [25], would be desirable, since one could compute the primitive cell wavefunctions only on coarse grids rather than on the fine grids needed to converge the RTs. In addition, since the broadening needed to converge the RTs increases with increasing carrier energy, using an adaptive broadening scheme could significantly speed up the RT calculations. One could use relatively coarse BZ grids and larger broadening values at higher carrier energies to save computational time, while using finer grids (and a smaller broadening) only at low carrier energy.

The method presented in this work can be extended in several ways. While our calculations focus on neutral point defects, our method can be generalized to treat charged defects, by adding their long-range Coulomb interaction to the local perturbation potential in reciprocal space, similar to what is done in e-ph calculations [28]. The method can also be extended, using spin-polarized DFT calculations, to treat e-d scattering processes involving spin and spin-orbit coupling. Example applications include spin-flip processes due to magnetic impurities and defects in topological materials.

The proposed e-d calculations are general since they take into account the atomic structure of the material (including important structural relaxation effects around the defect) and its electronic structure. Unlike empirical models, there is no particular extension needed to treat different types of point and extended defects or different materials, provided one can afford the large DFT calculations needed to obtain the perturbation potentials. For example, our method is suitable for extended defects such as dislocations or grain boundaries, but to study them one may need supercells with thousands of atoms.

While this work focuses on the lowest order of perturbation theory, by including higher order e-d interactions one could investigate a wide range of low-temperature phenomena, including weak localization or antilocalization and universal conductance fluctuations [51, 52]. Our method can also be a starting point for efficient inelastic e-d scattering calculations [53]. *Ab initio* calculations of e-d interactions are still in their infancy, and more work is needed to develop their potential and expand their scope.

V. CONCLUSION

We have presented an efficient approach that overcomes main technical challenges for *ab initio* calculations of e-d interactions. The method is applied to compute and systematically converge the elastic e-d RTs and the associated defect-limited carrier mobility below 150 K for vacancy and interstitial defects in silicon. The RTs exhibit a pronounced dependence on energy, carrier and defect type, and the defect-limited mobility is temperature dependent. These results cannot be explained using widely used empirical models of e-d interactions. Our approach can provide new microscopic insight into e-d scattering processes. It is broadly applicable and can be generalized to treat charged defects, magnetic impurities, and extended defects. We expect that this work will lay a solid foundation for efficient *ab initio* calculations of e-d interactions.

ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research through the Young Investigator Program Grant FA9550-18-1-0280. J.-J. Z. was supported by the National Science Foundation under Grant No. ACI-1642443, which provided for code development. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. I-T. L. thanks Dr. Luis Agapito, Dr. Davide Sangalli, Vatsal Jhalani, Jinsoo Park and Xiao Tong for fruitful discussions.

APPENDIX A: ELASTIC AND INCOHERENT ELECTRON-DEFECT SCATTERING RATE

The scattering rate $\Gamma_{n\boldsymbol{k}}$ can be written using Fermi's golden rule within lowest-order perturbation theory:

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar} \sum_{n'\boldsymbol{k}'} \left| \langle \psi_{n'\boldsymbol{k}'} | \Delta\hat{H} | \psi_{n\boldsymbol{k}} \rangle \right|^2 \delta\left(\varepsilon_{n'\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}\right), \quad \text{(A1)}$$

where the perturbation $\Delta\hat{H}$ is the difference between the Kohn-Sham Hamiltonian $\hat{H}^{(\mathrm{d})}$ of a crystal containing N_d identical defects and the Hamiltonian $\hat{H}^{(\mathrm{p})}$ of the same crystal with no defects, namely, $\Delta\hat{H} = \hat{H}^{(\mathrm{d})} - \hat{H}^{(\mathrm{p})}$. The crystal is made up by $N_{\boldsymbol{k}}$ primitive cells, and we apply Born-von Karman (BvK) periodic boundary conditions; the crystal volume is $\Omega_\mathrm{BvK} = N_{\boldsymbol{k}}\Omega_\mathrm{uc}$, where Ω_uc is the volume of the primitive cell. Above, $|\psi_{n\boldsymbol{k}}\rangle$ are unperturbed Bloch wavefunctions with energy $\varepsilon_{n\boldsymbol{k}}$, which in coordinate space read:

$$\langle \boldsymbol{r} | \psi_{n\boldsymbol{k}} \rangle = \frac{1}{\sqrt{N_{\boldsymbol{k}}}} \langle \boldsymbol{r} | n\boldsymbol{k} \rangle = \frac{1}{\sqrt{N_{\boldsymbol{k}}}} u_{n\boldsymbol{k}}(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \quad \text{(A2)}$$

where $|n\boldsymbol{k}\rangle$ is the Bloch wavefunction without the prefactor, and $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is the periodic part of the Bloch wavefunction, normalized in the primitive cell as

$$\int_{\Omega_\mathrm{uc}} d^3r \; u^*_{n\boldsymbol{k}}(\boldsymbol{r}) u_{n\boldsymbol{k}}(\boldsymbol{r}) = 1. \quad \text{(A3)}$$

Since the kinetic energy is the same in the pristine and defect-containing systems, the difference of their Hamiltonians equals the sum of the perturbations due to all defects:

$$\Delta\hat{H} = \hat{H}^{(\mathrm{d})} - \hat{H}^{(\mathrm{p})} = \sum_{i=1}^{N_\mathrm{d}} \Delta V_{\mathrm{e-d}}(\boldsymbol{r} - \boldsymbol{r}_i), \quad \text{(A4)}$$

where $\Delta V_{\mathrm{e-d}}(\boldsymbol{r} - \boldsymbol{r}_i)$ denotes the perturbation potential due to a defect located at $\boldsymbol{r}_i$, and we consider non-interacting defects of the same kind. Assuming that the scattering events are independent, we can write the scattering rate for elastic and incoherent scattering processes due to all defects as

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar} \frac{n_\mathrm{at} C_\mathrm{d}}{N_{\boldsymbol{k}'}} \sum_{n'\boldsymbol{k}'} \left| M_{n'\boldsymbol{k}',n\boldsymbol{k}} \right|^2 \delta\left(\varepsilon_{n'\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}\right), \quad \text{(A5)}$$

where n_at is the number of atoms in the primitive cell, C_d, which is formally equal to $N_\mathrm{d}/\left(N_{\boldsymbol{k}} \times n_\mathrm{at}\right)$, is in practice an assumed value of the defect concentration, and $M_{n'\boldsymbol{k}',n\boldsymbol{k}}$ is defined as the e-d matrix element for the perturbation due to a single defect:

$$M_{n'\boldsymbol{k}',n\boldsymbol{k}} = \langle n'\boldsymbol{k}' | \Delta V_{\mathrm{e-d}} | n\boldsymbol{k} \rangle. \quad \text{(A6)}$$

Within our approximations, the scattering rate is proportional to the defect concentration, and can be computed at any desired value of the defect concentration, provided that the scattering events remain uncorrelated and the defects non-interacting throughout the concentration range of interest.

APPENDIX B: LOCAL MATRIX ELEMENTS

The local matrix elements can be written as:

$$\begin{aligned} M^\mathrm{L}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \langle n'\boldsymbol{k}' | \Delta V_\mathrm{L}(\boldsymbol{r}) | n\boldsymbol{k} \rangle \\ &= \int_{\Omega_\mathrm{BvK}} d^3r \; u^*_{n'\boldsymbol{k}'}(\boldsymbol{r}) e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}} \Delta V_\mathrm{L}(\boldsymbol{r}) u_{n\boldsymbol{k}}(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \end{aligned} \quad \text{(B1)}$$

We define the forward and inverse Fourier transforms of the local perturbation potential, respectively, as

$$\Delta V_\mathrm{L}(\boldsymbol{r}) = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{q}} \Delta\widetilde{V}_\mathrm{L}(\boldsymbol{q}) e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \quad \text{(B2)}$$

and

$$\begin{aligned} \Delta\widetilde{V}_\mathrm{L}(\boldsymbol{q}) =& \frac{1}{\Omega_\mathrm{uc}} \int_{\Omega_\mathrm{BvK}} d^3r \; \Delta V_\mathrm{L}(\boldsymbol{r}) e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \\ \approx& \frac{1}{\Omega_\mathrm{uc}} \int_{\Omega_\mathrm{sup}} d^3r \; \Delta V_\mathrm{L}(\boldsymbol{r}) e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}, \end{aligned} \quad \text{(B3)}$$

where in the last line we replace the crystal volume Ω_BvK with the supercell volume Ω_sup, using the fact that the local perturbation potential vanishes at the supercell boundary, which is typically the case for supercells larger than a few primitive cells due to the localized nature of the perturbation potential. Inserting Eq. (B2) into Eq. (B1) and using the translational invariance of $u_{n\boldsymbol{k}}(\boldsymbol{r})$, we have

$$M^\mathrm{L}_{n'\boldsymbol{k}',n\boldsymbol{k}} = \sum_{\boldsymbol{G}} \Delta\widetilde{V}_\mathrm{L}(\boldsymbol{k}' - \boldsymbol{k} + \boldsymbol{G}) \langle u_{n'\boldsymbol{k}'} | e^{i\boldsymbol{G}\cdot\boldsymbol{r}} | u_{n\boldsymbol{k}} \rangle_\mathrm{uc}, \quad \text{(B4)}$$

with the plane-wave matrix elements defined as

$$\langle u_{n'\boldsymbol{k}'} | e^{i\boldsymbol{G}\cdot\boldsymbol{r}} | u_{n\boldsymbol{k}} \rangle_\mathrm{uc} = \int_{\Omega_\mathrm{uc}} d^3r \; u^*_{n'\boldsymbol{k}'}(\boldsymbol{r}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}}(\boldsymbol{r}), \quad \text{(B5)}$$

where $\boldsymbol{G}$ are reciprocal lattice vectors of the primitive cell. This formula is valid for any basis set. Here we use a plane wave basis for $u_{n\boldsymbol{k}}(\boldsymbol{r})$, and write

$$u_{n\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega_\mathrm{uc}}} \sum_{\boldsymbol{G}} C_{n\boldsymbol{k}}(\boldsymbol{G}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \quad \text{(B6)}$$

where $C_{n\boldsymbol{k}}(\boldsymbol{G})$ is the Fourier coefficient of $u_{n\boldsymbol{k}}(\boldsymbol{r})$ at the reciprocal lattice vector $\boldsymbol{G}$. The local matrix element formula in a plane wave basis becomes:

$$\begin{aligned} M^\mathrm{L}_{n'\boldsymbol{k}',n\boldsymbol{k}} =& \sum_{\boldsymbol{G}} \Delta\widetilde{V}_\mathrm{L}(\boldsymbol{k}' - \boldsymbol{k} + \boldsymbol{G}) \\ &\times \left[\sum_{\mathbf{G}''} \sum_{\mathbf{G}'} C^*_{n'\boldsymbol{k}'}(\mathbf{G}'') C_{n\boldsymbol{k}}(\mathbf{G}') \delta_{\mathbf{G}'',\mathbf{G}'+\boldsymbol{G}} \right]. \end{aligned} \quad \text{(B7)}$$

This is the formula implemented in our code and used in this work.

APPENDIX C: NONLOCAL MATRIX ELEMENTS

To obtain an expression for the nonlocal matrix elements, we focus on the scalar product

$$\begin{aligned}\langle \beta_j^{(s)} \,|\, n\boldsymbol{k}\rangle &= \int_{\Omega_{\mathrm{BvK}}} d^3r \beta_j^* \left(\boldsymbol{r}-\tau_s\right) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}}\left(\boldsymbol{r}\right) \\ &= \int_{\Omega_{\mathrm{BvK}}} d^3r \left[\frac{\beta_j\left(\boldsymbol{r}-\tau_s\right)}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\right]^* \sqrt{\Omega_{\mathrm{uc}}} u_{n\boldsymbol{k}}\left(\boldsymbol{r}\right).\end{aligned} \tag{C1}$$

We first fix the atomic position at the origin (by setting $\tau_s = 0$) and then generalize the result to arbitrary atomic positions. We define generalized forward and inverse Fourier transforms of the Kleinman-Bylander (KB) projectors, respectively, as:

$$\frac{\beta_j\left(\boldsymbol{r}\right)}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} = \frac{1}{\Omega_{\mathrm{BvK}}} \sum_{\boldsymbol{q}} B_{j\boldsymbol{k}}\left(\boldsymbol{q}\right) e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \tag{C2}$$

and

$$\begin{aligned}B_{j\boldsymbol{k}}\left(\boldsymbol{q}\right) &= \int_{\Omega_{\mathrm{BvK}}} d^3r \left[\frac{\beta_j\left(\boldsymbol{r}\right)}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\right] e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \\ &= \int_{\Omega_{\mathrm{sup}}} d^3r \left[\frac{\beta_j\left(\boldsymbol{r}\right)}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\right] e^{-i\boldsymbol{q}\cdot\boldsymbol{r}},\end{aligned} \tag{C3}$$

where we replace the crystal volume Ω_{BvK} with the supercell volume Ω_{sup} because the KB projector is localized around the core atomic region. If a general atomic position τ_s is chosen, the Fourier coefficient $B_{j\boldsymbol{k}}^{(s)}\left(\boldsymbol{q}\right)$ becomes, using the properties of the Fourier transforms,

$$B_{j\boldsymbol{k}}^{(s)}\left(\boldsymbol{q}\right) = e^{-i(\boldsymbol{k}+\boldsymbol{q})\cdot\tau_s} B_{j\boldsymbol{k}}\left(\boldsymbol{q}\right). \tag{C4}$$

The scalar product in Eq. (C1), after inserting Eq. (C2) into Eq. (C1), becomes

$$\langle \beta_j^{(s)} \,|\, n\boldsymbol{k}\rangle = \frac{1}{\sqrt{\Omega_{\mathrm{uc}}}} \sum_{\boldsymbol{G}} B_{j\boldsymbol{k}}^{(s)*}\left(\boldsymbol{G}\right) \langle e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \,|\, u_{n\boldsymbol{k}}\rangle_{\mathrm{uc}}, \tag{C5}$$

where we used the translational invariance of $u_{n\boldsymbol{k}}\left(\boldsymbol{r}\right)$. This formula is valid for any basis set. Here we use a plane wave basis, so the matrix elements of $\hat{V}_{\mathrm{NL}}$, for each of the pristine and defect-containing supercells (labelled by $\alpha = \mathrm{d}, \mathrm{p}$), read:

$$\begin{aligned}\langle n'\boldsymbol{k}'|\hat{V}_{\mathrm{NL}}^{(\alpha)}|n\boldsymbol{k}\rangle &= \sum_{s_\alpha=1} \sum_{ij} D_{ij}^{(s_\alpha)} \times \\ &\left[\sum_{\mathbf{G}'} B_{i\boldsymbol{k}'}^{(s_\alpha)}(\mathbf{G}') \, C_{n'\boldsymbol{k}'}^*(\mathbf{G}')\right] \left[\sum_{\boldsymbol{G}} B_{j\boldsymbol{k}}^{(s_\alpha)*}(\boldsymbol{G}) \, C_{n\boldsymbol{k}}(\boldsymbol{G})\right],\end{aligned} \tag{C6}$$

where we use the same notation as in Eq. (11). The nonlocal matrix elements are computed as the difference in Eq. (10). This is the formula implemented in our code and used in this work.

* bmarco@caltech.edu

[1] B. Radisavljevic and A. Kis, Nat. Mater. **12**, 815 (2013).
[2] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nat. Nanotech. **7**, 699 (2012).
[3] S.-L. Li, K. Tsukagoshi, E. Orgiu, and P. Samorì, Chem. Soc. Rev. **45**, 118 (2016).
[4] L. M. Herz, ACS Energy Lett. **2**, 1539 (2017).
[5] J. Mao, J. Shuai, S. Song, Y. Wu, R. Dally, J. Zhou, Z. Liu, J. Sun, Q. Zhang, C. dela Cruz, S. Wilson, Y. Pei, D. J. Singh, G. Chen, C.-W. Chu, and Z. Ren, Proc. Natl. Acad. Sci. U.S.A. **114**, 10548 (2017).
[6] P. Sun, B. Wei, J. Zhang, J. M. Tomczak, A. M. Strydom, M. Søndergaard, B. B. Iversen, and F. Steglich, Nat. Commun. **6**, 7425 (2015).
[7] Y. Xu, J. Chiu, L. Miao, H. He, Z. Alpichshev, A. Kapitulnik, R. R. Biswas, and L. A. Wray, Nat. Commun. **8**, 14081 (2017).
[8] P. M. Koenraad and M. E. Flatté, Nat. Mater. **10**, 91 (2011).
[9] D. Gunlycke and C. T. White, Phys. Rev. Lett. **106**, 136806 (2011).
[10] O. Yazyev and S. G. Louie, Nat. Mater. **9**, 806 (2010).
[11] G. Bergmann, Phys. Rep. **107**, 1 (1984).
[12] B. L. Altshuler, D. Khmel'nitzkii, A. I. Larkin, and P. A. Lee, Phys. Rev. B **22**, 5142 (1980).
[13] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge university press, 1997).
[14] P. A. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).
[15] E. Akkermans and G. Montambaux, *Mesoscopic Physics of Electrons and Photons* (Cambridge university press, 2007).
[16] D. Chattopadhyay and H. J. Queisser, Rev. Mod. Phys. **53**, 745 (1981).
[17] C. Erginsoy, Phys. Rev. **79**, 1013 (1950).
[18] N. Sclar, Phys. Rev. **104**, 1559 (1956).
[19] J. S. Blakemore, Phys. Rev. B **22**, 743 (1980).
[20] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2004).
[21] M. Bernardi, D. Vigil-Fowler, J. Lischner, J. B. Neaton, and S. G. Louie, Phys. Rev. Lett. **112**, 257402 (2014).
[22] M. Bernardi, D. Vigil-Fowler, C. S. Ong, J. B. Neaton, and S. G. Louie, Proc. Natl. Acad. Sci. U.S.A. **112**, 5291 (2015).
[23] M. Bernardi, J. Mustafa, J. B. Neaton, and S. G. Louie, Nat. Commun. **6**, 7044 (2015).
[24] H. Tanimura, J. Kanasaki, K. Tanimura, J. Sjakste, N. Vast, M. Calandra, and F. Mauri, Phys. Rev. B **93**, 161203 (2016).

[25] L. A. Agapito and M. Bernardi, Phys. Rev. B **97**, 235146 (2018).

[26] J. I. Mustafa, M. Bernardi, J. B. Neaton, and S. G. Louie, Phys. Rev. B **94**, 155105 (2016).

[27] C. H. Park, N. Bonini, T. Sohier, G. Samsonidze, B. Kozinsky, M. Calandra, F. Mauri, and N. Marzari, Nano Lett. **14**, 1113 (2014).

[28] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201 (2016).

[29] N.-E. Lee, J.-J. Zhou, L. A. Agapito, and M. Bernardi, Phys. Rev. B **97**, 115203 (2018).

[30] J.-J. Zhou, O. Hellman, and M. Bernardi, Phys. Rev. Lett. **121**, 226603 (2018).

[31] M. H. Evans, X.-G. Zhang, J. D. Joannopoulos, and S. T. Pantelides, Phys. Rev. Lett. **95**, 106802 (2005).

[32] O. D. Restrepo, K. Varga, and S. T. Pantelides, Appl. Phys. Lett. **94**, 212103 (2009).

[33] V. Lordi, P. Erhart, and D. Åberg, Phys. Rev. B **81**, 235204 (2010).

[34] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[35] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[36] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[37] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[38] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[39] Y. Kumagai and F. Oba, Phys. Rev. B **89**, 195205 (2014).

[40] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[41] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).

[42] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014).

[43] The code developed in this work will be released in the future at `http://perturbo.caltech.edu/`.

[44] M. Bernardi, Eur. Phys. J. B **89**, 239 (2016).

[45] V. A. Jhalani, J. J. Zhou, and M. Bernardi, Nano Lett. **17**, 5012 (2017).

[46] Converging the RTs within a few meV of the band edges may require even smaller broadenings and denser BZ grids, but it is not necessary for mobility calculations, even at very low temperatures, because the carrier velocity vanishes at the band edges.

[47] M. P. Vaughan, F. Murphy-Armando, and S. Fahy, Phys. Rev. B **85**, 165209 (2012).

[48] C. Hamaguchi, *Basic Semiconductor Physics* (Springer-Verlag Berlin Heidelberg, 2010).

[49] P. Norton, T. Braggins, and H. Levinstein, Phys. Rev. B **8**, 5632 (1973).

[50] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, Phys. Rev. B **12**, 2265 (1975).

[51] H. Bruus and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction* (Oxford university press, 2004).

[52] G. D. Mahan, *Many-Particle Physics* (Springer US, 2000).

[53] G. D. Barmparis, Y. S. Puzyrev, X. G. Zhang, and S. T. Pantelides, Phys. Rev. B **92**, 214111 (2015).