Wannier Function Perturbation Theory: Localized Representation and Interpolation of Wave Function Perturbation

Jae-Mo Lihm^{*} and Cheol-Hwan Park[†]

*Center for Correlated Electron Systems, Institute for Basic Science, Seoul 08826, Korea;
Department of Physics and Astronomy, Seoul National University, Seoul 08826, Korea;
and Center for Theoretical Physics, Seoul National University, Seoul 08826, Korea*

(Received 16 July 2021; revised 7 October 2021; accepted 8 October 2021; published 16 December 2021)

Thanks to the nearsightedness principle, the low-energy electronic structure of solids can be represented by localized states such as the Wannier functions. Wannier functions are actively being applied to a wide range of phenomena in condensed matter systems. However, the Wannier-function-based representation is limited to a small number of bands and thus cannot describe the change of wave functions due to various kinds of perturbations, which require sums over an infinite number of bands. Here, we introduce the concept of the Wannier function perturbation, which provides a localized representation of wave function perturbations. Wannier function perturbation theory allows efficient calculation of numerous quantities involving wave function perturbation, among which we provide three applications. First, we calculate the temperature-dependent indirect optical absorption spectra of silicon near the absorption edge nonadiabatically, i.e., differentiating phonon-absorption and phonon-emission processes, and without arbitrary temperature-dependent shifts in energy. Second, we establish a theory to calculate the shift spin conductivity without any band-truncation error. Unlike the shift charge conductivity, an exact calculation of the shift spin conductivity is not possible within the conventional Wannier function methods because it cannot be obtained from geometric quantities for low-energy bands. We apply the theory to monolayer WTe_2 . Third, we calculate the spin Hall conductivity of the same material again without any band-truncation error. Wannier function perturbation theory is a versatile method that can be readily applied to calculate a wide range of quantities related to various kinds of perturbations.

DOI: [10.1103/PhysRevX.11.041053](https://doi.org/10.1103/PhysRevX.11.041053)

Subject Areas: Computational Physics
Condensed Matter Physics
Optoelectronics

I. INTRODUCTION

Most physical and chemical systems are “nearsighted” [1,2]: Their properties can be studied by inspecting one local region at a time, and the long-range effects can be easily approximated. A localized representation of a system is highly desirable because it enables one to exploit the nearsightedness to its full extent. Wannier functions (WFs) provide a localized representation of wave functions [3,4]. The invention of the maximally localized Wannier function (MLWF) method [5,6] enabled the calculation of the localized WFs for electronic structures computed using density functional theory (DFT). Using the MLWFs, the

low-energy eigenvalues and eigenstates can be interpolated very efficiently.

Applications of MLWFs are stretched over all fields of condensed matter physics and materials science [7]. MLWFs are used to efficiently compute electron-phonon coupling (EPC) at a dense grid of electron and phonon wave vectors [8]. Wannier interpolation of EPC is actively being applied to study transport [9,10], phonon-assisted optical absorption [11], superconductivity [12], and ultrafast carrier dynamics [13]. Various linear and nonlinear responses of solids to dc and ac electric fields are also routinely calculated using Wannier interpolation [14–19].

However, one of the most severe limitations of WF interpolation is the limited number of bands being interpolated. Consequently, WFs cannot represent the wave function perturbation, which is the sum of contributions from an infinite number of bands. Obtaining MLWFs for the high-energy, plane-wave-like bands is a highly nontrivial problem. Even assuming that the MLWFs are obtained, calculations in the MLWF basis will be much more costly, negating the efficiency of Wannier interpolation. In DFT

^{*}jaemo.lihm@gmail.com
[†]cheolhwan@snu.ac.kr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

calculations, the wave function perturbation can be calculated without referring to the high-energy states by solving a linear equation known as the Sternheimer equation [20]. This method forms the basis of the calculation of the linear response properties of solids using density functional perturbation theory (DFPT) [21]. However, the Sternheimer equation is not directly applicable to WFs since the WFs do not accurately represent the wave function perturbation. Therefore, WF methods cannot deal with properties that require the calculation of wave function perturbations. Ge *et al.* [22] developed a local representation of wave function perturbation for the dielectric response of insulators. However, this formalism cannot be applied to WFs for entangled bands [6], while such generalization is necessary for many applications.

A representative example where the interpolation of wave function perturbations is needed is the real part of the phonon-induced electron self-energy. This self-energy causes the temperature dependence and zero-point renormalization of electron band structures [23,24]. The state-of-the-art method for the perturbative calculation of the real part of the self-energy is to solve the Sternheimer equation using a complete basis set such as the plane waves [24]. However, such a calculation is too costly to be done on a large number of k points. The high computational cost prohibits the consideration of phonon-induced renormalization at a dense k -point mesh in the study of optical and transport properties of materials. Also, band disentanglement [6] is necessary even for the valence band of insulators because the wave functions and energy eigenvalues of the low-lying conduction bands are required.

Another example of a quantity involving the wave function perturbation is the shift current [25]. For the shift charge current, the wave function perturbation can alternatively be expressed as a derivative with respect to the crystal momentum $\mathbf{k}$, allowing an exact WF-based calculation [16]. However, such an alternative expression cannot be made for the shift spin current in spin-orbit coupled systems [26]. Previous calculations of shift spin current in spin-orbit coupled systems truncated the infinite sum over bands [27,28] or used the tight-binding approximation [26]. Similarly, conventional Wannier interpolation schemes for the spin Hall conductivity [18,19] resorts to the band truncation. To our knowledge, the effect of band truncation on the shift spin or spin Hall conductivity has not been investigated.

In this work, we develop Wannier function perturbation theory (WFPT), which enables a localized representation of wave function perturbation, alleviating the conceptual and practical limitations of the WF method. We define the Wannier function perturbation (WFP) as the change of the WF due to a perturbation. The WFPs are spatially localized and thus can be used to efficiently interpolate the perturbation of eigenstate wave functions and the related matrix elements. We provide three applications of WFPT. The first

application is the temperature-dependent indirect optical absorption of silicon. We calculate the phonon renormalization of the electron band structure using WFPT. Using the renormalized band structures, we perform a predictive calculation of the temperature dependence of the phonon-assisted optical absorption spectra near the absorption edge. Our calculation takes into account nonadiabatic effects and does not require arbitrary shifts in energy depending on the temperature. The second and third applications are the shift spin and spin Hall conductivity of monolayer WTe_2 , respectively. In these applications, we show that WFPT is necessary to use an ultradense k -point grid and eliminate the band-truncation error at the same time.

II. FORMALISM OF WANNIER FUNCTION PERTURBATION THEORY

Before discussing the formalism of WFPT in detail, let us briefly summarize the key ideas. First, we Wannierize both the unperturbed and perturbed systems. We call the corresponding WFs the unperturbed WFs and the perturbed WFs, respectively. To study the perturbative properties of the WFs, the perturbed WFs must converge smoothly to the unperturbed WFs in the limit of a small perturbation. To achieve this property, we use the unperturbed WFs as the initial guesses and construct projection-only WFs [6] of the perturbed system. The resulting first-order derivatives of the WFs with respect to the perturbation strength are the WFPs. Next, we show that the perturbation of the eigenstate wave function can be efficiently interpolated using WFPs. Finally, we study the interpolation of matrix elements that involve wave function perturbations.

A. Wannier function perturbation

In this subsection, we first introduce the setting for our theory. Then, we derive the expression for the WFs of the perturbed system and the WFPs. Finally, we argue that the WFPs are spatially localized.

1. Setting

The WFs of a perfect crystal can be written as

$$|w_{i\mathbf{R}}^{(0)}\rangle = \frac{1}{\sqrt{N_k}} \sum_{n,\mathbf{k}} |\psi_{n\mathbf{k}}^{(0)}\rangle U_{ni;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (1)$$

Here, $|\psi_{n\mathbf{k}}^{(0)}\rangle$ is the unperturbed electron eigenstate for band n and crystal momentum $\mathbf{k}$, $i = 1, \dots, N_w$ the WF index, $\mathbf{R}$ the unit cell position, and N_k the size of the Born-von Karman supercell. The semi-unitary gauge transformation matrix $U_{ni;\mathbf{k}}^{(0)}$ is chosen to make $|w_{i\mathbf{R}}^{(0)}\rangle$ spatially localized. The unperturbed eigenstates satisfy

$$\langle \psi_{m\mathbf{k}}^{(0)} | \psi_{n\mathbf{k}'}^{(0)} \rangle = \delta_{mn} \delta_{\mathbf{k}\mathbf{k}'}. \quad (2)$$

We apply a perturbation $\lambda\hat{V}$, where λ is a small perturbation parameter. The perturbation $\hat{V}$ may or may not break the periodicity of the lattice but should respect the Born-von Karman periodic boundary condition. The only additional condition we impose on the perturbation $\hat{V}$ is that $\langle \mathbf{r} | \hat{V} | \mathbf{r}' \rangle$ should be small if $|\mathbf{r} - \mathbf{r}'|$ is large. This condition is necessary for the perturbed system to be nearsighted. Any local potentials, phonon perturbation potentials that contain contributions from nonlocal pseudopotentials, velocity operator $\hat{v}^a = [\hat{r}^a, \hat{H}^{(0)}]/i\hbar$, where $\hat{H}^{(0)}$ is the unperturbed Hamiltonian, and spin-velocity operator $\hat{j}^{s;a} = \frac{1}{2} \{\hat{S}_s, \hat{v}^a\}$ all fulfill this condition. Here, $\hat{S}_s$ is the spin operator for spin polarization along s .

Eigenstates of the perturbed system can be expanded as

$$|\psi_{n\mathbf{k}}\rangle = |\psi_{n\mathbf{k}}^{(0)}\rangle + \lambda|\psi_{n\mathbf{k}}^{(1)}\rangle + \mathcal{O}(\lambda^2), \quad (3)$$

where the wave function perturbation is

$$|\psi_{n\mathbf{k}}^{(1)}\rangle = \sum_{n'\mathbf{k}'}' |\psi_{n'\mathbf{k}'}^{(0)}\rangle \frac{\langle \psi_{n'\mathbf{k}'}^{(0)} | \hat{V} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{n'\mathbf{k}'}^{(0)}}. \quad (4)$$

Hereafter, the primed sum denotes that the terms making the denominator zero are excluded. For degenerate states, the true perturbed eigenstates in the $\lambda \rightarrow 0$ limit will, in general, converge to the unperturbed eigenstates that diagonalize $\hat{V}$ in the degenerate subspace, not to $|\psi_{n\mathbf{k}}^{(0)}\rangle$. In this case, we linearly combine the perturbed eigenstates so that they are no longer eigenstates of the perturbed Hamiltonian but satisfy Eqs. (3) and (4). This operation is allowed because our goal is to construct WFs of the perturbed system, and there is gauge freedom in constructing the WFs that allows such a linear combination. We note that this treatment is needed only to make the derivation more concise, while, in practice, one does not have any

issue with degeneracy because of the careful use of energy windows that will be described in the following.

2. WFs of the perturbed system and the WFPs

We aim to accurately describe the wave function perturbation of the low-energy states whose eigenvalues are inside the “perturbation window” $\mathcal{W}_P$. We denote the inner (frozen) and outer (disentanglement) windows for the unperturbed system by $\mathcal{W}_F$ and $\mathcal{W}_D$, respectively [6]. The following inclusion relation should hold:

$$\mathcal{W}_P \subseteq \mathcal{W}_F \subseteq \mathcal{W}_D. \quad (5)$$

This relation is illustrated in Fig. 1. In principle, $\mathcal{W}_P$ can be chosen to be identical to $\mathcal{W}_F$. Using $\mathcal{W}_P$ slightly narrower than $\mathcal{W}_F$ is useful when the perturbation mixes the states right above and below the boundary of the inner window for the unperturbed system. For later use, we define the projection operators onto the subspaces spanned by the unperturbed eigenstates inside $\mathcal{W}_P$ and $\mathcal{W}_D$ as $\hat{P}_P^{(0)}$ and $\hat{P}_D^{(0)}$, respectively, and we also define $\hat{Q}_D^{(0)} = \hat{1} - \hat{P}_D^{(0)}$. Additionally, N_X and $N_{X,\mathbf{k}}$ denote the number of unperturbed states inside $\mathcal{W}_X$ and those with crystal momentum $\mathbf{k}$, respectively, for $X = P, F, D$. We only consider small λ such that no eigenvalues cross the boundary of the windows as λ increases from 0. Another important subspace is the unperturbed Wannier subspace, of which the projection operator is

$$\hat{P}_W^{(0)} = \sum_{\mathbf{R}} \sum_{i=1}^{N_W} |w_{i\mathbf{R}}^{(0)}\rangle \langle w_{i\mathbf{R}}^{(0)}|. \quad (6)$$

By construction, $\hat{P}_P^{(0)} \subseteq \hat{P}_W^{(0)} \subseteq \hat{P}_D^{(0)}$ holds, as illustrated in Fig. 1. We define “subspaces ①, ②, ③, and ④” as shown in Fig. 1.

In general, the perturbed WFs need not make any reference to the unperturbed WFs. However, to study the

FIG. 1. Schematic illustration of the subspaces and energy windows. The vertical lines indicate the separation of the full Hilbert space into subspaces. Solid vertical lines separating different regions indicate the separation by energy eigenvalues. The dashed vertical line separates the subspace spanned by the unperturbed WFs from its complement and does not indicate an energetic separation. The “high-energy” states may also include deep semicore states which are excluded in the Wannierization process.

perturbative property of the WFs, the WFs must change smoothly around $\lambda = 0$. Our key idea to make the WFs smooth functions of λ is to use the unperturbed WFs as the initial guesses for the projection-only WFs of the perturbed system. For Wannierization of the perturbed system, the inner and outer windows are set to $\mathcal{W}_P$ and $\mathcal{W}_D$, respectively. Using the perturbation window as the inner window is natural because only the states inside the perturbation window need to be accurately represented. The outer window is set to make the conditions for Wannierization of the perturbed system as similar as possible to that of the unperturbed system. This condition guarantees that the perturbed WFs are close to the unperturbed WFs; hence, it makes the WFPs well localized. See Appendix B for the numerical verification that this choice is indeed beneficial.

Now, we explicitly construct the perturbed WFs. First, we select the subspace to be Wannierized following the initialization recipe described in Sec. III. G of Ref. [6]. We take the perturbed eigenstates inside the inner window $\mathcal{W}_P$ directly. For the states not included in $\mathcal{W}_P$, we project $\hat{P}_W^{(0)}$, the projection operator for the unperturbed Wannier subspace, onto the subspace $\mathcal{W}_D - \mathcal{W}_P$ and diagonalize it. Since we are Wannierizing a supercell, we need $N_k N_W$ WFs. Hence, we select $N_k N_W - N_P$ states inside $\mathcal{W}_D - \mathcal{W}_P$ with the largest eigenvalues [6] together with N_P states inside $\mathcal{W}_P$ and use them as the basis states for the perturbed WFs.

After we obtain the disentangled subspace, we compute the perturbed WFs. Following the scheme of Ref. [5], we project the initial guess functions, the unperturbed WFs, onto the disentangled subspace and perform Löwdin orthonormalization. The resulting WFs of the perturbed system are (see Appendix A for the detailed derivation)

$$|w_{i\mathbf{R}}(\lambda)\rangle = |w_{i\mathbf{R}}^{(0)}\rangle + \lambda |w_{i\mathbf{R}}^{(1)}\rangle + \mathcal{O}(\lambda^2), \quad (7)$$

where

$$|w_{i\mathbf{R}}^{(1)}\rangle = \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi,\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} + \frac{1}{\sqrt{N_k}} \sum_{\substack{m,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_P}} \left(\hat{P}_D^{(0)} - \hat{P}_W^{(0)} \right) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi,\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (8)$$

We call $|w_{i\mathbf{R}}^{(1)}\rangle$, the first-order derivative of a WF at $\lambda = 0$, the “Wannier function perturbation.” The concept of WFP and its explicit form [Eq. (8)] is the first main theoretical result of this paper.

The WFPs can be computed without explicitly referring to the states outside the outer window. The first term of Eq. (8) can be evaluated by solving the Sternheimer equation to obtain the wave function perturbation for states $m\mathbf{k}$ inside $\mathcal{W}_D$. The second term of Eq. (8) can be rewritten using

$$\begin{aligned} & \sum_{\substack{m,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_P}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle \\ &= \sum_{\substack{m,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_P}} \sum_{m'} \sum_{n',\mathbf{k}'} |\psi_{m'\mathbf{k}'}^{(0)}\rangle \times (P_D^{(0)} - P_W^{(0)})_{m'n';\mathbf{k}'} \frac{\langle \psi_{n'\mathbf{k}'}^{(0)} | \hat{V} | \psi_{m\mathbf{k}}^{(0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{n'\mathbf{k}'}^{(0)}}. \end{aligned} \quad (9)$$

Evaluating this expression requires only the eigenenergies and matrix elements of $\hat{V}$ and $\hat{P}_W^{(0)}$ for states inside $\mathcal{W}_D$.

Figure 2(a) schematically illustrates the definition of the WFP as the change of the WF due to a perturbation. Figures 2(b) and 2(c) show the actual WF of silicon and the corresponding WFP for an optical phonon perturbation, respectively. They visually demonstrate that the WFP is the

FIG. 2. (a) Schematic illustration of the definition of a WFP. (b) Isosurface of a WF of silicon. Isovalues are $+0.21$ and -0.21 times the maximum absolute value for the yellow and cyan isosurfaces, respectively. (c) Isosurface of a WFP for an optical phonon perturbation whose vibration mode is indicated by the arrows. Isovalues are $+0.13$ and -0.13 times the maximum absolute value for the yellow and cyan isosurfaces, respectively.

change of the WF due to the perturbation and that the WFP is spatially localized.

The formula for the WFPs [Eq. (8)] can be intuitively understood from the partition of subspaces shown in Fig. 1. First, consider subspaces ① and ②, whose union is the unperturbed Wannier subspace. The wave function perturbation in this subspace is not included in the WFPs because it can be reconstructed from the perturbation theory of the WF-based tight-binding Hamiltonian. Second, consider subspace ③, which includes the states inside the outer window but not selected for Wannierization. To accurately describe the wave function perturbation, these states should be included in the WFPs. However, the contribution of these states to the perturbation of eigenstates in subspace ② might diverge because the separation of subspaces ② and ③ is done by Wannierization not by the energy eigenvalues. In Eq. (8), this problem is absent as the contribution of subspace ③ is included only if $m\mathbf{k} \in \mathcal{W}_p$. From Eq. (9), one can find that this condition ensures that the denominator $\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{n'\mathbf{k}'}^{(0)}$ is not zero because states $m\mathbf{k}$ and $n'\mathbf{k}'$ are inside subspaces ① and ③, respectively. The last subspace, subspace ④, is energetically separated from other states by the outer window $\mathcal{W}_p$. Its contribution to the wave function perturbation is included for all states as shown in the first term of Eq. (8). The separation of energy ranges by the windows is the key idea in WFPT which makes the WFPs spatially localized.

3. Spatial localization of the WFPs

The most important property of the WFPs is their spatial localization. We provide two arguments for the localization of the WFPs. First, the localization of WFPs follows from the localization of the perturbed WFs. There exists much empirical evidence that the WFs are well localized in real space if a proper initial guess is used [7,15,29]. The initial guesses we used for the WFs of the perturbed system are the MLWFs of the unperturbed system, which is an ideal guess in the $\lambda \rightarrow 0$ limit. Thus, it is very likely that the perturbed WFs are also localized, at least for small λ . Since the WFPs are the derivative of the WFs at $\lambda = 0$, the localization of the WFs for a finite range of λ including $\lambda = 0$ implies the localization of the WFPs.

The second argument is that the summands of Eq. (8) are almost smooth functions of $\mathbf{k}$. The summand being smooth in reciprocal space implies localization in real space. We emphasize that the condition $m\mathbf{k} \in \mathcal{W}_p$ in the second term of Eq. (8) plays a crucial role. If this condition is absent, the denominator $\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{n'\mathbf{k}'}^{(0)}$ in Eq. (9) can be zero, leading to a divergence in the summand and rendering the WFPs delocalized. In our construction, the denominator never becomes zero because $\epsilon_{m\mathbf{k}}^{(0)}$ is forced to be inside $\mathcal{W}_p$ (subspace ①) while $\epsilon_{n'\mathbf{k}'}^{(0)}$ is forced to be in subspace ③.

Although these two arguments are not rigorous in a mathematical sense, they provide a solid justification that

the WFPs are spatially localized. We note that the localization of disentangled MLWFs is also an empirical fact without a mathematical proof [7,30]. Later, we numerically demonstrate that the WFPs and the matrix elements involving WFPs are localized.

One source of singularity in Eq. (8) is the discontinuity that occurs when $\epsilon_{m\mathbf{k}}^{(0)}$ crosses $\mathcal{W}_p$. In this case, the summand may change discontinuously because of the condition $m\mathbf{k} \in \mathcal{W}_p$, although it does not diverge. This discontinuity can be removed by smoothing the inner window, as described in Appendix C. However, for the systems studied in this work, we find that the WFPs are already sufficiently localized without this additional step. Hence, we use the WFPs as defined in Eq. (8).

We also note that if two bands lie close to the boundary of the outer window from inside and outside, respectively, at some k point, the first term of Eq. (8) can give a large contribution rapidly changing with $\mathbf{k}$. In such a rare case, one can locally, i.e., at that particular k point, extend the outer window to include the band outside the original outer window to make the summand change slowly. The local extension of the outer window can be done just for the construction of WFPs without affecting the unperturbed WFs. In the applications studied in this paper, we did not perform this modification as the WFPs were already sufficiently localized.

B. Interpolation of wave function perturbation

Next, we study how to interpolate the wave function perturbation of Bloch eigenstates using the WFPs. In this subsection, we first construct the perturbed tight-binding model using the matrix elements of the perturbed Hamiltonian in the perturbed WF basis. Then, we use first-order perturbation theory to obtain the eigenvector of this tight-binding model. Finally, we calculate the wave function perturbation.

1. Perturbed tight-binding Hamiltonian

Let us calculate the Hamiltonian matrix elements of the perturbed system in the perturbed WF basis. The Hamiltonian matrix element is

$$\begin{aligned} H_{i'\mathbf{R}',i\mathbf{R}} &= \langle w_{i'\mathbf{R}'} | (\hat{H}^{(0)} + \lambda \hat{V}) | w_{i\mathbf{R}} \rangle \\ &= H_{i'\mathbf{R}',i\mathbf{R}}^{(0)} + \lambda H_{i'\mathbf{R}',i\mathbf{R}}^{(1)} + \mathcal{O}(\lambda^2), \end{aligned} \quad (10)$$

where

$$H_{i'\mathbf{R}',i\mathbf{R}}^{(0)} = \langle w_{i'\mathbf{R}'}^{(0)} | \hat{H}^{(0)} | w_{i\mathbf{R}}^{(0)} \rangle \quad (11)$$

is the unperturbed Hamiltonian matrix element and

$$H_{i\mathbf{R}',i\mathbf{R}}^{(1)} = \langle w_{i\mathbf{R}'}^{(0)} | \hat{V} | w_{i\mathbf{R}}^{(0)} \rangle + \langle w_{i\mathbf{R}'}^{(1)} | \hat{H}^{(0)} | w_{i\mathbf{R}}^{(0)} \rangle + \langle w_{i\mathbf{R}'}^{(0)} | \hat{H}^{(0)} | w_{i\mathbf{R}}^{(1)} \rangle \quad (12)$$

is the perturbative correction. Note that $H_{i\mathbf{R}',i\mathbf{R}}^{(1)}$ contains a direct contribution from $\hat{V}$ as well as indirect contributions from the change of the WFs.

We write the perturbation $\hat{V}$ as a sum of monochromatic perturbations

$$\hat{V} = \sum_{\mathbf{q}} \hat{V}_{\mathbf{q}}, \quad (13)$$

where each monochromatic perturbation satisfies

$$\langle \mathbf{r} + \mathbf{R} | \hat{V}_{\mathbf{q}} | \mathbf{r}' + \mathbf{R} \rangle = e^{i\mathbf{q} \cdot \mathbf{R}} \langle \mathbf{r} | \hat{V}_{\mathbf{q}} | \mathbf{r}' \rangle \quad (14)$$

for every lattice vector $\mathbf{R}$. Correspondingly, the WFPs can be written as

$$|w_{i\mathbf{R}}^{(1)}\rangle = \sum_{\mathbf{q}} |w_{i\mathbf{R};\mathbf{q}}^{(1)}\rangle, \quad (15)$$

where

$$\langle \mathbf{r} + \mathbf{R}' | w_{i\mathbf{R}+\mathbf{R}';\mathbf{q}}^{(1)} \rangle = e^{i\mathbf{q} \cdot \mathbf{R}'} \langle \mathbf{r} | w_{i\mathbf{R};\mathbf{q}}^{(1)} \rangle. \quad (16)$$

We also define real-space matrix elements

$$g_{ij\mathbf{R};\mathbf{q}} \equiv \langle w_{i\mathbf{0}}^{(0)} | \hat{V}_{\mathbf{q}} | w_{j\mathbf{R}}^{(0)} \rangle, \quad (17)$$

$$\delta g_{ij\mathbf{R};\mathbf{q}} \equiv \langle w_{i\mathbf{0}}^{(0)} | \hat{H}^{(0)} | w_{j\mathbf{R};\mathbf{q}}^{(1)} \rangle + \langle w_{i\mathbf{0};-\mathbf{q}}^{(1)} | \hat{H}^{(0)} | w_{j\mathbf{R}}^{(0)} \rangle, \quad (18)$$

and

$$\tilde{g}_{ij\mathbf{R};\mathbf{q}} \equiv g_{ij\mathbf{R};\mathbf{q}} + \delta g_{ij\mathbf{R};\mathbf{q}}. \quad (19)$$

Here, $g_{\mathbf{R};\mathbf{q}}$ is the usual WF-based matrix element of $\hat{V}_{\mathbf{q}}$, and $\delta g_{\mathbf{R};\mathbf{q}}$ is a correction originating from the change of the WFs. This correction term is not needed for calculating the scattering matrix elements, but it is necessary when calculating the perturbed wave functions. We define the Fourier transformation of $g_{\mathbf{R};\mathbf{q}}$ as

$$g_{ij\mathbf{k};\mathbf{q}}^{(W)} = \sum_{\mathbf{R}} e^{i\mathbf{k} \cdot \mathbf{R}} g_{ij\mathbf{R};\mathbf{q}}, \quad (20)$$

and analogously, we define $\tilde{g}_{\mathbf{k};\mathbf{q}}^{(W)}$.

2. Perturbed tight-binding eigenvector

The Hamiltonian of the perturbed system in the Bloch representation is

$$H_{i'\mathbf{k}',i\mathbf{k}}^{(W)} \equiv \frac{1}{N_k} \sum_{\mathbf{R},\mathbf{R}'} H_{i'\mathbf{R}',i\mathbf{R}} e^{-i\mathbf{k}' \cdot \mathbf{R}'} e^{i\mathbf{k} \cdot \mathbf{R}}. \quad (21)$$

Since the unperturbed system has lattice translation symmetry, the zeroth-order part of $H^{(W)}$ is block diagonal:

$$H_{i'\mathbf{k}',i\mathbf{k}}^{(W0)} = \delta_{\mathbf{k},\mathbf{k}'} H_{i';i;\mathbf{k}}^{(W0)} = \delta_{\mathbf{k},\mathbf{k}'} \sum_{\mathbf{R}} H_{i'\mathbf{0},i\mathbf{R}}^{(0)} e^{i\mathbf{k} \cdot \mathbf{R}}. \quad (22)$$

The unperturbed Hamiltonian $H_{\mathbf{k}}^{(W0)}$ can be diagonalized as

$$(V_{\mathbf{k}}^{(0)\dagger} H_{\mathbf{k}}^{(W0)} V_{\mathbf{k}}^{(0)})_{mn} = \varepsilon_{m\mathbf{k}}^{(H0)} \delta_{m,n}, \quad (23)$$

where $V_{\mathbf{k}}^{(0)}$ is an $N_W \times N_W$ unitary matrix. In the Born-von Karman supercell, the unperturbed eigenvector is

$$V_{i\mathbf{R},n\mathbf{k}}^{(0)} = \frac{1}{\sqrt{N_k}} e^{i\mathbf{k} \cdot \mathbf{R}} V_{in;\mathbf{k}}^{(0)}. \quad (24)$$

Next, we consider the perturbative correction. Using Eqs. (14) and (16), one can show that the Fourier transformation of $H_{i'\mathbf{R}',i\mathbf{R}}^{(1)}$ is

$$H_{i'\mathbf{k}',i\mathbf{k}}^{(W1)} = \tilde{g}_{i'\mathbf{k};\mathbf{k}'-\mathbf{k}}^{(W)} \quad (25)$$

(see Appendix D for a proof). The matrix element of $H^{(W1)}$ in the unperturbed eigenstate basis is

$$H_{n'\mathbf{k}',n\mathbf{k}}^{(H1)} = (V_{\mathbf{k}'}^{(0)\dagger} \tilde{g}_{\mathbf{k};\mathbf{k}'-\mathbf{k}}^{(W)} V_{\mathbf{k}}^{(0)})_{n'n} \equiv \tilde{g}_{n'\mathbf{k};\mathbf{k}'-\mathbf{k}}^{(H)}. \quad (26)$$

We also define $g^{(H)}$ analogously.

We now apply first-order perturbation theory to the tight-binding Hamiltonian. The perturbed eigenvector in the WF basis is

$$V_{i\mathbf{R},n\mathbf{k}} = \frac{1}{\sqrt{N_k}} e^{i\mathbf{k} \cdot \mathbf{R}} V_{in;\mathbf{k}}^{(0)} + \lambda \frac{1}{\sqrt{N_k}} \sum_{\mathbf{k}'} \sum_{n'=1}^{N_W} e^{i\mathbf{k}' \cdot \mathbf{R}} V_{in';\mathbf{k}'}^{(0)} \times \frac{\tilde{g}_{n'\mathbf{k};\mathbf{k}-\mathbf{k}'}^{(H)}}{\varepsilon_{n\mathbf{k}}^{(H0)} - \varepsilon_{n'\mathbf{k}'}^{(H0)}} + \mathcal{O}(\lambda^2). \quad (27)$$

3. Calculation of wave function perturbation

We can now calculate the wave function perturbation as we know the basis functions (the perturbed WF $|w_{i\mathbf{R}}\rangle$) as well as the coefficients (the perturbed eigenvector $V_{i\mathbf{R},n\mathbf{k}}$) for the perturbed wave function. The perturbed wave function reads

$$\begin{aligned} |\psi_{n\mathbf{k}}^{(H)}\rangle &= \sum_{i,\mathbf{R}} |w_{i\mathbf{R}}\rangle V_{i\mathbf{R},n\mathbf{k}} \\ &= |\psi_{n\mathbf{k}}^{(H0)}\rangle + \lambda \sum_{\mathbf{q}} |\psi_{n\mathbf{k};\mathbf{q}}^{(H1)}\rangle + \mathcal{O}(\lambda^2), \end{aligned} \quad (28)$$

where

$$|\psi_{n\mathbf{k}}^{(H0)}\rangle = \frac{1}{\sqrt{N_k}} \sum_{i,\mathbf{R}} |w_{i\mathbf{R}}^{(0)}\rangle V_{in,\mathbf{k}}^{(0)} e^{i\mathbf{k}\cdot\mathbf{R}} \quad (29)$$

and

$$\begin{aligned} |\psi_{n\mathbf{k};\mathbf{q}}^{(H1)}\rangle &= \frac{1}{\sqrt{N_k}} \sum_{i,\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |w_{i\mathbf{R};\mathbf{q}}^{(1)}\rangle V_{in,\mathbf{k}}^{(0)} \\ &+ \sum_{m=1}^{N_w} |\psi_{m\mathbf{k}+\mathbf{q}}^{(H0)}\rangle \frac{\tilde{g}_{mn\mathbf{k};\mathbf{q}}^{(H)}}{\varepsilon_{n\mathbf{k}}^{(H0)} - \varepsilon_{m\mathbf{k}+\mathbf{q}}^{(H0)}}. \end{aligned} \quad (30)$$

Equation (30) contains contributions from both the WFP and the perturbation of the tight-binding eigenvector $V_{in;\mathbf{k}}^{(0)}$.

Equation (30) is the interpolation equation for the first-order wave function perturbation. This equation is the second main theoretical result of this paper. The contribution of the states outside $\hat{P}_W^{(0)}$, the unperturbed Wannier subspace, to the wave function perturbation is taken into account by the WFPs and is the first term of Eq. (30). The contribution of the states inside $\hat{P}_W^{(0)}$ is the second term, which is derived by the perturbation theory in the tight-

binding model. Since the WFPs are spatially localized, the wave function perturbation converges quickly with the number of lattice vectors used in the sum over $\mathbf{R}$.

C. Interpolation of matrix elements

Finally, we demonstrate how the matrix elements that involve wave function perturbations can be interpolated. We consider interpolation of two different types of quantities, $S_{mn\mathbf{k}}$ [Eq. (31)] and $K_{mn\mathbf{k}}$ [Eq. (37)].

1. Interpolation of $S_{mn\mathbf{k}}$

First, we consider quantities of the form

$$\begin{aligned} S_{mn\mathbf{k}} &= \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle \\ &= \sum_p' \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}'_{\mathbf{q}} | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle \langle \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}}, \end{aligned} \quad (31)$$

where $\hat{V}'_{\mathbf{q}}$ is a monochromatic perturbation that may or may not equal $\hat{V}_{\mathbf{q}}$. If $\hat{V}_{\mathbf{q}}$ is a phonon potential, interpolation of $S_{\mathbf{k}}$ enables an efficient calculation of the phonon-induced renormalization of the electron band structure. For $\hat{V}_0$ being the spin-velocity operator and $\hat{V}_{\mathbf{q}\neq 0} = 0$, $S_{\mathbf{k}}$ is relevant to the shift spin current.

For $m\mathbf{k}, n\mathbf{k} \in \mathcal{W}_p$, one can use the Wannier interpolation of the unperturbed wave function and the wave function perturbation [Eqs. (29) and (30)] to find

$$\begin{aligned} S_{mn\mathbf{k}} &= \langle \psi_{m\mathbf{k};\mathbf{q}}^{(H1)} | \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(H0)} \rangle \\ &= \frac{1}{N_k} \sum_{i,i',\mathbf{R},\mathbf{R}'} V_{mi',\mathbf{k}}^{(0)\dagger} \langle w_{i'\mathbf{R}';\mathbf{q}}^{(1)} | \hat{V}'_{\mathbf{q}} | w_{i\mathbf{R}}^{(0)} \rangle V_{in;\mathbf{k}}^{(0)} e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')} + \sum_{p=1}^{N_w} \frac{(\tilde{g}_{pm\mathbf{k};\mathbf{q}}^{(H)})^* \langle \psi_{p\mathbf{k}+\mathbf{q}}^{(H0)} | \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(H0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(H0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(H0)}} \\ &= \sum_{i,i',\mathbf{R}} V_{mi',\mathbf{k}}^{(0)\dagger} \langle w_{i'\mathbf{0};\mathbf{q}}^{(1)} | \hat{V}'_{\mathbf{q}} | w_{i\mathbf{R}}^{(0)} \rangle V_{in;\mathbf{k}}^{(0)} e^{i\mathbf{k}\cdot\mathbf{R}} + \sum_{p=1}^{N_w} \frac{(\tilde{g}_{pm\mathbf{k};\mathbf{q}}^{(H)})^* g'_{pn\mathbf{k};\mathbf{q}}^{(H)}}{\varepsilon_{m\mathbf{k}}^{(H0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(H0)}}. \end{aligned} \quad (32)$$

In the third equality, we use the discrete translation properties of $\hat{V}'_{\mathbf{q}}$ and the WFPs [Eq. (14) and (16)].

We define the real-space matrix element

$$s_{ij\mathbf{R};\mathbf{q}} \equiv \langle w_{i\mathbf{0};\mathbf{q}}^{(1)} | \hat{V}'_{\mathbf{q}} | w_{j\mathbf{R}}^{(0)} \rangle. \quad (33)$$

Since the WFs and the WFPs are localized, $s_{ij\mathbf{R};\mathbf{q}}$ rapidly decays as a function of $|\mathbf{R}|$. Defining

$$s_{ij\mathbf{k};\mathbf{q}}^{(W)} \equiv \frac{1}{N_k} \sum_{\mathbf{R}} s_{ij\mathbf{R};\mathbf{q}} e^{i\mathbf{k}\cdot\mathbf{R}} \quad (34)$$

and

$$s_{mn\mathbf{k};\mathbf{q}}^{(H)} \equiv (V_{\mathbf{k}}^{(0)\dagger} s_{\mathbf{k};\mathbf{q}}^{(W)} V_{\mathbf{k}}^{(0)})_{mn}, \quad (35)$$

one can rewrite Eq. (32) as

$$S_{mn\mathbf{k}} = s_{mn\mathbf{k};\mathbf{q}}^{(H)} + \sum_{p=1}^{N_w} \frac{(\tilde{g}_{pm\mathbf{k};\mathbf{q}}^{(H)})^* g'_{pn\mathbf{k};\mathbf{q}}^{(H)}}{\varepsilon_{m\mathbf{k}}^{(H0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(H0)}}. \quad (36)$$

Since all quantities appearing in Eq. (36) can be calculated using a WF- or WFP-based interpolation, $S_{mn\mathbf{k}}$ can also be interpolated.

TABLE I. Key quantities of WFPT.

Description	Real space, Wannier basis	Reciprocal space, eigenstate basis
Wave function perturbation	$ w_{i\mathbf{R}}^{(1)}\rangle$, Eq. (8)	$ \psi_{n\mathbf{k};\mathbf{q}}^{(H1)}\rangle$, Eq. (30)
Perturbation to the Hamiltonian matrix elements	WF contribution: $g_{ij\mathbf{R};\mathbf{q}}$, Eq. (17) WFP contribution: $\delta g_{ij\mathbf{R};\mathbf{q}}$, Eq. (18) Total: $\tilde{g}_{ij\mathbf{R};\mathbf{q}}$, Eq. (19)	$H_{n'\mathbf{k}',n\mathbf{k}}^{(H1)}$, Eq. (26)
Sternheimer matrix elements	$s_{ij\mathbf{R};\mathbf{q}}$, Eq. (33)	$S_{mn\mathbf{k}}$, Eqs. (31), (32)
Kubo matrix elements	$k_{ij\mathbf{R};\mathbf{q}}$, Eq. (38)	$K_{mn\mathbf{k}}$, Eqs. (37), (39)

2. Interpolation of $K_{mn\mathbf{k}}$

Next, we show that the matrix element

$$K_{mn\mathbf{k}} = \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \psi_{n\mathbf{k};\mathbf{q}}^{(1)} \rangle = \sum_p \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}_{\mathbf{q}}^\dagger | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle \langle \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}}' | \psi_{n\mathbf{k}}^{(0)} \rangle}{(\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(0)})(\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(0)})} \quad (37)$$

can similarly be interpolated using WFPT. This matrix element is relevant to the evaluation of the Kubo formula for noninteracting quasiparticles. When $\mathbf{q} = 0$ and $\hat{V}_0$ and $\hat{V}_0'$ are the velocity and spin-velocity operators, respectively, $K_{mn\mathbf{k}}$ is directly related to the spin Berry curvature and the spin Hall conductivity.

Defining the real-space matrix element

$$k_{ij\mathbf{R};\mathbf{q}} \equiv \langle w_{i0;\mathbf{q}}^{(1)} | w_{j\mathbf{R};\mathbf{q}}^{(1)} \rangle, \quad (38)$$

we show in Appendix E that $K_{mn\mathbf{k}}$ can be interpolated as

$$K_{mn\mathbf{k}} = k_{mn\mathbf{k};\mathbf{q}}^{(H)} + \sum_{p=1}^{N_W} \frac{(\tilde{g}_{pm\mathbf{k};\mathbf{q}}^{(H)})^* \tilde{g}_{pn\mathbf{k};\mathbf{q}}^{(H)}}{(\epsilon_{m\mathbf{k}}^{(H0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(H0)})(\epsilon_{n\mathbf{k}}^{(H0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(H0)})}. \quad (39)$$

Here, we define

$$k_{mn\mathbf{k};\mathbf{q}}^{(H)} \equiv (V_{\mathbf{k}}^{(0)\dagger} k_{\mathbf{k};\mathbf{q}}^{(W)} V_{\mathbf{k}}^{(0)})_{mn} \quad (40)$$

and

$$k_{ij\mathbf{k};\mathbf{q}}^{(W)} \equiv \frac{1}{N_k} \sum_{\mathbf{R}} k_{ij\mathbf{R};\mathbf{q}} e^{i\mathbf{k} \cdot \mathbf{R}}. \quad (41)$$

In the actual evaluation of response quantities, positive infinitesimal η that ensures the causality of the response is included in the denominator of the low-energy-band

contribution, as in Eqs. (49) and (61). Numerically, the value of η is decreased until the calculated quantity converges.

In Appendix F, we detail how to calculate the real-space matrix elements $g_{ij\mathbf{R};\mathbf{q}}$, $\tilde{g}_{ij\mathbf{R};\mathbf{q}}$, $s_{ij\mathbf{R};\mathbf{q}}$, and $k_{ij\mathbf{R};\mathbf{q}}$ from the Bloch eigenstates at the coarse k grid. Table I summarizes the key quantities of WFPT.

D. Proof-of-principles example: Berry curvature

Now, let us present a proof-of-principles example of WFPT interpolation by interpolating the band-resolved Berry curvature. Since the Berry curvature strongly oscillates in momentum space due to the contribution from nearby bands, we exclude the intra-valence-band contribution as follows:

$$\tilde{\Omega}_{m\mathbf{k}}^{ab} = -2\text{Im} \left\langle \frac{\partial u_{m\mathbf{k}}}{\partial k^a} \middle| \hat{Q}_{\mathbf{k},\text{valence}} \middle| \frac{\partial u_{m\mathbf{k}}}{\partial k^b} \right\rangle. \quad (42)$$

Here, $|u_{m\mathbf{k}}\rangle = e^{-i\mathbf{k} \cdot \hat{\mathbf{R}}} |\psi_{m\mathbf{k}}\rangle$ and $\hat{Q}_{\mathbf{k},\text{valence}} = \hat{1} - \hat{P}_{\mathbf{k},\text{valence}}$, with $\hat{P}_{\mathbf{k},\text{valence}} = \sum_{n \in \text{valence}} |u_{n\mathbf{k}}\rangle \langle u_{n\mathbf{k}}|$ the projection operator to the valence bands at $\mathbf{k}$. By “valence bands,” we mean only the occupied bands inside the outer window for Wannierization, $\mathcal{W}_D$: The low-lying bands that are outside $\mathcal{W}_D$ are not included in $\hat{P}_{\mathbf{k},\text{valence}}$.

The Berry curvature can be calculated using WFPT interpolation. Since the velocity operator is the derivative of the Hamiltonian with respect to $\mathbf{k}$, the derivative of the wave function is equivalent to the wave function perturbation due to a velocity operator perturbation. Hence, by choosing the velocity operators as the perturbations for the WFPs, $\hat{V}_{\mathbf{q}} = \hat{v}^a \delta_{\mathbf{q},0}$ and $\hat{V}_{\mathbf{q}}' = \hat{v}^b \delta_{\mathbf{q},0}$, we have

$$|\psi_{m\mathbf{k}}^{(1)}\rangle = e^{i\mathbf{k} \cdot \mathbf{r}} \left| \frac{\partial u_{m\mathbf{k}}}{\partial k^a} \right\rangle \quad \text{and} \quad |\psi_{m\mathbf{k}}'^{(1)}\rangle = e^{i\mathbf{k} \cdot \mathbf{r}} \left| \frac{\partial u_{m\mathbf{k}}}{\partial k^b} \right\rangle. \quad (43)$$

Using Eqs. (37) and (39), we find

$$\begin{aligned}
\tilde{\Omega}_{m\mathbf{k}}^{ab} &= -2\text{Im}\left(\left\langle\frac{\partial u_{m\mathbf{k}}}{\partial k^a}\middle|\frac{\partial u_{m\mathbf{k}}}{\partial k^b}\right\rangle - \sum_{n\in\text{valence}}\left\langle\frac{\partial u_{m\mathbf{k}}}{\partial k^a}\middle|u_{n\mathbf{k}}\right\rangle\left\langle u_{n\mathbf{k}}\middle|\frac{\partial u_{m\mathbf{k}}}{\partial k^b}\right\rangle\right) \\
&= -2\text{Im}\left(\langle\psi_{m\mathbf{k}}^{(1)}|\psi_{m\mathbf{k}}^{(1)}\rangle - \sum_{n\in\text{valence}}\langle\psi_{m\mathbf{k}}^{(1)}|\psi_{n\mathbf{k}}^{(0)}\rangle\langle\psi_{n\mathbf{k}}^{(0)}|\psi_{m\mathbf{k}}^{(1)}\rangle\right) \\
&= -2\text{Im}\left(K_{mm\mathbf{k}} - \sum_{\substack{n=1 \\ n\in\text{valence}}}^{N_W}\frac{(g_{nm\mathbf{k}}^{(H)})^*g_{nm\mathbf{k}}^{(H)}}{(\varepsilon_{m\mathbf{k}}^{(H0)} - \varepsilon_{n\mathbf{k}}^{(H0)})^2}\right) \\
&= -2\text{Im}\left(k_{mm\mathbf{k}}^{(H)} + \sum_{\substack{n=1 \\ n\notin\text{valence}}}^{N_W}\frac{(\tilde{g}_{nm\mathbf{k}}^{(H)})^*\tilde{g}_{nm\mathbf{k}}^{(H)}}{(\varepsilon_{m\mathbf{k}}^{(H0)} - \varepsilon_{n\mathbf{k}}^{(H0)})^2}\right). \tag{44}
\end{aligned}$$

This equation is the WFPT interpolation formula for $\tilde{\Omega}_{m\mathbf{k}}^{ab}$. In the last equality, we used the fact that in the limit of exact interpolation, $g_{nm\mathbf{k}}^{(H)} = \tilde{g}_{nm\mathbf{k}}^{(H)}$ holds for $m\mathbf{k}, n\mathbf{k} \in \mathcal{W}_P$, which can easily be shown to follow from the $P_D^{(0)} - P_W^{(0)}$ terms in Eq. (F2). Because of the interpolation error, $\delta g_{nm\mathbf{k}}^{(H)} = 0$ holds exactly only if $\mathbf{k}$ belongs to the coarse k -point grid for Wannier interpolation. But, the deviation at a generic $\mathbf{k}$ is small and can be diminished by using a denser coarse k -point grid for Wannierization, as with all other Wannier-interpolated quantities.

We also calculate the Berry curvature directly in the plane-wave basis by solving the following Sternheimer equation to calculate $\hat{Q}_{\mathbf{k},\text{valence}}|\partial u_{m\mathbf{k}}/\partial k^a\rangle$:

$$\begin{aligned}
&(\hat{H}(\mathbf{k}) - \varepsilon_{m\mathbf{k}})\hat{Q}_{\mathbf{k},\text{valence}}\left|\frac{\partial u_{m\mathbf{k}}}{\partial k^a}\right\rangle \\
&= -\left(\sum_{\substack{n=1 \\ n\notin\text{valence}}}^{\infty}|u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|\right)\frac{\partial \hat{H}(\mathbf{k})}{\partial k^a}|u_{m\mathbf{k}}\rangle, \tag{45}
\end{aligned}$$

FIG. 3. Band-resolved Berry curvature of monolayer WTe₂. VBM stands for the highest-energy valence band, or valence band maximum. The plane-wave calculation is done by solving the Sternheimer equation, Eq. (45), to directly evaluate Eq. (42).

where $\hat{H}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\hat{\mathbf{r}}}\hat{H}e^{i\mathbf{k}\cdot\hat{\mathbf{r}}}$. The conventional Wannier interpolation method is also exact for the calculation of Berry curvature [14]. There, $\tilde{\Omega}_{m\mathbf{k}}^{ab}$ is calculated by first calculating the full Berry curvature and then subtracting the intra-valence-band contribution.

Figure 3 shows the Berry curvature of four valence bands of monolayer WTe₂ calculated using WFPT interpolation, direct plane-wave calculations, and exact conventional Wannier interpolation. The WFPT result agrees perfectly with the plane-wave and Wannier interpolation results, demonstrating the validity of the WFPT interpolation method.

III. RESULTS

A. Application 1: Temperature-dependent electron band structure and indirect absorption

As the first application of WFPT, we calculate the temperature-dependent electronic structure and indirect optical absorption spectra of silicon. The electron self-energy in the dynamical Allen-Heine-Cardona formalism [31–33] is the sum of the Fan and the Debye-Waller (DW) terms [34]:

$$\Sigma_{nn'\mathbf{k}}(\omega) = \frac{1}{N_q} \sum_{\mathbf{q},\nu} [\Sigma_{nn'\mathbf{k};\mathbf{q}\nu}^{\text{Fan}}(\omega) + \Sigma_{nn'\mathbf{k};\mathbf{q}\nu}^{\text{DW}}], \tag{46}$$

$$\begin{aligned}
\Sigma_{nn'\mathbf{k};\mathbf{q}\nu}^{\text{Fan}}(\omega) &= \sum_m g_{mn\mathbf{k};\mathbf{q}\nu}^* g_{mn'\mathbf{k};\mathbf{q}\nu} \\
&\times \sum_{\pm} \frac{n_{\mathbf{q}\nu} + [1 \pm (2f_{m\mathbf{k}+\mathbf{q}} - 1)]/2}{\omega - \varepsilon_{m\mathbf{k}+\mathbf{q}} \pm \omega_{\mathbf{q}\nu} + i\eta}, \tag{47}
\end{aligned}$$

$$\Sigma_{nn'\mathbf{k};\mathbf{q}\nu}^{\text{DW}} = \frac{n_{\mathbf{q}\nu} + \frac{1}{2}}{2\omega_{\mathbf{q}\nu}} \sum_{\substack{m,\kappa,a \\ \kappa',a'}} \mathcal{D}_{nn'}^{\kappa a \kappa' a'}(\mathbf{k}, \mathbf{q}) U_{\kappa a, \nu}^*(\mathbf{q}) U_{\kappa' a', \nu}(\mathbf{q}). \tag{48}$$

Here, $U_{\kappa a, \nu}(\mathbf{q})$ is the displacement pattern for the phonon eigenmode with branch index ν and wave vector $\mathbf{q}$ in units of inverse square root of mass, $n_{\mathbf{q}\nu}$ the Bose-Einstein occupation factor, $f_{m\mathbf{k}+\mathbf{q}}$ the Fermi-Dirac occupation

factor, and $\mathcal{D}_{nn'}^{\kappa\kappa'd'}(\mathbf{k}, \mathbf{q})$ the DW matrix element [see Eq. (5) of Ref. [35] for the definition]. In practice, the DW matrix element is calculated within the rigid-ion approximation [31,36].

Under the on-the-mass-shell approximation [37,38], the diagonal part of the self-energy of the state $n\mathbf{k}$ is $\Sigma_{n\mathbf{k}} = \Sigma_{nn\mathbf{k}}(\epsilon_{n\mathbf{k}})$. In this work, we only consider the diagonal part of the self-energy because the off-diagonal part has little effect on silicon, which has a band gap much larger than the band energy renormalization [35].

In practical calculations, the Fan term is separated into the lower Fan and the upper Fan terms, which are the contributions from the low-energy and high-energy bands, respectively. We write the set of low-energy states as $\mathcal{W}_{\text{Lower}}$. The upper bound of the energy window for $\mathcal{W}_{\text{Lower}}$ is chosen to be a few eV above $\epsilon_{n\mathbf{k}}^{(0)}$. We set $\mathcal{W}_{\text{Lower}}$ to be inside the perturbation window, which is necessary because the lower Fan term [Eq. (49)] requires an explicit Wannier interpolation of the EPC matrix elements for bands inside $\mathcal{W}_{\text{Lower}}$.

In the upper Fan term, one neglects the phonon frequency $\omega_{\mathbf{q}\nu}$ and η in the denominator, which is a good approximation because $|\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}+\mathbf{q}}|$ is always much larger than $\omega_{\mathbf{q}\nu}$ and η as one is only interested in the self-energy of low-energy states whose energy is much lower than the upper energy bound for $\mathcal{W}_{\text{Lower}}$. The lower and upper Fan terms read

$$\Sigma_{n\mathbf{k};\mathbf{q}\nu}^{\text{lower Fan}}(\eta) = \sum_{m\mathbf{k}+\mathbf{q} \in \mathcal{W}_{\text{Lower}}} |g_{mn\mathbf{k};\mathbf{q}\nu}|^2 \times \sum_{\pm} \frac{n_{\mathbf{q}\nu} + [1 \pm (2f_{m\mathbf{k}+\mathbf{q}} - 1)]/2}{\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}+\mathbf{q}} \pm \omega_{\mathbf{q}\nu} + i\eta} \quad (49)$$

and

$$\Sigma_{n\mathbf{k};\mathbf{q}\nu}^{\text{upper Fan}} = (2n_{\mathbf{q}\nu} + 1) \sum_{m\mathbf{k}+\mathbf{q} \notin \mathcal{W}_{\text{Lower}}} \frac{|g_{mn\mathbf{k};\mathbf{q}\nu}|^2}{\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}+\mathbf{q}}}, \quad (50)$$

respectively.

The DW self-energy under the rigid-ion approximation [31,36] can be easily Wannier interpolated without any infinite sum by exploiting the recently developed momentum-operator-based representation of the DW matrix elements [35]. The lower Fan term [Eq. (49)] can also be calculated using the conventional Wannier interpolation method.

Calculation of the upper Fan term [Eq. (50)] is where the WF interpolation fails and WFPT becomes necessary. The reason is that the upper Fan term involves a sum over an infinite number of bands. Typically, a few hundred bands are needed to reach convergence [23,24]. Wannier

interpolation of such a large number of bands is impractical, if not impossible.

Using WFPT, the upper Fan term can be easily interpolated without explicitly performing the sum over bands. By choosing $\hat{V}_{\mathbf{q}}$ and $\hat{V}'_{\mathbf{q}}$ as the perturbation potential for the same phonon mode ν in the definition of $S_{mn\mathbf{k}}$ [Eq. (31)] and using the WFPT interpolation formula [Eq. (36)], we find

$$\begin{aligned} \sum_{m\mathbf{k}+\mathbf{q} \notin \mathcal{W}_{\text{Lower}}} \frac{|g_{mn\mathbf{k};\nu\mathbf{q}}|^2}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{m\mathbf{k}+\mathbf{q}}^{(0)}} &= S_{nn\mathbf{k};\nu\mathbf{q}} - \sum_{m=1}^{N_W} \frac{(g_{mn\mathbf{k};\nu\mathbf{q}}^{(H)})^* g_{mn\mathbf{k};\nu\mathbf{q}}^{(H)}}{\epsilon_{n\mathbf{k}}^{(H0)} - \epsilon_{m\mathbf{k}+\mathbf{q}}^{(H0)}} \\ &= S_{nn\mathbf{k};\nu\mathbf{q}}^{(H)} + \sum_{m=1}^{N_W} \frac{(\tilde{g}_{mn\mathbf{k};\nu\mathbf{q}}^{(H)})^* g_{mn\mathbf{k};\nu\mathbf{q}}^{(H)}}{\epsilon_{n\mathbf{k}}^{(H0)} - \epsilon_{m\mathbf{k}+\mathbf{q}}^{(H0)}}. \end{aligned} \quad (51)$$

Equation (51) is the WFPT interpolation formula for the upper Fan self-energy. In the second equality, we use the fact that in the limit of exact interpolation, $\delta g_{mn\mathbf{k}}^{(H)} = 0$ and $g_{mn\mathbf{k}}^{(H)} = \tilde{g}_{mn\mathbf{k}}^{(H)}$ if $m\mathbf{k}, n\mathbf{k} \in \mathcal{W}_p$, as explained in Sec. IID.

We calculate the upper Fan and DW terms on a coarse $6 \times 6 \times 6$ q -point grid, as it converges much faster than the lower Fan term. We develop and use an improved double-grid method to speed up the convergence with respect to the q grid used to calculate the upper Fan and the DW self-energies. The improved double-grid method utilizes the fact that the low-energy acoustic phonons should give a small contribution to the total self-energy due to the translational invariance. With this improved double-grid method, one can achieve convergence at a considerably coarser q grid than the conventional double-grid method [35,39] (see Appendix G for details).

Figures 4(a) and 4(b) show the renormalized band structure of silicon at $T = 1000$ K computed using WFPT. For comparison, we also show the renormalized band energies calculated directly in the plane-wave basis. WFPT gives an accurate result at a much smaller computational cost.

Since WFPT enables a cheap calculation of the eigenvalue renormalization, we can now calculate quantities that require renormalized eigenvalues at a large number of k points. As a prototypical example, we study the temperature dependence of the indirect absorption spectrum. Without the phonon-induced band-structure renormalization, the perturbative expression for the phonon-assisted absorption coefficient is [11,41]

$$\begin{aligned} \alpha(\Omega) &= \frac{\pi q^2}{c\epsilon_0 V \Omega n(\Omega)} \sum_{\nu, m, n, \mathbf{k}, \mathbf{q}, \pm} \left(n_{\mathbf{q}\nu} + \frac{1}{2} \pm \frac{1}{2} \right) \\ &\times (f_{n\mathbf{k}} - f_{m\mathbf{k}+\mathbf{q}}) |\boldsymbol{\lambda} \cdot (\mathbf{S}_1 + \mathbf{S}_2)|^2 \\ &\times \delta(\epsilon_{m\mathbf{k}+\mathbf{q}} - \epsilon_{n\mathbf{k}} - \hbar\Omega \pm \hbar\omega_{\mathbf{q}\nu}), \end{aligned} \quad (52)$$

FIG. 4. (a,b) Band structure of silicon around the (a) VBM and (b) conduction band minimum (CBM) computed without and with the EPC at $T = 1000$ K. The red shades around the red curves indicate the phonon-induced broadening, with the width the same as the imaginary part of the self-energy. (c,d) Indirect absorption spectra of silicon computed (c) without and (d) with the EPC-induced band renormalization, and the experimental data from Ref. [40]. The curves correspond to temperatures 20, 77, 90, 112, 170, 195, 249, 291, 363, and 415 K, from right to left.

where

$$S_1 = \sum_{p=1}^{\infty} \frac{\mathbf{v}_{np\mathbf{k}} g_{p\mathbf{k};-\mathbf{q}\nu}}{\varepsilon_{p\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}} \mp \hbar\omega_{\mathbf{q}\nu} + i\Gamma_{p\mathbf{k}}} \quad (53)$$

and

$$S_2 = \sum_{p=1}^{\infty} \frac{g_{np\mathbf{k};-\mathbf{q}\nu} \mathbf{v}_{p\mathbf{k}+\mathbf{q}}}{\varepsilon_{p\mathbf{k}+\mathbf{q}} - \varepsilon_{n\mathbf{k}} \pm \hbar\omega_{\mathbf{q}\nu} + i\Gamma_{p\mathbf{k}+\mathbf{q}}}. \quad (54)$$

Here, ϵ_0 is the vacuum permittivity, Ω the frequency of the incident photon, c the speed of light, λ the light polarization, and $n(\Omega)$ the frequency-dependent refractive index of the material. Note that we use the energy conservation imposed by the delta function to rewrite the denominator of Eq. (53) in a form different from Eq. (2) of Ref. [41].

When the phonon-induced band-structure renormalization is included, the absorption coefficient becomes [42]

$$\begin{aligned} \alpha(\Omega) = & 2 \frac{4\pi^2 q^2}{\Omega c n(\Omega) V} \sum_{\nu, m, n, \mathbf{k}, \mathbf{q}, \pm} \left(n_{\mathbf{q}\nu} + \frac{1}{2} \pm \frac{1}{2} \right) \\ & \times (f_{m\mathbf{k}} - f_{n\mathbf{k}+\mathbf{q}}) |\lambda \cdot (\mathbf{S}_1 + \mathbf{S}_2)|^2 \\ & \times \delta(\varepsilon_{m\mathbf{k}+\mathbf{q}} + \text{Re}\Sigma_{m\mathbf{k}+\mathbf{q}} - \varepsilon_{n\mathbf{k}} - \text{Re}\Sigma_{n\mathbf{k}} - \hbar\Omega \pm \hbar\omega_{\mathbf{q}\nu}). \end{aligned} \quad (55)$$

The phonon-induced real part of the self-energy appears only in the delta function while $\mathbf{S}_1$ and $\mathbf{S}_2$ are unaffected [34,42].

Often, the absorption spectra are insensitive to the imaginary part of the self-energy of the intermediate state

[see Fig. S2(d)]. Following Ref. [11], we approximate $\Gamma_{p\mathbf{k}}$ as a constant:

$$\Gamma_{p\mathbf{k}} = \Gamma_{p\mathbf{k}+\mathbf{q}} = \Gamma. \quad (56)$$

Also, we broaden the delta function by a Gaussian:

$$\delta(\varepsilon) \approx \frac{1}{\gamma\sqrt{\pi}} e^{-\varepsilon^2/\gamma^2}, \quad (57)$$

where γ is the broadening parameter.

We correct the band structure and the velocity matrix elements using the G_0W_0 method while calculating the EPC and the phonon-induced renormalization at the DFT level. The G_0W_0 -corrected band gap is 1.28 eV, which is consistent with a previous calculation [43].

Figures 4(c) and 4(d) show the indirect absorption spectra near the onset calculated without [Fig. 4(c)] and with [Fig. 4(d)] the phonon-induced band renormalization. The spectra calculated using the WFPT-interpolated band renormalization successfully reproduce the experimental spectra for the entire range of temperatures, from 20 K to 415 K. Previous perturbative calculation of the indirect absorption spectra used an arbitrary temperature-dependent shift of the calculated spectra [11] because the phonon contribution to the electron self-energy was not considered. Our calculation is free from such temperature-dependent parameters and thus is an accurate and predictive calculation of the phonon-assisted indirect absorption.

An alternative method to study EPC is the adiabatic supercell method, where supercells with distorted lattice structures are used [44–48]. The adiabatic supercell

methods are limited in that the phonon emission and absorption processes are not separated. Consequently, such methods cannot reproduce the characteristic two-slope behavior near the onset of indirect absorption due to the different onsets of phonon emission and absorption [11]. Also, the absorption onset is hard to resolve due to the large computational cost of supercell calculations and the need for a fine q -point sampling. For infrared-active materials, adiabatic supercell methods also miss the non-adiabatic effect of the Fröhlich interaction, which is the predominant mechanism for band renormalization in many materials [38].

We note that we used two empirical parameters regarding quasiparticle energies to compute the spectra in Figs. 4(c) and 4(d). First, we shifted all theoretical spectra by -54 meV to match the experimental ones, which reflects the typical accuracy of G_0W_0 calculations [49]. The spectra computed without EPC were additionally shifted by -93 meV to match the experimental ones at low temperatures. Second, we scaled the EPC-induced self-energy by a factor of 1.36, correcting the underestimation due to the neglect of electron-electron interactions in calculating EPC [50]. The scaling factor was obtained by dividing the experimental slope of the temperature-dependent indirect band-gap renormalization at the high-temperature region [40,51] by the calculated one. We emphasize that these parameters are temperature independent. The need for these parameters is the limitation of semilocal DFT and the G_0W_0 method. More sophisticated methods to treat electron-electron interactions may allow a completely parameter-free calculation but this is beyond the scope of this study.

The real-space matrix elements calculated using WFs and WFPs are well localized, as shown in Fig. 5. This fact explains the success of the WF- and WFP-based interpolation.

We summarized the computational cost for calculating the phonon-induced electron self-energy in Table II. We find a speedup of several orders of magnitude. The major speedup of the interpolation scheme comes from the calculation of the lower Fan term, where both the k and q points are finely sampled using electron and phonon WFs. Although this speedup comes from the conventional WF interpolation, WFPT interpolation enables one to consistently calculate the self-energy in the WF basis. If WFPT is not used, the upper and lower Fan parts of the self-energy should have been calculated in the plane-wave and the WF bases, respectively. Assigning band indices to those two calculations when the bands are *nearly* degenerate is a nontrivial issue. The WFPT interpolation of the upper Fan self-energy for the fine k points (23 cpu · h) also gives a considerable speedup of above 50 times compared to solving the Sternheimer equation in the plane-wave basis (1241 cpu · h). In our calculation, the number of fine k points is quite small (379), as we sampled only the states

FIG. 5. Spatial decay of (a) Hamiltonian, (b) EPC, and (c) WFP matrix elements. Each data point corresponds to the maximum absolute value over all WF pairs, phonon momentum, and atomic displacements for each lattice vector $\mathbf{R}$. The y axes are on a logarithmic scale.

inside a small energy window and the crystal is highly symmetric. The speedup of the WFPT interpolation will be much greater if a wider energy window is considered, or if the valence and conduction bands have a larger effective mass. Even for silicon, the speedup will reach the above-mentioned factor of over 50 as the number of points in the fine k grid is increased.

We note that the WFP interpolation of EPC-induced electron self-energy is conceptually completely different from the Wannier interpolation of the GW quasiparticle band structure [52]. In the latter, the GW quasiparticle energy is first calculated for all bands to be Wannierized and then Wannier interpolated. However, in the WFP interpolation, the matrix elements are Wannier interpolated, and then the self-energy is calculated. We briefly discuss why one should not directly interpolate the phonon-induced electron self-energy. First, one may try to calculate the self-energy on the coarse k grid before Wannierization. However, doing so is very inefficient because one needs to sample q points on a very fine grid to converge the self-energy. Second, one may try to interpolate only the sum of the upper Fan and DW terms, as it converges at a relatively coarse q grid. However, to do so, one must set $\mathcal{W}_{\text{Lower}}$ to be outside the *outer* window $\mathcal{W}_{\text{D}}$ because the upper Fan term should be calculated for all states inside $\mathcal{W}_{\text{D}}$ for Wannierization. Then, the lower Fan terms cannot

TABLE II. Time required to calculate the phonon-induced electron self-energy for 379 k points using direct plane-wave calculations and WFPT interpolation.

	Time (cpu · h)	
	Plane wave	WFPT
SCF, DFPT (coarse q)	269	269
NSCF + Sternheimer (coarse k , coarse q)	...	424
Wannierization	...	171
WFPT interpolation (fine k , coarse q)	...	23
NSCF + Sternheimer (fine k , coarse q)	1241 ^a	...
Lower Fan (fine k , fine q)	4,450,591 ^a	2846
Total	4,452,101 ^a	3733

^aEstimated from a calculation with smaller numbers of k and q points.

be calculated using Wannier interpolation because the contribution from the states that are excluded from Wannierization, which are outside $\mathcal{W}_F$ but inside $\mathcal{W}_{\text{Lower}}$, is not captured. In our method, we get around all these problems by using WFPT to interpolate the upper Fan self-energy and by using electron and phonon WFs to efficiently calculate the EPC on a fine k and q grid.

B. Application 2: Shift spin current

The second application of WFPT is the shift spin current, which is a second-order dc spin current response to an ac electric field. The shift spin and shift charge conductivity can be written as [26]

$$\sigma_{\text{shift}}^{s,a;bc}(\Omega) = \frac{i\pi q^3}{2\hbar^2 V \Omega^2} \sum_{\mathbf{k}, m, n} (d_{m\mathbf{k}}^{s,b;a} v_{n\mathbf{k}}^c - d_{n\mathbf{k}}^{s,c;a} v_{m\mathbf{k}}^b) \times (f_{m\mathbf{k}} - f_{n\mathbf{k}}) \delta(\Omega + (\varepsilon_{m\mathbf{k}} - \varepsilon_{n\mathbf{k}})/\hbar), \quad (58)$$

where Ω is the frequency of the incident photon, s and a the spin polarization axis and flow direction of the current, and b and c the directions of the external electric field. In an actual calculation, we broaden the delta function by a Gaussian as in Eq. (57). The matrix element $d_{m\mathbf{k}}^{s,b;a}$ is called the generalized spin-velocity derivative [26] and is defined as

$$d_{m\mathbf{k}}^{s,b;a} = j_{m\mathbf{k}}^{s,ab} + \sum_p \frac{j_{mp\mathbf{k}}^{s;a} v_{p\mathbf{k}}^b}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}}^{(0)}} + \sum_p' \frac{v_{mp\mathbf{k}}^b j_{pn}^{s;a}}{\varepsilon_{n\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}}^{(0)}}, \quad (59)$$

where $\hat{j}^{s,ab} = -\{\hat{S}_s, [\hat{r}^a, [\hat{r}^b, \hat{H}^{(0)}]]\}/2\hbar^2$. We define $\hat{S}_0 = \hat{1}$ so that $s = 0$ represents the charge current. In contrast to its charge counterpart, the generalized spin-velocity derivative is not geometric: It cannot be written as a derivative with respect to $\mathbf{k}$. Thus, the shift spin current cannot be calculated using the conventional WF interpolation method.

By choosing $\hat{V}_{\mathbf{q}} = \hat{j}^{s;a} \delta_{\mathbf{q},0}$ and $\hat{V}'_{\mathbf{q}} = \hat{v}^b \delta_{\mathbf{q},0}$ in the definition of $S_{m\mathbf{k}}$ [Eq. (31)], we find

$$d_{m\mathbf{k}}^{s;b,a} = j_{m\mathbf{k}}^{s,ab} + S_{m\mathbf{k}} + S_{m\mathbf{k}}^*. \quad (60)$$

The matrix elements of $\hat{j}^{s,ab}$ can easily be Wannier interpolated via a direct calculation of the matrix elements in the plane-wave basis for coarse k points [53]. WFPT interpolation of $S_{m\mathbf{k}}$ enables the interpolation of $d_{m\mathbf{k}}^{s;b,a}$. With the matrix elements interpolated to a dense grid of k points, one can calculate the shift spin and shift charge current spectra using Eq. (58).

We calculate the shift spin and shift charge current of monolayer WTe_2 in the 2H phase [57]. In Fig. 6, we show the interpolated matrix elements for the shift spin and shift charge currents. For comparison, we also evaluate Eq. (59) exactly by solving the Sternheimer equation in the plane-wave basis and approximately by truncating Eq. (59) to include only the Wannierized bands. We find that this band truncation gives incorrect matrix elements. In contrast, the WFPT interpolation accurately reproduces the exact plane-wave basis result. We also find that the matrix element calculated with an exact WF-based interpolation of the generalized derivative of the interband dipole matrix element [16] agrees with the result of the WFPT

FIG. 6. (a) Band structure of monolayer WTe_2 . (b,c) Matrix element of the (b) generalized spin-velocity derivative and (c) generalized derivative of the velocity operator. The conduction band m is indicated by the red dashed curve in panel (a). The valence bands n for panels (b) and (c) are indicated by the green dotted and blue dash-dotted curves in panel (a), respectively.

FIG. 7. (a) Shift spin and (b) shift charge current of monolayer WTe_2 calculated using WFPT, WF with band truncation, and the difference between the two. In panel (b), the shift current spectrum calculated using an exact WF-based interpolation [16] is also shown.

interpolation for the charge case. Such an exact WF-based interpolation is not possible for the spin case.

By solving the Sternheimer equation, one can compute the shift spin current without any band-truncation error. WFPT enables such a band-truncation-error-free calculation at a much lower computational cost. To our knowledge, the shift spin current has not been studied using the Sternheimer equation-based method, which is devoid of the band-truncation error. Here, we use WFPT to calculate the shift spin conductivity without the truncation error and study the effect of band truncation.

Figure 7 shows the calculated shift spin and shift charge conductivity. We again find that the truncation of Eq. (59) gives considerable error in the shift spin and shift charge current spectra. Also, for the shift charge current, the result of the exact WF-based interpolation [16] agrees with the result of our WFPT interpolation.

The WF and WFP for WTe_2 are shown in Figs. 8(a) and 8(b). The WF shown is a d_{z^2} -orbital-like state of the W atom, with its spin almost fully polarized (99%) along the $+z$ direction. The WFP shown is for the spin-velocity perturbation $\hat{j}^{x,x}$. The result can be qualitatively understood by considering the spin and orbital parts separately. The spin operator $\hat{S}_x$ rotates the spin polarization by 180° around the x axis so that the spin of the WFP is almost fully polarized (97%) along the $-z$ direction. The velocity operator acts on the orbital part. Wave function perturbation due to the velocity operator is equivalent to the application of the position operator to the wave function [58]. Hence, the orbital part of the WFP will approximately be the orbital part of WF times x , which is indeed the case in Fig. 8(b). Figures 8(c)–8(e) show that the real-space matrix elements are spatially localized.

We summarize the computational cost for calculating the shift current in Table III. The WFPT interpolation method speeds up the shift current calculation by 3 orders of magnitude. We note that the calculation can be further optimized by using crystal symmetry to compute the Sternheimer matrix elements only in the irreducible wedge of the Brillouin zone, for both the coarse and the fine k grids. We leave such an optimization for future work.

C. Application 3: Spin Hall conductivity

Finally, we apply WFPT to the calculation of spin Hall conductivity. The key quantity for calculating the spin Hall conductivity is the spin Berry curvature [59]

FIG. 8. (a) Isosurface of the spin-up part of a WF. Isovalues are $+0.087$ and -0.087 times the maximum absolute value for the yellow and cyan isosurfaces, respectively. (b) Isosurface of the spin-down part of a WFP for the spin-velocity perturbation $j^{x,x}$. Isovalues are $+0.25$ and -0.25 times the maximum absolute value for the yellow and cyan isosurfaces, respectively. (c)–(e) Spatial decay of the (c) Hamiltonian, (d) spin-velocity, and (e) WFP matrix elements. Each data point corresponds to the maximum absolute value over all WF pairs and the directions a , b , and s for each $\mathbf{R}$. The y axes are on a logarithmic scale.

TABLE III. Time required to calculate the shift spin current of the WTe₂ monolayer using direct plane-wave calculations and WFPT interpolation.

	Time (cpu · h)	
	Plane wave	WFPT
SCF	1.9	1.9
NSCF + Sternheimer (coarse k)	...	47.2
Wannierization	...	0.2
WFPT interpolation (fine k)	...	3.0
NSCF + Sternheimer (fine k)	52,454.9 ^a	...
Total	52,456.8 ^a	52.4

^aEstimated from a calculation with a smaller number of k points.

$$\Omega_{m\mathbf{k}}^{s,ab} = 2 \sum_{n \neq m} \frac{\text{Im}[j_{m\mathbf{k}}^{s,a} v_{n\mathbf{k}}^b]}{(\varepsilon_{m\mathbf{k}} - \varepsilon_{n\mathbf{k}})^2 + \eta^2}. \quad (61)$$

Here, η is a positive infinitesimal, and a , b , and s are the directions of the external field, current flow, and the spin polarization, respectively. The intrinsic dc spin Hall conductivity is given by the Kubo formula as [59,60]

$$\sigma_{ab}^s = \frac{q}{\hbar V} \sum_{\mathbf{k}, m} f_{m\mathbf{k}} \Omega_{m\mathbf{k}}^{s,ab}. \quad (62)$$

Here, V is the volume of the unit cell, q the electron charge, and $f_{m\mathbf{k}}$ the Fermi-Dirac occupation factor. Contrary to the Berry curvature, the spin Berry curvature is not a geometric quantity. Thus, in principle, an infinite number of bands should be included to evaluate Eq. (61). For example, Ref. [60] included states up to 50 eV above the valence band maximum to converge the spin Hall conductivity of Si, Ge, GaAs, and AlAs. Moreover, a very fine k -point grid is required to achieve convergence, which calls for efficient interpolation methods such as Wannier interpolation. However, conventional Wannier interpolation methods for computing spin Hall conductivity [18,19] truncate the sum over band n in Eq. (61) to include only the Wannier-interpolated bands.

By solving the Sternheimer equation, one can accurately calculate the spin Berry curvature and the spin Hall conductivity without explicitly including the high-energy bands. To our knowledge, the spin Hall conductivity has not been studied using the Sternheimer-equation-based method. However, the computational cost of the Sternheimer approach is much higher than Wannier interpolation for the same k -point grid. WFPT enables an efficient Wannier-interpolation calculation of the spin Hall conductivity without any band-truncation error, taking the advantages of both the Sternheimer method and the WF method.

The spin Berry curvature can be calculated by WFPT interpolation similarly to the case of the spin-velocity

derivative. We set $\hat{V}_{\mathbf{q}} = \hat{j}^{s,a} \delta_{\mathbf{q},0}$ and $\hat{V}'_{\mathbf{q}} = \hat{v}^b \delta_{\mathbf{q},0}$ and use Eq. (39) to find

$$\frac{1}{2} \Omega_{m\mathbf{k}}^{s,ab} \approx K_{m\mathbf{k}} - \beta_{m\mathbf{k}}(0) + \beta_{m\mathbf{k}}(\eta) = k_{m\mathbf{k}}^{(H)} + \beta_{m\mathbf{k}}(\eta), \quad (63)$$

where

$$\beta_{m\mathbf{k}}(\eta) = \sum_{n=1}^{N_w} \frac{\text{Im}[\tilde{j}_{m\mathbf{k}}^{s,a(H)} \tilde{v}_{n\mathbf{k}}^{b(H)}]}{(\varepsilon_{m\mathbf{k}}^{(H)} - \varepsilon_{n\mathbf{k}}^{(H)})^2 + \eta^2}. \quad (64)$$

Here, we approximate that for $m\mathbf{k} \in \mathcal{W}_p$ and $n\mathbf{k} \notin \mathcal{W}_F$, $|\varepsilon_{m\mathbf{k}} - \varepsilon_{n\mathbf{k}}| \gg \eta$, so the η in the denominator can be ignored. In practice, we avoid the sums over pairs of fully occupied states $m\mathbf{k}$ and $n\mathbf{k}$ in Eqs. (61) and (64) since such contributions get canceled in Eq. (62) [14].

We calculate the spin Hall conductivity of monolayer WTe₂ in the 2H phase. In Fig. 9, we compare spin Hall conductivity at $T = 50$ K computed using the WFPT interpolation method, conventional Wannier interpolation with band truncation [19], and the slow but exact Sternheimer method in the plane-wave basis. The WFPT result agrees perfectly with the exact results, while the WF interpolation with band truncation shows some discrepancy, albeit small. Interestingly, the band-truncation error is present even for the hole-doped case, where all the relevant bands are separated from the boundary of the inner window by more than 3 eV. Table IV shows the speedup of WFPT interpolation compared to the plane-wave calculation for a $200 \times 200 \times 1$ k -point grid. WFPT gives a speedup of over 2 orders of magnitude, without lowering the accuracy by band truncation.

FIG. 9. Spin Hall conductivity of monolayer WTe₂ computed using WFPT, using WF with band truncation, and by solving the Sternheimer equation in the plane-wave basis. The dashed vertical lines indicate the valence band maximum and the conduction band minimum energies. The temperature is set to 50 K. The sum over k points in Eq. (62) is performed over an $N_k \times N_k \times 1$ grid, and η is the broadening parameter in Eq. (61). The plane-wave calculation is not performed for $N_k = 200$ due to the high computational cost.

TABLE IV. Time required to calculate the spin Hall conductivity of the WTe_2 monolayer using a $200 \times 200 \times 1$ k -point grid by direct plane-wave calculations and WFPT interpolation.

	Time (cpu · h)	
	Plane wave	WFPT
SCF	1.9	1.9
NSCF + Sternheimer (coarse k)	...	47.2
Wannierization	...	0.2
WFPT interpolation (fine k)	...	0.1
NSCF + Sternheimer (fine k)	13,113.7 ^a	...
Total	13,115.6 ^a	49.4

^aEstimated from a calculation with a smaller number of k points.

IV. DISCUSSION

If one is only interested in the valence bands of an insulator, one may choose all the windows, $\mathcal{W}_P$, $\mathcal{W}_F$, and $\mathcal{W}_D$, to include only the occupied bands. If WFPT is applied to the electronic dielectric response of such insulators, the WFP derived in this work simplifies to the result of Ref. [22], i.e., Eq. (8) without the second term on the right-hand side. We note, however, that even the Wannierization of only the valence bands may not be possible in the case of narrow-gap semiconductors or topological insulators where the orbital characters of the valence bands change rapidly with $\mathbf{k}$. In the case of such semiconductors, where the low-energy conduction bands need to be interpolated and thus disentanglement is required, or that of metals, the simplified results do not apply. Our formulation of WFPT is applicable to both isolated and entangled bands and provides a clear physical understanding of why the WFPs are spatially localized. The generalization of WFPT from isolated bands to entangled bands requires the deliberate use of energy windows and subspaces to separate the energy ranges. In addition, the changes in the Wannier basis functions and the resulting changes in the Hamiltonian matrix [Eq. (12)] must be addressed to correctly interpolate matrix elements involving the sum over an infinite number of bands as done in Eqs. (32) and (39). For WFs constructed without disentanglement, the Hamiltonian matrix elements do not change because the WFPs are spanned by eigenstates outside the outer window so that the Hamiltonian matrix elements between the WFPs and the unperturbed WFs [Eq. (18)] are zero. Only as these highly nontrivial problems are solved, the methods for accurately calculating the proper sum over an infinite number of bands can be developed (as summarized in Table I), and WFPT enables several new kinds of calculations that were beyond the computational reach.

WFPT could find wide applications. For EPC, WFPT opens up a way to systematically study the effect of phonon-induced band renormalization to quantities that require integration over fine k points. Phonon-assisted

optical properties are one example. We emphasize that the predictive calculation of the indirect absorption onset [Fig. 4(d)] is currently possible only with the WFPT interpolation. Another example is electronic transport. In transport calculations, the EPC-induced band renormalization is ignored or included only partially as a correction to the effective mass [61,62]. WFPT interpolation will allow the complete consideration of the renormalization by calculating the band renormalization at every k point in the fine grid for transport calculations. Also, a self-consistent calculation of the phonon-induced electron self-energy [39] without neglecting the k dependence of the self-energy may become feasible.

Zero-point renormalization and temperature dependence of band structures are also interesting by themselves. Calculation of the zero-point renormalization for a large set of materials began only very recently [38]. Also, most calculations are limited to the correction of the valence band maximum and conduction band minimum, partly due to the large computational cost of calculations for a large number of k points. WFPT will considerably reduce the barrier of the calculation of the zero-point and temperature-dependent renormalization for a large number of materials or for many k points. Combined with the automatic generation of Wannier functions [63], an efficient, automatized high-throughput calculation will also be possible.

Similar developments can also be made for the shift spin current. The interpolation scheme based on WFPs is necessary to calculate converged shift spin current spectra at a reasonable computational cost. Considering the non-geometric property of the spin-transport-related matrix elements, WFPT will be useful for studying a wide range of spin transport phenomena.

Another promising application of WFPT is the calculation of the nonadiabatic Born effective charge, a concept recently introduced in Refs. [64–66]. Calculation of the nonadiabatic Born effective charge of metals and doped semiconductors requires a fine k -point sampling of wave function perturbations for phonon perturbations [66]. WFPT interpolation allows such a calculation for all kinds of materials, not just for the special case where the low-energy conduction bands are isolated from the high-energy bands.

Finally, we discuss two ways the WFPT formalism itself can be expanded. First, in this work, we did not consider the localized representation of the perturbations. The concept of WFPs may be generalized to localized perturbations, for example, by using phonon Wannier functions [8,67]. Such a generalization may enable the study of phonon-induced band renormalization in complex heterostructures by stitching WF- and WFP-based real-space matrix elements. Second, considering time-periodic perturbations will allow the localized representation of time-dependent wave function perturbations. This generalization will be useful for the study of dynamic properties such as the

dielectric function and the higher-harmonic generation. Also, these two generalizations may enable the application of WFPT to *GW* calculations, especially in the calculation of frequency- and momentum-dependent screened Coulomb interaction using the Sternheimer equation [68] or the Lanczos method [69].

V. CONCLUSION

In conclusion, we introduced the concept of WFPT, which is the extension of the nearsightedness principle to wave function perturbations. We defined the WFP and showed that it gives an accurate localized representation of wave function perturbations. Using WFPs, various matrix elements involving the wave function perturbation can be efficiently interpolated. We applied WFPT to the calculation of the temperature-dependent electron band structures, the shift spin conductivity, and the spin Hall conductivity. The significant speedup gained thanks to WFPT enabled the study of phenomena such as the temperature-dependent onset of the indirect optical absorption and the shift spin currents, which were beyond the reach of previous methods. As density-functional perturbation theory has significantly expanded the applications of density functional theory, WFPT will also widen the scope of Wannier function methods for the theoretical and computational study of condensed matter.

VI. METHODS

We used the QUANTUM ESPRESSO package to perform plane-wave pseudopotential DFT calculations [70,71]. We used optimized norm-conserving Vanderbilt pseudopotentials [72] taken from the PseudoDojo library (v0.4) [73]. For silicon, we used the Perdew-Wang functional [74] and a scalar-relativistic pseudopotential. We used a kinetic energy cutoff of 70 Ry and a $14 \times 14 \times 14$ k -point grid for DFT and DFPT calculations. DFPT calculation was done on a $6 \times 6 \times 6$ q -point grid. The experimental lattice parameter at 300 K ($a = 5.43$ Å) was used [75]. Thermal expansion was not taken into account as its effect on the indirect band gap is negligible [11,76]. For the WTe₂ monolayer, we used the Perdew-Burke-Ernzerhof functional [77] and included spin-orbit coupling using a fully relativistic pseudopotential. We used a kinetic energy cutoff of 70 Ry and a $12 \times 12 \times 1$ k -point grid. The lattice structure was optimized until the internal stresses and forces were below 0.01 kbar and 2×10^5 Ry/Bohr, respectively. The optimized lattice parameters of a rectangular in-plane supercell were $a = 3.56$ Å and $b = 6.17$ Å, consistent with a previous calculation [57]. The supercell length along the nonperiodic direction was set to 15 Å.

To construct the MLWFs, we used the Wannier90 package [78]. For silicon, we constructed 10 MLWFs from 18 bands to accurately interpolate the bands near the conduction band minimum. We used sp^3 orbitals centered

at the Si atom and s orbitals centered at the two interstitial sites as the initial guesses. The upper bounds of the perturbation window ($\mathcal{W}_P$), the inner window of the unperturbed system ($\mathcal{W}_F$), and the outer window ($\mathcal{W}_D$) were 5.6 eV, 6.1 eV, and 18.5 eV above the conduction band minimum, respectively. A $12 \times 12 \times 12$ k -point grid was used. For WTe₂, we constructed 28 MLWFs from 32 bands, using W s and d , and Te s and p orbitals as initial guesses. We excluded the semicore states in the Wannierization. The upper bound of the perturbation window and the inner window of the unperturbed system were 2.8 eV and 3.3 eV above the valence band maximum, respectively. The outer window was not set so that all 32 bands are used for Wannierization. A $12 \times 12 \times 1$ k -point grid was used for the shift current and spin Hall applications. In the Berry curvature calculation, we used a $24 \times 24 \times 1$ grid to reduce the finite-difference error in the position matrix elements [5]. For plotting the isosurfaces of the WFs and WFPs for WTe₂, we used a coarser $6 \times 6 \times 6$ k -point grid due to the high computational cost. For all Wannier interpolation, we used the minimal-distance replica selection based on the Wigner-Seitz supercell that improves the quality of interpolation [15].

The lower Fan term in the phonon-induced electron self-energy of silicon was calculated using a $125 \times 125 \times 125$ q -point grid. The denominator was smoothed with $\eta = 10$ meV [Eq. (49)]. We used the on-the-mass-shell approximation (Rayleigh-Schrödinger perturbation), which is known to give a better estimate of the zero-point renormalization than the Dyson-Migdal approach [37,39]. The upper Fan and the DW self-energies were calculated on a coarse $6 \times 6 \times 6$ q -point grid without interpolation over q points. To speed up the convergence of the sum of the upper Fan and the DW self-energies, we used the improved double-grid method as detailed in Appendix G. See Sec. S2A of the Supplemental Material [53] for the convergence study of the lower Fan term.

The G_0W_0 calculation of the quasiparticle band structure of silicon was done using BerkeleyGW [79], within the plasmon-pole approximation [43]. We converged the direct and indirect band gaps to within 5 meV. We used a $12 \times 12 \times 12$ q -point grid, a cutoff of 30 Ry and 402 bands for the dielectric matrix, and 308 bands with the static remainder correction [80] for the self-energy. The quasiparticle energies were calculated on a $12 \times 12 \times 12$ k -point grid and then Wannier interpolated [52] using the EPW code [81]. The velocity matrix elements for the indirect optical absorption were renormalized following Ref. [82].

The indirect absorption spectra of silicon were calculated using the EPW code [81]. We used a 16-band WF model to include a sufficient number of intermediate bands. We checked that the indirect absorption spectra change only slightly when more bands are included. We used $60 \times 60 \times 60$ k - and q -point grids. We sampled the initial

and final states only in the ± 1 eV window around the center of the band gap. Using the lattice symmetry, we only sampled the k points in the irreducible wedge of the Brillouin zone. In total, 379k points were sampled. The imaginary part of the self-energy of intermediate electronic states was set to 50 meV. Delta functions for energy conservation were broadened using a Gaussian function of width 15 meV. We used the experimental frequency-dependent refractive index at 300 K [83]. See Sec. S2B of the Supplemental Material [53] for the convergence study.

The shift conductivity of monolayer WTe_2 was calculated using a $400 \times 400 \times 1$ k -point grid. Delta functions for energy conservation were broadened using a Gaussian function of width 10 meV. Following the scheme of Refs. [16,84], we regularized the denominators including an intermediate-state energy with $\eta = 10$ meV, which is in the range where the spectrum calculated using the given k -point grid remains stable [16,84]. See Sec. S2C of the Supplemental Material [53] for the convergence study.

The spin Hall conductivity of monolayer WTe_2 was calculated using a $200 \times 200 \times 1$ k -point grid. Note that $\eta = 10$ meV was used in the calculation of the spin Berry curvature [Eq. (61)]. We used the Smoother module of the WannierBerri package [85] to compute the spin Hall conductivity at a finite temperature. The response was scaled using an effective single-layer thickness of 7.1 Å,

which is taken from the bulk lattice parameter [86]. See Sec. S2D of the Supplemental Material [53] for the convergence study.

The WFPs and the corresponding matrix elements in the coarse k grid were calculated using an in-house modified version of the ph.x program of QUANTUM ESPRESSO [71]. We used in-house modified versions of the EPW [81] and Wannier90 [78] code for the EPC, and the shift spin and spin Hall current applications, respectively. In the EPC calculation, we used the crystal symmetry to efficiently unfold the electron-phonon, DW, and Sternheimer matrix elements calculated for the q points in the irreducible wedge to the full coarse q -point grid. We also implemented the optimization [85] of the minimal-distance replica selection method that significantly reduces the computational cost in the in-house versions of EPW and Wannier90. Isosurface plots were created using VESTA [87].

ACKNOWLEDGMENTS

This work was supported by the Creative-Pioneering Research Program through Seoul National University, the NRF of Korea No. 2020R1A2C1014760, and the Institute for Basic Science (No. IBSR009-D1). Computational resources have been provided by KISTI (KSC-2020-INO-0078).

APPENDIX A: DERIVATION OF THE WANNIER FUNCTION PERTURBATION

In this appendix, we derive Eq. (8), the formula for the WFP. First, we find the disentanglement subspace following the initialization scheme of Ref. [6]. Concretely, we project $\hat{P}_W^{(0)}$, the projection operator for the unperturbed Wannier subspace, onto the subspace $\mathcal{W}_D - \mathcal{W}_P$ and diagonalize it. We select $N_k N_W - N_P$ states with the largest eigenvalues [6] and use them as the basis states for the perturbed WFs, together with N_P states inside $\mathcal{W}_P$.

From Eqs. (1), (3), and (4), we find that for states $n\mathbf{k}, n'\mathbf{k}' \in \mathcal{W}_D$, the matrix element of $\hat{P}_W^{(0)}$ is

$$\begin{aligned}
 P_{n'\mathbf{k}', n\mathbf{k}} &= \langle \psi_{n'\mathbf{k}'} | \hat{P}_W^{(0)} | \psi_{n\mathbf{k}} \rangle = \sum_{\mathbf{R}} \sum_{i=1}^{N_W} \langle \psi_{n'\mathbf{k}'} | w_{i\mathbf{R}}^{(0)} \rangle \langle w_{i\mathbf{R}}^{(0)} | \psi_{n\mathbf{k}} \rangle \\
 &= \frac{1}{N_k} \sum_{\mathbf{R}} \sum_{i=1}^{N_W} U_{n'i;\mathbf{k}'}^{(0)} U_{in;\mathbf{k}}^{(0)\dagger} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{R}} \\
 &\quad + \lambda \frac{1}{N_k} \sum_{\mathbf{R}} \sum_{m, \mathbf{k}''} \sum_{i=1}^{N_W} \left[U_{n'i;\mathbf{k}'}^{(0)} U_{im;\mathbf{k}''}^{(0)\dagger} e^{i(\mathbf{k}''-\mathbf{k}') \cdot \mathbf{R}} \frac{\langle \psi_{m\mathbf{k}''}^{(0)} | \hat{V} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{m\mathbf{k}''}^{(0)}} + (n\mathbf{k} \leftrightarrow n'\mathbf{k}')^* \right] + \mathcal{O}(\lambda^2) \\
 &= P_{W, n'n; \mathbf{k}}^{(0)} \delta_{\mathbf{k}, \mathbf{k}'} + \lambda \sum_m \left(P_{W, n'm; \mathbf{k}'}^{(0)} \frac{\langle \psi_{m\mathbf{k}'}^{(0)} | \hat{V} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{m\mathbf{k}'}^{(0)}} - \frac{\langle \psi_{n'\mathbf{k}'}^{(0)} | \hat{V} | \psi_{m\mathbf{k}}^{(0)} \rangle}{\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{n'\mathbf{k}'}^{(0)}} P_{W, mn; \mathbf{k}}^{(0)} \right) + \mathcal{O}(\lambda^2). \tag{A1}
 \end{aligned}$$

Here, we defined

$$P_{W, n'n; \mathbf{k}}^{(0)} = \sum_{i=1}^{N_W} U_{n'i;\mathbf{k}}^{(0)} U_{in;\mathbf{k}}^{(0)\dagger}. \tag{A2}$$

We can obtain the eigenstates of $P_{n'\mathbf{k}',n\mathbf{k}}$ using first-order perturbation theory. The zeroth-order matrix is $P_{W,n';n\mathbf{k}}^{(0)} \delta_{\mathbf{k},\mathbf{k}'}$. We write its eigenvectors as $\tilde{U}_{np;\mathbf{k}}^{(0)}$, where the eigenvalues for $p = 1, \dots, N_W$ are 1 (in subspace ① and ②) and those for $p = N_W + 1, \dots, N_{D,\mathbf{k}}$ are 0 (in subspace ③). Since $\mathcal{W}_F$ is the inner window for unperturbed WFs, one can choose the eigenvectors in $\mathcal{W}_F$ to satisfy

$$\tilde{U}_{np;\mathbf{k}}^{(0)} = \delta_{n,p} \quad \text{for } p = 1, \dots, N_{F,\mathbf{k}}. \quad (\text{A3})$$

Since the basis states are the perturbed wave functions, the zeroth-order eigenstates of $P_{n'\mathbf{k}',n\mathbf{k}}$ are

$$|\phi_{p\mathbf{k}}^{(0)}\rangle = \sum_{n=1}^{N_{D,\mathbf{k}}} |\psi_{n\mathbf{k}}\rangle \tilde{U}_{np;\mathbf{k}}^{(0)}. \quad (\text{A4})$$

The N_P energy eigenstates inside $\mathcal{W}_P$, the inner window for the perturbed system, are chosen without further disentanglement:

$$\{|\phi_{p\mathbf{k}}^{(0)}\rangle : p = 1, \dots, N_{P,\mathbf{k}}\}. \quad (\text{A5})$$

Next, we choose the remaining $N_{\mathbf{k}}N_W - N_P$ states for the disentangled subspace as follows. We apply first-order perturbation theory to $|\phi_{p\mathbf{k}}^{(0)}\rangle$ for $p = N_{P,\mathbf{k}} + 1, \dots, N_W$, which are the zeroth-order eigenstates of $P_{n'\mathbf{k}',n\mathbf{k}}$ inside $\mathcal{W}_D - \mathcal{W}_P$ with eigenvalue 1. This choice maximizes the perturbed eigenvalues because the unperturbed eigenvalues of the states that are not selected are 0. The mixing between the selected states is not important because our goal here is to determine the subspace spanned by them. The WFs that are obtained by projection depend only on the selected subspace, not on its basis [6]. Hence, we use nondegenerate perturbation theory and only consider the mixing of the eigenstates of $P_{W;\mathbf{k}}^{(0)}$ with eigenvalue 1 with those with eigenvalue 0.

Writing the first-order matrix elements in Eq. (A1) as $P_{n'\mathbf{k}',n\mathbf{k}}^{(1)}$, we find that $|\phi_{p\mathbf{k}}^{(0)}\rangle$ is perturbed as

$$\begin{aligned} |\phi_{p\mathbf{k}}\rangle - |\phi_{p\mathbf{k}}^{(0)}\rangle &= \lambda \sum_{\mathbf{k}'} \sum_{q=N_W+1}^{N_{D,\mathbf{k}'}} \sum_{n,n'} |\phi_{q\mathbf{k}'}^{(0)}\rangle \langle \phi_{q\mathbf{k}'}^{(0)} | \psi_{n'\mathbf{k}'} \rangle P_{n'\mathbf{k}',n\mathbf{k}}^{(1)} \langle \psi_{n\mathbf{k}} | \phi_{p\mathbf{k}}^{(0)} \rangle + \mathcal{O}(\lambda^2) \\ &= \lambda \sum'_{n,m,n',\mathbf{k}'} \sum_{q=N_W+1}^{N_{D,\mathbf{k}'}} |\phi_{q\mathbf{k}'}^{(0)}\rangle \tilde{U}_{qn';\mathbf{k}'}^{(0)\dagger} \left(P_{W,n';m;\mathbf{k}'}^{(0)} \frac{\langle \psi_{m\mathbf{k}'}^{(0)} | \hat{V} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{m\mathbf{k}'}^{(0)}} - \frac{\langle \psi_{n'\mathbf{k}'}^{(0)} | \hat{V} | \psi_{m\mathbf{k}}^{(0)} \rangle}{\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{n'\mathbf{k}'}^{(0)}} P_{W,mn;\mathbf{k}}^{(0)} \right) \tilde{U}_{np;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) \\ &= -\lambda \sum_m (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2). \end{aligned} \quad (\text{A6})$$

In the third equality, we used the fact that for $p = N_{P,\mathbf{k}} + 1, \dots, N_W$ and $q = N_W + 1, \dots, N_{D,\mathbf{k}}$, $\tilde{U}_{np;\mathbf{k}}^{(0)}$ and $\tilde{U}_{qn';\mathbf{k}'}^{(0)\dagger}$ are the right and left eigenvectors of $P_{W;\mathbf{k}}^{(0)}$ with eigenvalues 1 and 0, respectively.

The unperturbed state $|\phi_{p\mathbf{k}}^{(0)}\rangle$ defined in Eq. (A4) also contains an $\mathcal{O}(\lambda^1)$ contribution because the basis states are $|\psi_{n\mathbf{k}}\rangle$'s, the perturbed eigenstates. Using Eq. (4), the perturbative expansion of the disentangled state $|\phi_{p\mathbf{k}}\rangle$ becomes

$$\begin{aligned} |\phi_{p\mathbf{k}}\rangle &= \sum_n |\psi_{n\mathbf{k}}^{(0)}\rangle \tilde{U}_{np;\mathbf{k}}^{(0)} \\ &\quad + \lambda \sum_m (\hat{\mathbb{I}} - (\hat{P}_D^{(0)} - \hat{P}_W^{(0)})) |\psi_{m\mathbf{k}}^{(1)}\rangle \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) \end{aligned} \quad (\text{A7})$$

for $p = N_{P,\mathbf{k}} + 1, \dots, N_W$.

Combining results in Eqs. (A3) and (A7), the final disentangled subspace for the perturbed system is spanned by

$$\begin{aligned} &\cup_{\mathbf{k}} (\{|\phi_{p\mathbf{k}}^{(0)}\rangle : p = 1, \dots, N_{P,\mathbf{k}}\} \\ &\cup \{|\phi_{p\mathbf{k}}\rangle : p = N_{P,\mathbf{k}} + 1, \dots, N_W\}). \end{aligned} \quad (\text{A8})$$

The dimension of the disentangled subspace is $N_{\mathbf{k}}N_W$, as desired.

Next, to make computation simpler, we linearly combine the basis states and write them in a simpler form. For $\lambda = 0$, the disentangled subspace [Eq. (A8)] is identical to the unperturbed Wannier subspace. Hence, we can linearly combine the basis states to make their first-order terms orthogonal to the unperturbed Wannier subspace. Mathematically, this orthogonalization is equivalent to applying $(\hat{\mathbb{I}} - \hat{P}_W^{(0)})$ to the first-order terms:

$$|\tilde{\phi}_{p\mathbf{k}}\rangle = \begin{cases} \sum_m [\psi_{m\mathbf{k}}^{(0)} + \lambda(\hat{1} - \hat{P}_W^{(0)})|\psi_{m\mathbf{k}}^{(1)}\rangle] \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) & \text{for } p = 1, \dots, N_{P,\mathbf{k}}, \\ \sum_m [\psi_{m\mathbf{k}}^{(0)} + \lambda(\hat{1} - \hat{P}_D^{(0)})|\psi_{m\mathbf{k}}^{(1)}\rangle] \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) & \text{for } p = N_{P,\mathbf{k}} + 1, \dots, N_W. \end{cases} \quad (\text{A9})$$

This application of $(\hat{1} - \hat{P}_W^{(0)})$ guarantees that the basis states for the disentangled subspace, $|\tilde{\phi}_{p\mathbf{k}}\rangle$'s, are orthonormal to each other with errors of order $\mathcal{O}(\lambda^2)$. Although the final form of the WFs of the perturbed system is not affected by this orthonormalization, it makes the following analysis simpler.

Using Eq. (A3) and the fact that $1 \leq m \leq N_{P,\mathbf{k}}$ is equivalent to $m\mathbf{k} \in \mathcal{W}_P$, we find

$$|\tilde{\phi}_{p\mathbf{k}}\rangle = \sum_m [\psi_{m\mathbf{k}}^{(0)} + \lambda \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle] \tilde{U}_{mp;\mathbf{k}}^{(0)} + \lambda \sum_{m\mathbf{k} \in \mathcal{W}_P} [(\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle] \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2). \quad (\text{A10})$$

Finally, we compute the perturbed WFs. We follow the scheme of Ref. [5], where the initial guess functions are projected onto the disentangled subspace and then orthonormalized by Löwdin orthogonalization. This procedure does not change the subspace but fixes the unitary transformation within the subspace to give localized WFs. The overlap matrix A between the initial guesses and the disentangled basis states is

$$A_{n\mathbf{k},i\mathbf{R}} \equiv \langle \tilde{\phi}_{n\mathbf{k}} | w_{i\mathbf{R}}^{(0)} \rangle = \frac{1}{\sqrt{N_k}} (\tilde{U}_{\mathbf{k}}^{(0)\dagger} U_{\mathbf{k}}^{(0)})_{ni} e^{-i\mathbf{k} \cdot \mathbf{R}} + \mathcal{O}(\lambda^2). \quad (\text{A11})$$

The first-order term of $A_{n\mathbf{k},i\mathbf{R}}$ vanishes because we linearly combined the disentangled basis functions so that their first-order term is orthogonal to the unperturbed Wannier subspace [Eq. (A9)]. We perform Löwdin orthogonalization to find the gauge matrix and multiply it by $|\tilde{\phi}_{n\mathbf{k}}\rangle$ to obtain the WFs for the perturbed system:

$$|w_{i\mathbf{R}}(\lambda)\rangle = \sum_{n\mathbf{k}} |\tilde{\phi}_{n\mathbf{k}}\rangle [A(A^\dagger A)^{-1/2}]_{n\mathbf{k},i\mathbf{R}}. \quad (\text{A12})$$

From Eq. (A11), one can show that the gauge matrix satisfies

$$[A(A^\dagger A)^{-1/2}]_{n\mathbf{k},i\mathbf{R}} = A_{n\mathbf{k},i\mathbf{R}} + \mathcal{O}(\lambda^2). \quad (\text{A13})$$

Using Eqs. (A10)–(A13), we obtain Eqs. (7) and (8).

APPENDIX B: NECESSITY OF ALL TERMS IN THE WFP

In this section, we discuss four possible modifications to the WFP formula [Eq. (8)] that might seem reasonable but turn out to be undesirable. The two most important properties of the WFPs are that they accurately represent the wave function perturbations and that they are spatially localized. We demonstrate that the modifications make the WFPs lose either one of these two properties.

Let us consider the following cases.

Case 0: The true WFPs as defined in Eq. (8).

Case 1: Keep only the first term in Eq. (8). This case is equivalent to using no inner window for the Wannierization of the perturbed system or, in other words, $\mathcal{W}_P = \emptyset$.

$$|w_{i\mathbf{R}}^{(1),\text{Case 1}}\rangle = \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}. \quad (\text{B1})$$

Case 2: Keep only the second term in Eq. (8). In other words, ignore the contribution from states in subspace ④, which are outside the outer window $\mathcal{W}_D$.

$$|w_{i\mathbf{R}}^{(1),\text{Case 2}}\rangle = \frac{1}{\sqrt{N_k}} \sum_{\substack{m,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_P}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}. \quad (\text{B2})$$

Case 3: Do not impose the condition $m\mathbf{k} \in \mathcal{W}_P$ in the second term of Eq. (8).

$$\begin{aligned} |w_{i\mathbf{R}}^{(1),\text{Case 3}}\rangle &= \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} \\ &\quad + \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} \\ &= \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} (\hat{1} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}. \end{aligned} \quad (\text{B3})$$

Case 4: Impose the condition $m\mathbf{k} \in \mathcal{W}_P$ not only in the second term but also in the first term of Eq. (8). This case is equivalent to not using an outer window for the Wannierization of the perturbed system to include all bands or, in other words, $\mathcal{W}_D = (-\infty, \infty)$.

$$\begin{aligned}
|w_{\mathbf{R}}^{(1), \text{Case 4}}\rangle &= \frac{1}{\sqrt{N_k}} \sum_{\substack{m, \mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_p}} \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi, \mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} + \frac{1}{\sqrt{N_k}} \sum_{\substack{m, \mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_p}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi, \mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} \\
&= \frac{1}{\sqrt{N_k}} \sum_{\substack{m, \mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_p}} (\hat{1} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi, \mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}.
\end{aligned} \tag{B4}$$

To study whether the modified WFPs can accurately and efficiently interpolate the wave function perturbation, we calculate the following matrix element, which we call the “smoothed Sternheimer matrix element”:

$$\begin{aligned}
X_{mn\mathbf{k}; \mathbf{q}} &= \sum_{\substack{p=1 \\ p\mathbf{k} + \mathbf{q} \in \mathcal{W}_p}}^{\infty} \frac{(h_{p\mathbf{m}\mathbf{k}; \mathbf{q}\nu})^* h_{p\mathbf{n}\mathbf{k}; \mathbf{q}\nu}}{\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(0)}} \times f(\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(0)}) \\
&+ \sum_{\substack{p=1 \\ p\mathbf{k} + \mathbf{q} \notin \mathcal{W}_p}}^{\infty} \frac{(h_{p\mathbf{m}\mathbf{k}; \mathbf{q}\nu})^* h_{p\mathbf{n}\mathbf{k}; \mathbf{q}\nu}}{\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(0)}},
\end{aligned} \tag{B5}$$

where

$$f(x) = 1 - e^{-(x/\sigma)^2}. \tag{B6}$$

Here, we defined

$$h_{mn\mathbf{k}; \mathbf{q}\nu} = \sum_{\kappa, a} \langle u_{m\mathbf{k} + \mathbf{q}} | \partial_{\mathbf{q}\kappa a} \hat{H}_{\text{KS}} | u_{n\mathbf{k}} \rangle U_{\kappa a, \nu}(\mathbf{q}), \tag{B7}$$

where $\partial_{\mathbf{q}\kappa a} \hat{H}_{\text{KS}}$ is the derivative of the Kohn-Sham Hamiltonian with respect to a monochromatic displacement of atom κ along Cartesian direction a with wave vector $\mathbf{q}$ and $U_{\kappa a, \nu}(\mathbf{q})$ the displacement pattern for phonon eigenmode ν in units of the inverse square root of mass. The h matrix element is related to the electron-phonon matrix element g , which is in units of energy, as

$$g_{mn\mathbf{k}; \mathbf{q}\nu} = \frac{1}{\sqrt{2\omega_{\mathbf{q}\nu}}} h_{mn\mathbf{k}; \mathbf{q}\nu}. \tag{B8}$$

The form of the matrix element $X_{mn\mathbf{k}; \mathbf{q}}$ in Eq. (B5) is similar to the static Fan self-energy except that the smoothing factor $f(\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(0)})$ suppresses the contribution from states $p\mathbf{k} + \mathbf{q}$ that are energetically close to $m\mathbf{k}$. Since our goal is to study whether the contribution of the high-energy bands is accurately captured, we apply this factor to suppress the contribution from the low-energy bands. In the following, we set $\sigma = 0.5$ eV.

We calculate the smoothed Sternheimer matrix elements using the WFPT interpolation and the exact plane-wave evaluation of Eq. (B5). For the WFPT calculation, we use Eq. (36) to find

$$\begin{aligned}
X_{mn\mathbf{k}; \mathbf{q}} &= \sum_{\substack{p=1 \\ p\mathbf{k} + \mathbf{q} \in \mathcal{W}_p}}^{N_W} \frac{(\tilde{h}_{p\mathbf{m}\mathbf{k}; \mathbf{q}\nu}^{(H)})^* h_{p\mathbf{n}\mathbf{k}; \mathbf{q}\nu}^{(H)}}{\epsilon_{m\mathbf{k}}^{(H0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(H0)}} \times f(\epsilon_{m\mathbf{k}}^{(H0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(H0)}) \\
&+ \sum_{\substack{p=1 \\ p\mathbf{k} + \mathbf{q} \notin \mathcal{W}_p}}^{N_W} \frac{(\tilde{h}_{p\mathbf{m}\mathbf{k}; \mathbf{q}\nu}^{(H)})^* h_{p\mathbf{n}\mathbf{k}; \mathbf{q}\nu}^{(H)}}{\epsilon_{m\mathbf{k}}^{(H0)} - \epsilon_{p\mathbf{k} + \mathbf{q}}^{(H0)}} + s_{mn\mathbf{k}; \mathbf{q}\nu}^{(H)}.
\end{aligned} \tag{B9}$$

Note that $\tilde{h}$ is equal to h for states inside $\mathcal{W}_p$, just as we discussed for $\tilde{g}$ and g in Sec. IID. For the plane-wave calculation, we solve the following Sternheimer equation to evaluate the infinite sum in the second term of Eq. (B5), in which we use the fact that $p\mathbf{k} + \mathbf{q} \notin \mathcal{W}_p$ if $p \geq 19$.

$$\begin{aligned}
&(\hat{H}_{\text{KS}} - \epsilon_{m\mathbf{k}}) |\delta_{\mathbf{q}\kappa a} \psi_{m\mathbf{k}}\rangle \\
&= - \left(\sum_{p=19}^{\infty} |\psi_{p\mathbf{k} + \mathbf{q}}\rangle \langle \psi_{p\mathbf{k} + \mathbf{q}}| \right) (\partial_{\mathbf{q}\kappa a} \hat{H}_{\text{KS}}) |\psi_{m\mathbf{k}}\rangle.
\end{aligned} \tag{B10}$$

The contribution of the bands with $p \leq 18$ to Eq. (B5) is obtained by directly calculating the eigenvalues and the matrix elements.

We calculate the smoothed Sternheimer matrix element of silicon for $\mathbf{q} = 0$ and $m = n$ along the band path shown in Fig. 10. Figure 11 shows the comparison for the five cases and for all six phonon modes. For case 0, the Wannier interpolation is accurate for all k points and phonon modes. For cases 1 and 2, the Wannier interpolation does not reproduce the exact result, even at the coarse k points denoted by the vertical dashed lines. This result indicates

FIG. 10. Band structure of silicon. Matrix elements shown in Fig. 11 are for the states on the band denoted with the red dashed curve.

FIG. 11. Smoothed Sternheimer matrix elements for silicon obtained from an exact calculation using the plane-wave basis and from interpolations using different modifications of the WFP formula. The vertical dashed lines indicate k points on the coarse k grid where the interpolation should be exact.

that the two terms in Eq. (8) are both required to make the WFPs accurately represent the perturbation of the wave functions. For cases 3 and 4, the Wannier interpolation is accurate at the coarse k points, indicating that the WFPs at least represent the wave function perturbations. This property can be easily shown from Eqs. (B3) and (B4) as the summands are exact wave function perturbations for $m\mathbf{k} \in \mathcal{W}_p$. However, the interpolation displays wiggles [Figs. 11(v)–(ab)], which imply that the WFPs and the matrix elements are not localized enough.

The spatial decay of the WFP matrix elements for cases 0, 3, and 4 are shown in Fig. 12. (We did not plot the decay for cases 1 and 2 because the WFPs for cases 1 and 2 fail to represent the wave function perturbation so they are of no use even if they are spatially localized.) We find that the matrix elements in cases 3 and 4 are much more delocalized than those in case 0, consistent with the interpolation quality in Fig. 11.

FIG. 12. Spatial decay of the WFP matrix elements for silicon. Each data point corresponds to the maximum absolute value over all WF pairs and atomic displacements for each lattice vector $\mathbf{R}$. The y axes are on a logarithmic scale. Only the $\mathbf{q} = 0$ case is considered.

This result can be understood as follows. For case 3, because the condition $m\mathbf{k} \in \mathcal{W}_P$ is removed, the denominator in Eq. (9) can be close to zero, making the summand a rapidly varying and diverging function of $\mathbf{k}$. Thus, the corresponding WFPs are highly delocalized. For case 4, the outer window for Wannierization is set for the unperturbed system but not for the perturbed system. Hence, the difference between the unperturbed and perturbed WFs will be greater than in case 0 where the outer window is consistently used. Therefore, the WFPs constructed with case 4 are more delocalized than case 0.

In passing, we note that the good quality of the interpolation for the acoustic modes for case 3 [Figs. 11(s)–11(u)] is an exceptional case resulting from the acoustic sum rule. Because of the translational invariance, the matrix elements of the acoustic phonon modes at $\mathbf{q} = 0$ satisfy

$$h_{mn\mathbf{k};0\nu} \sim \langle \psi_{m\mathbf{k}} | i[\hat{H}, \hat{p}] | \psi_{n\mathbf{k}} \rangle = i(\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{n\mathbf{k}}^{(0)}) p_{mn\mathbf{k}}, \quad (\text{B11})$$

where $p_{mn\mathbf{k}}$ is the matrix element of the momentum operator [35]. Thus, when the denominator in Eq. (9) becomes close to zero, the numerator also becomes small.

To summarize, all four modifications to the definition of the WFPs considered in this section either fail to produce WFPs that accurately represent the wave function perturbations or generate delocalized WFPs.

APPENDIX C: SMOOTHED INNER WINDOW FOR WFPT

In this section, we discuss how to remove the discontinuity due to the $m\mathbf{k} \in \mathcal{W}_P$ condition in Eq. (8) by smoothing the inner window in the disentanglement step for constructing the projection-only WFs. The issue we aim to solve is that the inner window is applied in an “all-or-nothing” manner. So, the solution is to assign weights to the eigenstates that are 1 for states inside the inner window and continuously decay to 0 outside the inner window.

Concretely, we define a coefficient $l_{n\mathbf{k}}$ for each state $n\mathbf{k}$ inside $\mathcal{W}_F - \mathcal{W}_P$. The coefficient should decay from 1 to 0 as the unperturbed eigenvalue moves away from the boundary energy for $\mathcal{W}_P$. In Appendix A, in the disentanglement step, we projected $\hat{P}_W^{(0)}$ to the perturbed eigenstates inside $\mathcal{W}_D - \mathcal{W}_P$ and diagonalized the resulting

matrix $P_{n'\mathbf{k}',n\mathbf{k}}$ [Eq. (A1)]. Here, instead of projecting to $\mathcal{W}_D - \mathcal{W}_P$, we use the following operator:

$$\hat{Q}_{\text{smooth}} = \sum_{n\mathbf{k} \in \mathcal{W}_D - \mathcal{W}_F} |\psi_{n\mathbf{k}}\rangle \langle \psi_{n\mathbf{k}}| + \sum_{n\mathbf{k} \in \mathcal{W}_F - \mathcal{W}_P} \frac{1}{1 - l_{n\mathbf{k}}} |\psi_{n\mathbf{k}}\rangle \langle \psi_{n\mathbf{k}}|. \quad (\text{C1})$$

For each state $n\mathbf{k}$ inside $\mathcal{W}_F - \mathcal{W}_P$, the weight of the projector $|\psi_{n\mathbf{k}}\rangle \langle \psi_{n\mathbf{k}}|$ in $\hat{Q}_{\text{smooth}}$ increases as $\varepsilon_{n\mathbf{k}}^{(0)}$ approaches the boundary energy of $\mathcal{W}_P$ and $l_{n\mathbf{k}}$ approaches 1. Then, we diagonalize

$$\hat{P}_{\text{smooth}} = \hat{Q}_{\text{smooth}} \hat{P}_W^{(0)} \hat{Q}_{\text{smooth}} \quad (\text{C2})$$

and take the $N_{\mathbf{k}}N_W - N_P$ states with the largest eigenvalues. If all $l_{n\mathbf{k}} = 0$,

$$\hat{Q}_{\text{smooth}} = \sum_{n\mathbf{k} \in \mathcal{W}_D - \mathcal{W}_P} |\psi_{n\mathbf{k}}\rangle \langle \psi_{n\mathbf{k}}| \quad (\text{C3})$$

holds and the original scheme of Ref. [6] is obtained.

Now, we calculate the eigenstates of $\hat{P}_{\text{smooth}}$ using perturbation theory. Writing the zeroth-order eigenstates as $|\phi_{p\mathbf{k}}^{(0)}\rangle$ for $p = N_{P,\mathbf{k}} + 1, \dots, N_{D,\mathbf{k}}$, and using Eq. (A3), we have

$$\hat{P}_{\text{smooth}}(\lambda = 0) |\phi_{p\mathbf{k}}^{(0)}\rangle = \begin{cases} \frac{|\phi_{p\mathbf{k}}^{(0)}\rangle}{(1 - l_{p\mathbf{k}})^2} & : N_{P,\mathbf{k}} < p \leq N_{F,\mathbf{k}} \\ |\phi_{p\mathbf{k}}^{(0)}\rangle & : N_{F,\mathbf{k}} < p \leq N_W \\ 0 & : N_W < p \leq N_{D,\mathbf{k}} \end{cases} \quad (\text{C4})$$

The total number of states with nonzero unperturbed eigenvalues of $\hat{P}_{\text{smooth}}$ is $N_{\mathbf{k}}N_W - N_P$, and the perturbed version of these states will be selected for Wannierization. Since the mixing between the selected states is not important because our goal here is to determine the subspace spanned by them, we only consider the mixing between the eigenstates of $\hat{P}_{\text{smooth}}(\lambda = 0)$ with nonzero eigenvalues and those with zero eigenvalues.

Using first-order perturbation theory in a similar manner to Eq. (A6), the perturbed eigenstates of $\hat{P}_{\text{smooth}}$ are

$$|\phi_{p\mathbf{k}}\rangle - |\phi_{p\mathbf{k}}^{(0)}\rangle = \begin{cases} -\lambda(1 - l_{p\mathbf{k}}) \sum_m (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) & \text{for } p = N_{P,\mathbf{k}} + 1, \dots, N_{F,\mathbf{k}} \\ -\lambda \sum_m (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) & \text{for } p = N_{F,\mathbf{k}} + 1, \dots, N_W \end{cases} \quad (\text{C5})$$

The only difference with Eq. (A6) is the $(1 - l_{p\mathbf{k}})$ factor for $p = N_{P,\mathbf{k}} + 1, \dots, N_{F,\mathbf{k}}$.

Finally, we include the states inside $\mathcal{W}_P$ ($|\phi_{p\mathbf{k}}^{(0)}\rangle$ for $p = 1, \dots, N_{P,\mathbf{k}}$) and orthogonalize them to $\mathcal{O}(\lambda)$ as done in Eq. (A9). With Eq. (A3), the resulting basis states for the disentangled subspace are

$$|\tilde{\phi}_{p\mathbf{k}}\rangle = \begin{cases} |\psi_{p\mathbf{k}}^{(0)}\rangle + \lambda(\hat{1} - \hat{P}_W^{(0)})|\psi_{p\mathbf{k}}^{(1)}\rangle + \mathcal{O}(\lambda^2) & \text{for } p = 1, \dots, N_{P,\mathbf{k}} \\ |\psi_{p\mathbf{k}}^{(0)}\rangle + \lambda(\hat{1} - \hat{P}_D^{(0)} + l_{p\mathbf{k}}(\hat{P}_D^{(0)} - \hat{P}_W^{(0)}))|\psi_{p\mathbf{k}}^{(1)}\rangle + \mathcal{O}(\lambda^2) & \text{for } p = N_{P,\mathbf{k}} + 1, \dots, N_{F,\mathbf{k}} \\ \sum_m [|\psi_{m\mathbf{k}}^{(0)}\rangle + \lambda(\hat{1} - \hat{P}_D^{(0)})|\psi_{m\mathbf{k}}^{(1)}\rangle] \tilde{U}_{mp;\mathbf{k}}^{(0)} + \mathcal{O}(\lambda^2) & \text{for } p = N_{F,\mathbf{k}} + 1, \dots, N_W. \end{cases} \quad (\text{C6})$$

Note the additional $l_{p\mathbf{k}}(\hat{P}_D^{(0)} - \hat{P}_W^{(0)})$ factor for the $p = N_{P,\mathbf{k}} + 1, \dots, N_{F,\mathbf{k}}$ case, which is absent in Eq. (A9).

By projecting and orthogonalizing the unperturbed WFs to the disentangled subspace, the WFPs become

$$|w_{i\mathbf{R}}^{(1)}\rangle = \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} \hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} + \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ + \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} l_{m\mathbf{k}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) |\psi_{m\mathbf{k}}^{(1)}\rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (\text{C7})$$

The only difference between Eq. (C7) and the original formula for WFPs [Eq. (8)] is that Eq. (C7) has the third term. The factor $l_{m\mathbf{k}}$ in this term smoothly changes from 1 to 0 as $\varepsilon_{m\mathbf{k}}^{(0)}$ moves from the boundary energy of $\mathcal{W}_P$ to that of $\mathcal{W}_F$. Hence, this third term eliminates the discontinuity generated by the second term of Eq. (C7) while maintaining the desired property that the WFPs should accurately represent the wave function perturbation of the states inside $\mathcal{W}_P$. Therefore, Eq. (C7) is an alternative way to create WFPs and is expected to give more localized WFPs than Eq. (8). We detail how the real-space matrix elements can be calculated with this modified definition of WFPs in Appendix F 2.

For the systems studied in this work, we find that the WFPs constructed without smoothing the inner window are already sufficiently localized. Smoothing the inner window will be useful in systems where the WFP delocalization due to the discontinuity at the boundary of the inner window is a problem.

APPENDIX D: WFP CORRECTION TO THE HAMILTONIAN MATRIX ELEMENT

In this appendix, we prove Eq. (25).

$$H_{i'\mathbf{k}',i\mathbf{k}}^{(W1)} = \frac{1}{N_k} \sum_{\mathbf{R},\mathbf{R}',\mathbf{q}} [\langle w_{i'\mathbf{R}'}^{(0)} | \hat{V}_{\mathbf{q}} | w_{i\mathbf{R}}^{(0)} \rangle + \langle w_{i'\mathbf{R}'}^{(0)} | \hat{H}^{(0)} | w_{i\mathbf{R}+\mathbf{q}}^{(1)} \rangle + \langle w_{i'\mathbf{R}';-\mathbf{q}}^{(1)} | \hat{H}^{(0)} | w_{i\mathbf{R}}^{(0)} \rangle] e^{-i\mathbf{k}'\cdot\mathbf{R}'} e^{i\mathbf{k}\cdot\mathbf{R}} \\ = \frac{1}{N_k} \sum_{\mathbf{R},\mathbf{R}',\mathbf{q}} [\langle w_{i'\mathbf{R}'}^{(0)} | \hat{V}_{\mathbf{q}} | w_{i\mathbf{R}+\mathbf{R}'}^{(0)} \rangle + \langle w_{i'\mathbf{R}'}^{(0)} | \hat{H}^{(0)} | w_{i\mathbf{R}+\mathbf{R}';\mathbf{q}}^{(1)} \rangle + \langle w_{i'\mathbf{R}';-\mathbf{q}}^{(1)} | \hat{H}^{(0)} | w_{i\mathbf{R}+\mathbf{R}'}^{(0)} \rangle] e^{-i\mathbf{k}'\cdot\mathbf{R}'} e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{R}')} \\ = \frac{1}{N_k} \sum_{\mathbf{R},\mathbf{R}',\mathbf{q}} (g_{i'\mathbf{R};\mathbf{q}} + \delta g_{i'\mathbf{R};\mathbf{q}}) e^{i\mathbf{q}\cdot\mathbf{R}'} e^{-i\mathbf{k}'\cdot\mathbf{R}'} e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{R}')} \\ = \sum_{\mathbf{R},\mathbf{q}} (g_{i'\mathbf{R};\mathbf{q}} + \delta g_{i'\mathbf{R};\mathbf{q}}) e^{i\mathbf{k}\cdot\mathbf{R}} \delta_{\mathbf{k}+\mathbf{q},\mathbf{k}'} \\ = \tilde{g}_{i'\mathbf{k};\mathbf{k}'-\mathbf{k}}^{(W)}. \quad (\text{D1})$$

APPENDIX E: INTERPOLATION OF MATRIX ELEMENTS FOR KUBO FORMULA

In this section, we discuss the interpolation of the $K_{m\mathbf{n}\mathbf{k}}$ matrix element defined in Eq. (37). Wannier interpolation of $K_{m\mathbf{n}\mathbf{k}}$ can be done analogously to that of $S_{m\mathbf{n}\mathbf{k}}$ in Eqs. (33)–(36). For $n\mathbf{k}, m\mathbf{k} \in \mathcal{W}_P$, one can use the Wannier interpolation of the wave function perturbation [Eq. (30)] to find

$$\begin{aligned}
K_{mn\mathbf{k}} &= \langle \psi_{m\mathbf{k};\mathbf{q}}^{(\text{H1})} | \psi_{n\mathbf{k};\mathbf{q}}^{(\text{H1})} \rangle \\
&= \frac{1}{N_k} \sum_{i,i',\mathbf{R},\mathbf{R}'} e^{i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')} V_{m'i';\mathbf{k}}^{\dagger(0)} \langle w_{i'\mathbf{R}';\mathbf{q}}^{(1)} | w_{i\mathbf{R};\mathbf{q}}^{(1)} \rangle V_{in;\mathbf{k}}^{(0)} + \sum_{p=1}^{N_W} \frac{(\tilde{g}_{p\mathbf{k};\mathbf{q}}^{(\text{H})})^* \tilde{g}_{p\mathbf{k};\mathbf{q}}^{(\text{H})}}{(\epsilon_{m\mathbf{k}}^{(\text{H0})} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(\text{H0})})(\epsilon_{n\mathbf{k}}^{(\text{H0})} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(\text{H0})})} \\
&= \sum_{i,i',\mathbf{R}} V_{m'i';\mathbf{k}}^{\dagger(0)} k_{i'\mathbf{R};\mathbf{q}} V_{in;\mathbf{k}}^{(0)} e^{i\mathbf{k} \cdot \mathbf{R}} + \sum_{p=1}^{N_W} \frac{(\tilde{g}_{p\mathbf{k};\mathbf{q}}^{(\text{H})})^* \tilde{g}_{p\mathbf{k};\mathbf{q}}^{(\text{H})}}{(\epsilon_{m\mathbf{k}}^{(\text{H0})} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(\text{H0})})(\epsilon_{n\mathbf{k}}^{(\text{H0})} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(\text{H0})})}. \tag{E1}
\end{aligned}$$

In the third equality, we used the discrete translation properties of the WFP [Eq. (16)] and renamed $\mathbf{R} - \mathbf{R}'$ to $\mathbf{R}$. We also used the real-space matrix element defined in Eq. (38).

Using Eqs. (41) and (40), one can rewrite Eq. (E1) as Eq. (39). Since all quantities appearing in Eq. (39) can be calculated using a WF- or WFP-based interpolation, $K_{mn\mathbf{k}}$ can also be interpolated.

APPENDIX F: CALCULATION OF REAL-SPACE MATRIX ELEMENTS

1. Without the smoothed inner window

In this appendix, we describe how to calculate the real-space matrix elements using the Bloch eigenstates computed at a coarse k grid. For the potential matrix element $g_{\mathbf{R};\mathbf{q}}$, Ref. [8] gives

$$g_{ij\mathbf{R};\mathbf{q}} = \frac{1}{N_k} \sum_{m,n,\mathbf{k}} U_{im;\mathbf{k}+\mathbf{q}}^{(0)\dagger} \langle \psi_{m\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}, \tag{F1}$$

which easily follows from the definition of $g_{\mathbf{R};\mathbf{q}}$ [Eq. (17)]. Here, we derive analogous expressions for the newly introduced matrices $\delta g_{\mathbf{R};\mathbf{q}}$ and $s_{\mathbf{R};\mathbf{q}}$.

First, for $\delta g_{\mathbf{R};\mathbf{q}}$, we start from its definition [Eq. (18)] and use Eqs. (1) and (8) to find

$$\begin{aligned}
\delta g_{ij\mathbf{R};\mathbf{q}} &= \frac{1}{N_k} \sum_{\substack{m,n,p,\mathbf{k} \\ n\mathbf{k} \in \mathcal{W}_p}} U_{im;\mathbf{k}+\mathbf{q}}^{(0)\dagger} \frac{\epsilon_{m\mathbf{k}+\mathbf{q}}^{(0)} (P_D^{(0)} - P_W^{(0)})_{mp;\mathbf{k}+\mathbf{q}} \langle \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\epsilon_{n\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} \\
&\quad + \frac{1}{N_k} \sum_{\substack{m,n,p,\mathbf{k} \\ m\mathbf{k}+\mathbf{q} \in \mathcal{W}_p}} U_{im;\mathbf{k}+\mathbf{q}}^{(0)\dagger} \frac{\langle \psi_{m\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{p\mathbf{k}}^{(0)} \rangle (P_D^{(0)} - P_W^{(0)})_{pn;\mathbf{k}} \epsilon_{n\mathbf{k}}^{(0)}}{\epsilon_{m\mathbf{k}+\mathbf{q}}^{(0)} - \epsilon_{p\mathbf{k}}^{(0)}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}. \tag{F2}
\end{aligned}$$

In Eq. (F2), all band indices are limited to the bands inside $\mathcal{W}_D$. The $\hat{Q}_D^{(0)} |\psi_{m\mathbf{k}}^{(1)}\rangle$ term in Eq. (8) does not appear because the WFs are orthogonal to the $\hat{Q}_D^{(0)}$ subspace. Thus, $\delta g_{ij\mathbf{R};\mathbf{q}}$ can be evaluated using only the energies and matrix elements for the states inside $\mathcal{W}_D$.

For $s_{\mathbf{R};\mathbf{q}}$, using the definitions [Eqs. (1), (8), and (33)], we find

$$\begin{aligned}
s_{ij\mathbf{R};\mathbf{q}} &= \frac{1}{N_k} \sum_{m,n,\mathbf{k}} U_{im;\mathbf{k}}^{(0)\dagger} \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \hat{Q}_D^{(0)} \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}} \\
&\quad + \frac{1}{N_k} \sum_{\substack{m,n,p,q,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_p}} U_{im;\mathbf{k}}^{(0)\dagger} \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle}{\epsilon_{m\mathbf{k}}^{(0)} - \epsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} (P_D^{(0)} - P_W^{(0)})_{pq;\mathbf{k}+\mathbf{q}} \langle \psi_{q\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}'_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k} \cdot \mathbf{R}}. \tag{F3}
\end{aligned}$$

The first term involves the wave function perturbation, which can be calculated by solving the Sternheimer equation. The second term refers only to the bands inside the outer window.

For $k_{ij\mathbf{R};\mathbf{q}}$, we similarly find

$$k_{ij\mathbf{R};\mathbf{q}} = \sum_{m,n,\mathbf{k}} U_{im;\mathbf{k}}^{(0)\dagger} \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \hat{Q}_D^{(0)} | \psi_{n\mathbf{k};\mathbf{q}}^{(1)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ + \frac{1}{N_k} \sum_{\substack{m,n,p,q,\mathbf{k} \\ n\mathbf{k}, m\mathbf{k} \in \mathcal{W}_F}} U_{im;\mathbf{k}}^{(0)\dagger} \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}_{\mathbf{q}}^\dagger | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} (P_D^{(0)} - P_W^{(0)})_{pq;\mathbf{k}+\mathbf{q}} \frac{\langle \psi_{q\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}}' | \psi_{n\mathbf{k}}^{(0)} \rangle}{\varepsilon_{n\mathbf{k}}^{(0)} - \varepsilon_{q\mathbf{k}+\mathbf{q}}^{(0)}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (\text{F4})$$

2. With the smoothed inner window

In this subsection, we detail how the formulas for the real-space matrix elements change if the WFPs are constructed with the smoothed inner window as described in Appendix C. For brevity, we define

$$\tilde{l}_{m\mathbf{k}} = \begin{cases} 1 & : m\mathbf{k} \in \mathcal{W}_F \\ l_{m\mathbf{k}} & : m\mathbf{k} \in \mathcal{W}_F - \mathcal{W}_F \end{cases} \quad (\text{F5})$$

and rewrite Eq. (C7) as

$$|w_{i\mathbf{R}}^{(1)}\rangle = \frac{1}{\sqrt{N_k}} \sum_{m,\mathbf{k}} \hat{Q}_D^{(0)} | \psi_{m\mathbf{k}}^{(1)} \rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} + \frac{1}{\sqrt{N_k}} \sum_{\substack{m,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_F}} \tilde{l}_{m\mathbf{k}} (\hat{P}_D^{(0)} - \hat{P}_W^{(0)}) | \psi_{m\mathbf{k}}^{(1)} \rangle U_{mi;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (\text{F6})$$

Then, one finds that the real-space matrix elements are

$$\delta g_{ij\mathbf{R};\mathbf{q}} = \frac{1}{N_k} \sum_{\substack{m,n,p,\mathbf{k} \\ n\mathbf{k} \in \mathcal{W}_F}} U_{im;\mathbf{k}+\mathbf{q}}^{(0)\dagger} \frac{\varepsilon_{m\mathbf{k}+\mathbf{q}}^{(0)} (P_D^{(0)} - P_W^{(0)})_{mp;\mathbf{k}+\mathbf{q}} \langle \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle}{\varepsilon_{n\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} \tilde{l}_{n\mathbf{k}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ + \frac{1}{N_k} \sum_{\substack{m,n,p,\mathbf{k} \\ m\mathbf{k}+\mathbf{q} \in \mathcal{W}_F}} U_{im;\mathbf{k}+\mathbf{q}}^{(0)\dagger} \tilde{l}_{m\mathbf{k}+\mathbf{q}} \frac{\langle \psi_{m\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}} | \psi_{p\mathbf{k}}^{(0)} \rangle (P_D^{(0)} - P_W^{(0)})_{pm;\mathbf{k}} \varepsilon_{n\mathbf{k}}^{(0)}}{\varepsilon_{m\mathbf{k}+\mathbf{q}}^{(0)} - \varepsilon_{p\mathbf{k}}^{(0)}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}, \quad (\text{F7})$$

$$s_{ij\mathbf{R};\mathbf{q}} = \frac{1}{N_k} \sum_{m,n,\mathbf{k}} U_{im;\mathbf{k}}^{(0)\dagger} \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \hat{Q}_D^{(0)} \hat{V}_{\mathbf{q}} | \psi_{n\mathbf{k}}^{(0)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ + \frac{1}{N_k} \sum'_{\substack{m,n,p,q,\mathbf{k} \\ m\mathbf{k} \in \mathcal{W}_F}} U_{im;\mathbf{k}}^{(0)\dagger} \tilde{l}_{m\mathbf{k}} \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}_{\mathbf{q}}^\dagger | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} (P_D^{(0)} - P_W^{(0)})_{pq;\mathbf{k}+\mathbf{q}} \langle \psi_{q\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}}' | \psi_{n\mathbf{k}}^{(0)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}, \quad (\text{F8})$$

and

$$k_{ij\mathbf{R};\mathbf{q}} = \sum_{m,n,\mathbf{k}} U_{im;\mathbf{k}}^{(0)\dagger} \langle \psi_{m\mathbf{k};\mathbf{q}}^{(1)} | \hat{Q}_D^{(0)} | \psi_{n\mathbf{k};\mathbf{q}}^{(1)} \rangle U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ + \frac{1}{N_k} \sum_{\substack{m,n,p,q,\mathbf{k} \\ n\mathbf{k}, m\mathbf{k} \in \mathcal{W}_F}} U_{im;\mathbf{k}}^{(0)\dagger} \tilde{l}_{m\mathbf{k}} \frac{\langle \psi_{m\mathbf{k}}^{(0)} | \hat{V}_{\mathbf{q}}^\dagger | \psi_{p\mathbf{k}+\mathbf{q}}^{(0)} \rangle}{\varepsilon_{m\mathbf{k}}^{(0)} - \varepsilon_{p\mathbf{k}+\mathbf{q}}^{(0)}} (P_D^{(0)} - P_W^{(0)})_{pq;\mathbf{k}+\mathbf{q}} \frac{\langle \psi_{q\mathbf{k}+\mathbf{q}}^{(0)} | \hat{V}_{\mathbf{q}}' | \psi_{n\mathbf{k}}^{(0)} \rangle}{\varepsilon_{n\mathbf{k}}^{(0)} - \varepsilon_{q\mathbf{k}+\mathbf{q}}^{(0)}} \tilde{l}_{n\mathbf{k}} U_{nj;\mathbf{k}}^{(0)} e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (\text{F9})$$

When $l_{p\mathbf{k}} = 0$, which means that the inner window is not smoothed at all, Eq. (F7), Eq. (F8), and Eq. (F9) reduce to Eq. (F2), Eq. (F3), and Eq. (F4), respectively.

APPENDIX G: IMPROVED DOUBLE-GRID METHOD FOR PHONON-INDUCED ELECTRON SELF-ENERGY

In this section, we discuss the improved double-grid method for accelerating the convergence of the phonon-induced electron self-energy with respect to the coarse k grid size. Here, we focus on the diagonal part of the self-energy, but the results can also be applied to the off-diagonal one [35]. Our aim is to calculate $\Sigma_{n\mathbf{k}}$ for all $n\mathbf{k} \in \mathcal{W}_p$.

The convergence of the self-energy with respect to the q -point sampling is slow. This convergence rate is dominated by the convergence of the lower Fan self-energy [88]. The convergence of the upper Fan term with respect to the q -point sampling is much faster because the denominator $\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}}$ in Eq. (50) is never small. Therefore, one can save considerable computational cost by calculating the rapidly convergent upper Fan self-energy at a coarse q -point grid and the slowly convergent lower Fan self-energy at a finer q -point grid.

$$\Sigma_{n\mathbf{k}}(\eta) = \sum_{\mathbf{q}, \nu}^{\text{coarse}} [\Sigma_{n\mathbf{k}}^{\text{upper Fan}} + \Sigma_{n\mathbf{k}}^{\text{DW}}] + \sum_{\mathbf{q}, \nu}^{\text{fine}} [\Sigma_{n\mathbf{k}}^{\text{lower Fan}}(\eta)]. \quad (\text{G1})$$

This method is the double-grid method [39]. To distinguish this method from the improved version we propose later, we call it the “conventional double-grid method.” Note that one should also calculate the DW self-energy at the coarse grid when using the double-grid method. The reason for this is that the convergence of the sum of the DW and the upper Fan self-energies is much faster than the convergence of each term because of the approximate cancellation between the two terms [31,35,89].

However, we found that the conventional double-grid method is not very successful in the case of the finite-temperature self-energy of silicon. Figure 13 shows that the

sum of the upper Fan and DW self-energy (black dots) converges as $1/N_q$ if an $N_q \times N_q \times N_q$ q -point grid is used. When a $6 \times 6 \times 6$ coarse grid is used, the error in the band-gap renormalization at 400 K is 14 meV, around 13% of the converged value. Therefore, we need to speed up the convergence.

By investigating the contribution of each phonon mode to the self-energy, we find that the contribution of the low-energy acoustic phonons dominates the slow convergence rate as shown in Fig. 14(a). The slow $1/N_q$ convergence can be understood as follows. Let us focus on the acoustic modes at small $|\mathbf{q}|$ satisfying $\omega_{\mathbf{q}\nu} \approx v|\mathbf{q}|$. We set the momentum cutoff $q_{\text{cut}} \ll 2\pi/a$ for the low-energy modes where a is the lattice parameter. We assume that the temperature is high enough to satisfy $vq_{\text{cut}} \ll k_B T$ so that $2n_{\mathbf{q}\nu} + 1 \approx 2k_B T / \hbar \omega_{\mathbf{q}\nu}$. To make the q dependence explicit, we write $g_{m\mathbf{n}\mathbf{k}, \mathbf{q}\nu}$ in terms of $h_{m\mathbf{n}\mathbf{k}, \mathbf{q}\nu}$ using Eq. (B8) and approximate $h_{m\mathbf{n}\mathbf{k}, \mathbf{q}\nu}$ by $h_{m\mathbf{n}\mathbf{k}, 0\nu}$. Note that h , not g , should be approximated because g has an additional $|\mathbf{q}|$ dependence in the $1/\sqrt{\omega_{\mathbf{q}\nu}}$ factor in Eq. (B8). Then, the sum of Eq. (50) for $|\mathbf{q}| < q_{\text{cut}}$ becomes

$$\begin{aligned} \Sigma_{n\mathbf{k}}^{\text{upper Fan, acoustic}} &\approx \sum_{m \notin \mathcal{W}_p} \frac{|h_{m\mathbf{n}\mathbf{k}, 0\nu}|^2}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}}} \int_{|\mathbf{q}| < q_{\text{cut}}} d\mathbf{q} \frac{k_B T}{\hbar v^2 |\mathbf{q}|^2} \\ &= q_{\text{cut}} \frac{4\pi k_B T}{\hbar v^2} \sum_{m \notin \mathcal{W}_p} \frac{|h_{m\mathbf{n}\mathbf{k}, 0\nu}|^2}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}}}. \end{aligned} \quad (\text{G2})$$

The acoustic phonon contribution to the upper Fan self-energy is proportional to q_{cut} . Using a coarse q grid gives an effective cutoff $q_{\text{cut}} \sim 1/N_q$ because the acoustic phonon modes with $q < 1/N_q \times 2\pi/a$ are not sampled. Therefore, the conventional double-grid method converges as $1/N_q$. Note that the convergence is much faster at $T = 0$ because then $2n_{\mathbf{q}\nu} + 1 = 1$ does not give a $1/q$ factor, so $\Sigma_{n\mathbf{k}}^{\text{upper Fan, acoustic}}$ becomes proportional to $q_{\text{cut}}^2 \sim 1/N_q^2$.

FIG. 13. Phonon-induced electron self-energy of silicon at $T = 400$ K calculated using the conventional and improved double-grid (DG) methods for the (a) VBM, (b) CBM, and the lowest conduction band (CB) at Γ . The fine q grid size and η for the self-energy are fixed to $125 \times 125 \times 125$ and 100 meV, respectively.

FIG. 14. (a) Contribution of each phonon mode to the sum of the DW and upper Fan self-energies of silicon for the electron state at the valence band maximum. The size and color of the circles indicate the magnitude and sign of the phonon-mode-resolved electron self-energy, where red (blue) means a positive (negative) contribution. (b) Sum of the upper Fan (UF), DW, and optionally lower Fan (LF) self-energies for the lowest-frequency ($\nu = 1$) acoustic mode multiplied by $|\mathbf{q}|^2$ for q points along the Γ -L line.

To speed up the convergence, we utilize the fact that the acoustic modes around $\mathbf{q} = 0$ actually have little effect on the total self-energy because of the translational symmetry. Since the acoustic modes at $\mathbf{q} = 0$ describe a uniform translation of the whole lattice, the band eigenvalues should not change. However, the sum of the phonon-mode-resolved upper Fan term and the DW term is not small and even diverges at $\mathbf{q} = 0$. This divergence is canceled by the divergence of the lower Fan term in the opposite sign, so their sum, the total self-energy, is zero at $\mathbf{q} = 0$.

To make use of this cancellation, we rewrite the total self-energy as follows:

$$\Sigma_{n\mathbf{k}}(\eta) = \sum_{\mathbf{q},\nu}^{\text{coarse}} [\Sigma_{n\mathbf{k}}^{\text{upper Fan}} + \Sigma_{n\mathbf{k}}^{\text{DW}} + \Sigma_{n\mathbf{k}}^{\text{lower Fan}}(\eta_{\text{large}})] + \sum_{\mathbf{q},\nu}^{\text{fine}} [\Sigma_{n\mathbf{k}}^{\text{lower Fan}}(\eta) - \Sigma_{n\mathbf{k}}^{\text{lower Fan}}(\eta_{\text{large}})]. \quad (\text{G3})$$

Here, η is the small parameter that should be converged to 0, while η_{large} is a large value around 1 eV and does not need to be converged. Equation (G3) is the “improved double-grid method” we propose in this work. The key property of Eq. (G3) is that the first term converges much faster with respect to the q grid than the first term of Eq. (G1), which is just the sum of the upper Fan and DW self-energies. The reason for this is that the diverging contribution of the low-energy acoustic modes in the upper Fan and DW terms is suppressed by the additional lower Fan term. Figure 14(b) shows the self-energy contributions multiplied by the phase-space volume factor $|\mathbf{q}|^2$. We find that the contribution of modes with small $|\mathbf{q}|$ is indeed suppressed by adding the lower Fan term.

The cancellation becomes exact as η_{large} is decreased. However, we need to choose η_{large} to be large enough so that the denominator in the lower Fan term [Eq. (49)] is always far from zero and does not create additional convergence issues. For example, Fig. 13(c) shows a slow convergence of the self-energy for the lowest conduction band at Γ for $\eta_{\text{large}} = 0.5$ eV. The slow convergence is more pronounced for states far from band extrema because there are more resonant couplings mediated by phonons to other electron states, leading to a rapid variation in the phonon-mode-resolved lower Fan self-energy. Practically, we find that η_{large} in [1 eV, 2 eV] is a good choice for silicon.

Figure 13 shows that the total self-energy converges much faster with respect to the coarse q grid when using the improved double-grid method. By using the improved double-grid method with $\eta = 1$ eV, the absolute (relative) error in the band-gap renormalization at 400 K is only 2.4 meV (2%) with a $6 \times 6 \times 6$ coarse q grid, in contrast to 14 meV (13%) for the conventional double-grid method. Also, the choice of η gives only a minor effect on the convergence rate, and more importantly, the convergence is consistently faster than the conventional double-grid method for a wide range of η in [1 eV, 2 eV].

-
- [1] W. Kohn, *Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms*, *Phys. Rev. Lett.* **76**, 3168 (1996).
 - [2] E. Prodan and W. Kohn, *Nearsightedness of Electronic Matter*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11635 (2005).
 - [3] G.H. Wannier, *The Structure of Electronic Excitation Levels in Insulating Crystals*, *Phys. Rev.* **52**, 191 (1937).
 - [4] W. Kohn, *Analytic Properties of Bloch Waves and Wannier Functions*, *Phys. Rev.* **115**, 809 (1959).

- [5] N. Marzari and D. Vanderbilt, *Maximally-Localized Generalized Wannier Functions for Composite Energy Bands*, *Phys. Rev. B* **56**, 12847 (1997).
- [6] I. Souza, N. Marzari, and D. Vanderbilt, *Maximally Localized Wannier Functions for Entangled Energy Bands*, *Phys. Rev. B* **65**, 035109 (2001).
- [7] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, *Maximally Localized Wannier Functions: Theory and Applications*, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [8] F. Giustino, M. L. Cohen, and S. G. Louie, *Electron-Phonon Interaction Using Wannier Functions*, *Phys. Rev. B* **76**, 165108 (2007).
- [9] J.-J. Zhou and M. Bernardi, *Ab Initio Electron Mobility and Polar Phonon Scattering in GaAs*, *Phys. Rev. B* **94**, 201201(R) (2016).
- [10] S. Poncé, E. R. Margine, and F. Giustino, *Towards Predictive Many-Body Calculations of Phonon-Limited Carrier Mobilities in Semiconductors*, *Phys. Rev. B* **97**, 121201(R) (2018).
- [11] J. Noffsinger, E. Kioupakis, C. G. Van de Walle, S. G. Louie, and M. L. Cohen, *Phonon-Assisted Optical Absorption in Silicon from First Principles*, *Phys. Rev. Lett.* **108**, 167402 (2012).
- [12] E. R. Margine and F. Giustino, *Anisotropic Migdal-Eliashberg Theory Using Wannier Functions*, *Phys. Rev. B* **87**, 024505 (2013).
- [13] V. A. Jhalani, J.-J. Zhou, and M. Bernardi, *Ultrafast Hot Carrier Dynamics in GaN and Its Impact on the Efficiency Droop*, *Nano Lett.* **17**, 5012 (2017).
- [14] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, *Ab Initio Calculation of the Anomalous Hall Conductivity by Wannier Interpolation*, *Phys. Rev. B* **74**, 195118 (2006).
- [15] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, *Spectral and Fermi Surface Properties from Wannier Interpolation*, *Phys. Rev. B* **75**, 195121 (2007).
- [16] J. Ibañez-Azpiroz, S. S. Tsirkin, and I. Souza, *Ab Initio Calculation of the Shift Photocurrent by Wannier Interpolation*, *Phys. Rev. B* **97**, 245143 (2018).
- [17] S. S. Tsirkin, P. A. Puente, and I. Souza, *Gyrotropic Effects in Trigonal Tellurium Studied from First Principles*, *Phys. Rev. B* **97**, 035158 (2018).
- [18] J. Qiao, J. Zhou, Z. Yuan, and W. Zhao, *Calculation of Intrinsic Spin Hall Conductivity by Wannier Interpolation*, *Phys. Rev. B* **98**, 214402 (2018).
- [19] J. H. Ryoo, C.-H. Park, and I. Souza, *Computation of Intrinsic Spin Hall Conductivities from First Principles Using Maximally Localized Wannier Functions*, *Phys. Rev. B* **99**, 235113 (2019).
- [20] R. M. Sternheimer, *Electronic Polarizabilities of Ions from the Hartree-Fock Wave Functions*, *Phys. Rev.* **96**, 951 (1954).
- [21] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, *Phonons and Related Crystal Properties from Density-Functional Perturbation Theory*, *Rev. Mod. Phys.* **73**, 515 (2001).
- [22] X. Ge and D. Lu, *Local Representation of the Electronic Dielectric Response Function*, *Phys. Rev. B* **92**, 241107(R) (2015).
- [23] F. Giustino, S. G. Louie, and M. L. Cohen, *Electron-Phonon Renormalization of the Direct Band Gap of Diamond*, *Phys. Rev. Lett.* **105**, 265501 (2010).
- [24] X. Gonze, P. Boulanger, and M. Côté, *Theoretical Approaches to the Temperature and Zero-Point Motion Effects on the Electronic Band Structure*, *Ann. Phys. (N.Y.)* **523**, 168 (2011).
- [25] J. E. Sipe and A. I. Shkrebtii, *Second-Order Optical Response in Semiconductors*, *Phys. Rev. B* **61**, 5337 (2000).
- [26] J.-M. Lihm and C.-H. Park (to be published).
- [27] H. Xu, H. Wang, J. Zhou, and J. Li, *Pure Spin Photocurrent in Non-centrosymmetric Crystals: Bulk Spin Photovoltaic Effect*, *Nat. Commun.* **12**, 4330 (2021).
- [28] X. Mu, Y. Pan, and J. Zhou, *Pure Bulk Orbital and Spin Photocurrent in Two-Dimensional Ferroelectric Materials*, *npj Comput. Mater.* **7**, 61 (2021).
- [29] L. He and D. Vanderbilt, *Exponential Decay Properties of Wannier Functions and Related Quantities*, *Phys. Rev. Lett.* **86**, 5341 (2001).
- [30] H. D. Cornean, D. Gontier, A. Levitt, and D. Monaco, *Localised Wannier Functions in Metallic Systems*, in *Ann. Henri Poincaré* **20**, 1367 (2019).
- [31] P. B. Allen and V. Heine, *Theory of the Temperature Dependence of Electronic Band Structures*, *J. Phys. C* **9**, 2305 (1976).
- [32] P. B. Allen and M. Cardona, *Theory of the Temperature Dependence of the Direct Gap of Germanium*, *Phys. Rev. B* **23**, 1495 (1981).
- [33] P. B. Allen and M. Cardona, *Temperature Dependence of the Direct Gap of Si and Ge*, *Phys. Rev. B* **27**, 4760 (1983).
- [34] F. Giustino, *Electron-Phonon Interactions from First Principles*, *Rev. Mod. Phys.* **89**, 015003 (2017).
- [35] J.-M. Lihm and C.-H. Park, *Phonon-Induced Renormalization of Electron Wave Functions*, *Phys. Rev. B* **101**, 121102(R) (2020).
- [36] S. Poncé, G. Antonius, Y. Gillet, P. Boulanger, J. L. Janssen, A. Marini, M. Côté, and X. Gonze, *Temperature Dependence of Electronic Eigenenergies in the Adiabatic Harmonic Approximation*, *Phys. Rev. B* **90**, 214304 (2014).
- [37] J. P. Nery, P. B. Allen, G. Antonius, L. Reining, A. Miglio, and X. Gonze, *Quasiparticles and Phonon Satellites in Spectral Functions of Semiconductors and Insulators: Cumulants Applied to the Full First-Principles Theory and the Fröhlich Polaron*, *Phys. Rev. B* **97**, 115145 (2018).
- [38] A. Miglio, V. Brousseau-Couture, E. Godbout, G. Antonius, Y.-H. Chan, S. G. Louie, M. Côté, M. Giantomassi, and X. Gonze, *Predominance of Non-adiabatic Effects in Zero-Point Renormalization of the Electronic Band Gap*, *npj Comput. Mater.* **6**, 167 (2020).
- [39] F. Brown-Altvater, G. Antonius, T. Rangel, M. Giantomassi, C. Draxl, X. Gonze, S. G. Louie, and J. B. Neaton, *Band Gap Renormalization, Carrier Mobilities, and the Electron-Phonon Self-Energy in Crystalline Naphthalene*, *Phys. Rev. B* **101**, 165102 (2020).
- [40] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts, *Fine Structure in the Absorption-Edge Spectrum of Si*, *Phys. Rev.* **111**, 1245 (1958).

- [41] E. Kioupakis, P. Rinke, A. Schleife, F. Bechstedt, and C. G. Van de Walle, *Free-Carrier Absorption in Nitrides from First Principles*, *Phys. Rev. B* **81**, 241201(R) (2010).
- [42] C. E. Patrick and F. Giustino, *Unified Theory of Electron-Phonon Renormalization and Phonon-Assisted Optical Absorption*, *J. Phys. Condens. Matter* **26**, 365503 (2014).
- [43] M. S. Hybertsen and S. G. Louie, *Electron Correlation in Semiconductors and Insulators: Band Gaps and Quasiparticle Energies*, *Phys. Rev. B* **34**, 5390 (1986).
- [44] M. Zacharias, C. E. Patrick, and F. Giustino, *Stochastic Approach to Phonon-Assisted Optical Absorption*, *Phys. Rev. Lett.* **115**, 177401 (2015).
- [45] M. Zacharias and F. Giustino, *One-Shot Calculation of Temperature-Dependent Optical Spectra and Phonon-Induced Band-Gap Renormalization*, *Phys. Rev. B* **94**, 075125 (2016).
- [46] B. Monserrat and D. Vanderbilt, *Temperature Effects in the Band Structure of Topological Insulators*, *Phys. Rev. Lett.* **117**, 226801 (2016).
- [47] V. Gorelov, D. M. Ceperley, M. Holzmann, and C. Pierleoni, *Electronic Structure and Optical Properties of Quantum Crystals from First Principles Calculations in the Born-Oppenheimer Approximation*, *J. Chem. Phys.* **153**, 234117 (2020).
- [48] M. Zacharias, M. Scheffler, and C. Carbogno, *Fully Anharmonic Nonperturbative Theory of Vibronically Renormalized Electronic Band Structures*, *Phys. Rev. B* **102**, 045126 (2020).
- [49] T. Rangel, M. Del Ben, D. Varsano, G. Antonius, F. Bruneval, F. H. da Jornada, M. J. van Setten, O. K. Orhan, D. D. O'Regan, A. Canning, A. Ferretti, A. Marini, G.-M. Rignanese, J. Deslippe, S. G. Louie, and J. B. Neaton, *Reproducibility in G_0W_0 Calculations for Solids*, *Comput. Phys. Commun.* **255**, 107242 (2020).
- [50] G. Antonius, S. Poncé, P. Boulanger, M. Côté, and X. Gonze, *Many-Body Effects on the Zero-Point Renormalization of the Band Structure*, *Phys. Rev. Lett.* **112**, 215501 (2014).
- [51] W. Bludau, A. Onton, and W. Heinke, *Temperature Dependence of the Band Gap of Silicon*, *J. Appl. Phys.* **45**, 1846 (1974).
- [52] D. R. Hamann and D. Vanderbilt, *Maximally Localized Wannier Functions for GW Quasiparticles*, *Phys. Rev. B* **79**, 045109 (2009).
- [53] See Supplemental Material, which includes Refs. [54–56], at <http://link.aps.org/supplemental/10.1103/PhysRevX.11.041053> for the details of the calculation of the velocity and related matrix elements and the convergence study.
- [54] S. Poncé, F. Macheda, E. R. Margine, N. Marzari, N. Bonini, and F. Giustino, *First-Principles Predictions of Hall and Drift Mobilities in Semiconductors*, *arXiv:2105.04192*.
- [55] B. Adolph, V. I. Gavrilenko, K. Tenelsen, F. Bechstedt, and R. Del Sole, *Nonlocality and Many-Body Effects in the Optical Properties of Semiconductors*, *Phys. Rev. B* **53**, 9797 (1996).
- [56] M. G. Lopez, D. Vanderbilt, T. Thonhauser, and I. Souza, *Wannier-Based Calculation of the Orbital Magnetization in Crystals*, *Phys. Rev. B* **85**, 014435 (2012).
- [57] H. H. Huang, X. Fan, D. J. Singh, H. Chen, Q. Jiang, and W. T. Zheng, *Controlling Phase Transition for Single-Layer MTe_2 ($\text{M} = \text{Mo}$ and W): Modulation of the Potential Barrier under Strain*, *Phys. Chem. Chem. Phys.* **18**, 4086 (2016).
- [58] S. Baroni and R. Resta, *Ab Initio Calculation of the Macroscopic Dielectric Constant in Silicon*, *Phys. Rev. B* **33**, 7017 (1986).
- [59] G. Y. Guo, S. Murakami, T.-W. Chen, and N. Nagaosa, *Intrinsic Spin Hall Effect in Platinum: First-Principles Calculations*, *Phys. Rev. Lett.* **100**, 096401 (2008).
- [60] G. Y. Guo, Y. Yao, and Q. Niu, *Ab Initio Calculation of the Intrinsic Spin Hall Effect in Semiconductors*, *Phys. Rev. Lett.* **94**, 226601 (2005).
- [61] M. Schlipf, S. Poncé, and F. Giustino, *Carrier Lifetimes and Polaronic Mass Enhancement in the Hybrid Halide Perovskite $\text{CH}_3\text{NH}_3\text{PbI}_3$ from Multiphonon Fröhlich Coupling*, *Phys. Rev. Lett.* **121**, 086402 (2018).
- [62] R. D'Souza, J. Cao, J. D. Querales-Flores, S. Fahy, and I. Savić, *Electron-Phonon Scattering and Thermoelectric Transport in p-Type PbTe from First Principles*, *Phys. Rev. B* **102**, 115204 (2020).
- [63] V. Vitale, G. Pizzi, A. Marrazzo, J. R. Yates, N. Marzari, and A. A. Mostofi, *Automated High-Throughput Wannierisation*, *npj Comput. Mater.* **6**, 66 (2020).
- [64] O. Bistoni, P. Barone, E. Cappelluti, L. Benfatto, and F. Mauri, *Giant Effective Charges and Piezoelectricity in Gapped Graphene*, *2D Mater.* **6**, 045015 (2019).
- [65] L. Binci, P. Barone, and F. Mauri, *First-Principles Theory of Infrared Vibrational Spectroscopy of Metals and Semimetals: Application to Graphite*, *Phys. Rev. B* **103**, 134304 (2021).
- [66] C. E. Dreyer, S. Coh, and M. Stengel, *Nonadiabatic Born effective charges in metals and the Drude weight*, *arXiv:2103.04425*.
- [67] A. Eiguren and C. Ambrosch-Draxl, *Wannier Interpolation Scheme for Phonon-Induced Potentials: Application to Bulk MgB_2 , W , and the (1×1) H -Covered $\text{W}(110)$ Surface*, *Phys. Rev. B* **78**, 045124 (2008).
- [68] F. Giustino, M. L. Cohen, and S. G. Louie, *GW Method with the Self-Consistent Sternheimer Equation*, *Phys. Rev. B* **81**, 115105 (2010).
- [69] P. Umari, G. Stenuit, and S. Baroni, *GW Quasiparticle Spectra from Occupied States Only*, *Phys. Rev. B* **81**, 115104 (2010).
- [70] P. Giannozzi *et al.*, *QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials*, *J. Phys. Condens. Matter* **21**, 395502 (2009).
- [71] P. Giannozzi *et al.*, *Advanced Capabilities for Materials Modelling with Quantum ESPRESSO*, *J. Phys. Condens. Matter* **29**, 465901 (2017).
- [72] D. R. Hamann, *Optimized Norm-Conserving Vanderbilt Pseudopotentials*, *Phys. Rev. B* **88**, 085117 (2013).
- [73] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, *The PseudoDojo: Training and Grading a 85 Element Optimized Norm-Conserving Pseudopotential Table*, *Comput. Phys. Commun.* **226**, 39 (2018).
- [74] J. P. Perdew and Y. Wang, *Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy*, *Phys. Rev. B* **45**, 13244 (1992).

- [75] R. R. Reeber and K. Wang, *Thermal Expansion and Lattice Parameters of Group IV Semiconductors*, *Mater. Chem. Phys.* **46**, 259 (1996).
- [76] P. Lautenschlager, P. B. Allen, and M. Cardona, *Temperature Dependence of Band Gaps in Si and Ge*, *Phys. Rev. B* **31**, 2163 (1985).
- [77] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [78] G. Pizzi *et al.*, *Wannier90 as a Community Code: New Features and Applications*, *J. Phys. Condens. Matter* **32**, 165902 (2020).
- [79] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, *BerkeleyGW: A Massively Parallel Computer Package for the Calculation of the Quasiparticle and Optical Properties of Materials and Nanostructures*, *Comput. Phys. Commun.* **183**, 1269 (2012).
- [80] J. Deslippe, G. Samsonidze, M. Jain, M. L. Cohen, and S. G. Louie, *Coulomb-Hole Summations and Energies for GW Calculations with Limited Number of Empty Orbitals: A Modified Static Remainder Approach*, *Phys. Rev. B* **87**, 165124 (2013).
- [81] S. Ponc, E. R. Margine, C. Verdi, and F. Giustino, *EPW: Electron-Phonon Coupling, Transport and Superconducting Properties Using Maximally Localized Wannier Functions*, *Comput. Phys. Commun.* **209**, 116 (2016).
- [82] M. Rohlfing and S. G. Louie, *Electron-Hole Excitations and Optical Spectra from First Principles*, *Phys. Rev. B* **62**, 4927 (2000).
- [83] M. A. Green and M. J. Keevers, *Optical Properties of Intrinsic Silicon at 300 K*, *Prog. Photovoltaics* **3**, 189 (1995).
- [84] F. Nastos and J. E. Sipe, *Optical Rectification and Shift Currents in GaAs and GaP Response: Below and Above the Band Gap*, *Phys. Rev. B* **74**, 035201 (2006).
- [85] S. S. Tsirkin, *High Performance Wannier Interpolation of Berry Curvature and Related Quantities with WannierBerri Code*, *npj Comput. Mater.* **7**, 33 (2021).
- [86] C.-H. Lee, E. C. Silva, L. Calderin, M. A. T. Nguyen, M. J. Hollander, B. Bersch, T. E. Mallouk, and J. A. Robinson, *Tungsten Ditelluride: A Layered Semimetal*, *Sci. Rep.* **5**, 10013 (2015).
- [87] K. Momma and F. Izumi, *VESTA3 for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data*, *J. Appl. Crystallogr.* **44**, 1272 (2011).
- [88] S. Ponc, Y. Gillet, J. L. Janssen, A. Marini, M. Verstraete, and X. Gonze, *Temperature Dependence of the Electronic Structure of Semiconductors and Insulators*, *J. Chem. Phys.* **143**, 102813 (2015).
- [89] J. D. Querales-Flores, J. Cao, S. Fahy, and I. Savi, *Temperature Effects on the Electronic Band Structure of PbTe from First Principles*, *Phys. Rev. Mater.* **3**, 055405 (2019).