

# Hamiltonian Transformation for Band Structure Calculations

Kai Wu,<sup>1</sup> Yingzhou Li,<sup>2</sup> Wentiao Wu,<sup>1</sup> Lin Lin,<sup>3,4</sup> Wei Hu,<sup>1,\*</sup> and Jinlong Yang<sup>1,†</sup>

<sup>1</sup>*School of Future Technology, Department of Chemical Physics, and Anhui Center for Applied Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, China*

<sup>2</sup>*School of Mathematical Sciences, Fudan University, Shanghai 200433, China*

<sup>3</sup>*Department of Mathematics, University of California, Berkeley, California 94720, United States*

<sup>4</sup>*Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

First-principles electronic band structure calculations are essential for understanding periodic systems in condensed matter physics and materials science. We propose an accurate and parameter-free method, called Hamiltonian transformation (HT), to calculate band structures in both density functional theory (DFT) and post-DFT calculations with plane-wave basis sets. The cost of HT is independent of the choice of the density functional and scales as  $\mathcal{O}(N_e^3 N_k \log N_k)$ , where  $N_e$  and  $N_k$  are the number of electrons and the number of  $\mathbf{k}$ -points. Compared to the widely used Wannier interpolation (WI), HT adopts an eigenvalue transformation to construct a spatial localized representation of the spectrally truncated Hamiltonian. HT also uses a non-iterative algorithm to change the basis sets to circumvent the construction of the maximally localized Wannier functions. As a result, HT can significantly outperform WI in terms of the accuracy of the band structure calculation. We also find that the eigenvalue transformation can be of independent interest, and can be used to improve the accuracy of the WI for systems with entangled bands.

*Introduction* – Band structure plays a fundamental role in condensed matter physics and materials science. In the Kohn-Sham density functional theory (DFT) [1, 2], the band structure calculations with  $\mathbf{k}$ -point sampling consist of two steps: Self-consistent field (SCF) electronic structure calculations to obtain the eigenvalues and Kohn-Sham orbitals (eigenvectors) on a uniform  $\mathbf{k}$ -point grid  $\{\mathbf{k}\}$ , and non-SCF calculations to obtain the eigenvalues on an additional nonuniform  $\mathbf{k}$ -point grid (or path)  $\{\mathbf{q}\}$  for band structures. The accuracy and complexity of band structure calculations strongly depend on the choice of exchange-correlation functionals. For example, calculations with local or semilocal functionals, such as local density approximation (LDA) [3, 4] and generalized gradient approximation (GGA) [5–7], scale up to  $\mathcal{O}(N_e^3 N_k)$ , where  $N_e$  and  $N_k$  are the number of electrons and the number of  $\mathbf{k}$ -points, respectively. However, nonlocal functionals such as hybrid functionals [8–10] and post-DFT calculations such as the GW approximation [11–14], usually results in much larger computational cost ( $\mathcal{O}(N_e^4 N_k^2)$  [15] or higher). Though several algorithms have been proposed to reduce the cost of SCF calculations, such as interpolative separable density fitting (ISDF) [16–22], the resolution-of-the-identity (RI) [23–27], adaptively compressed exchange (ACE) operator [28, 29], the chain of spheres algorithm [30, 31] and tensor hypercontraction (THC) [32–34], the subsequent non-SCF band structure calculations remain expensive. This becomes a bottleneck for band structure calculations beyond local or semilocal functionals in condensed matter physics and materials science.

In addition to non-SCF calculations which constructs the Hamiltonian implicitly through the SCF orbitals, one can also construct the Hamiltonian  $H_{\mathbf{k}}$  explicitly and

Fourier interpolating  $H_{\mathbf{q}}$  from  $H_{\mathbf{k}}$ :

$$H_{\mathbf{q}} = \frac{1}{N_k} \sum_{\mathbf{k}, \mathbf{R}} H_{\mathbf{k}} e^{i(\mathbf{q}-\mathbf{k})\mathbf{R}}, \quad (1)$$

where  $\mathbf{R}$  is the Bravais lattice vector. The success of such an interpolation procedure relies on two assumptions: the size of the matrix is small, and matrix elements are smooth in reciprocal space or localized in real space. The Fourier interpolation can be directly performed using a small basis set. When large basis sets such as plane-waves are used, the Hamiltonian is often first projected to a subspace spanned by the maximally localized Wannier functions (MLWFs) [35–37], and the Fourier interpolation is then performed with respect to the projected Hamiltonian matrix. This procedure is called Wannier interpolation (WI).

The WI is a popular interpolation method in condensed matter physics and plays a key role in constructing model Hamiltonians [38, 39] and computing various physical observables of solids [40–42]. However, constructing MLWFs is a challenging nonlinear optimization problem due to the existence of multiple local minima [36], hence the results can be sensitive to initial guesses and users may need detailed knowledge of the system to provide the initial guess. Despite many progresses in improving the robustness of numerical algorithms for finding localized Wannier functions [43–48], constructing MLWFs can still be challenging in a number of scenarios, such as topological insulators [49, 50] and entangled band structures [46, 51, 52].

In this Letter, we have two main contributions. First, if we only keep the eigenpairs of the original Hamiltonian within a certain energy window, the resulting spectrally truncated Hamiltonian can be more delocal-

ized than original Hamiltonian matrix. We introduce an eigenvalue transformation to find a localized representation of this spectrally truncated Hamiltonian. We also find that the eigenvalue transformation introduced here can be of independent interest, and can be used to improve the accuracy of WI for systems with entangled bands.

Second, WI finds a minimal set of orthogonal orbitals that are smooth in  $\mathbf{k}$ , so that the projected Hamiltonian is spatially localized. However, for band structure calculations, we may entirely circumvent the construction of MLWFs, and obtain instead a suitably larger but much more localized Hamiltonian through changing basis sets, thus improving the accuracy and efficiency of band structure calculations. This strategy can be particularly effective for entangled and/or topologically obstructed band structures.

Our method is called Hamiltonian transformation (HT), which is a combination of eigenvalue transformation and changing basis sets. It has a functional-independent time complexity of  $\mathcal{O}(N_e^3 N_k \log(N_k))$ . As an accurate, efficient, and parameter-free method for band structure calculations, HT can be applied to both DFT and post-DFT calculations.

*Eigenvalue transformation* – In addition to the well known MLWFs which optimizes the basis functions, there is another direction to localize Hamiltonian: eigenvalues. An example of 1-D atomic chain could show the effect of eigenvalue transformation, whose Hamiltonian  $T$  is the tridiagonal Toeplitz matrix [53]. Fig. 1(a) is  $T$ , its main diagonal elements are 1, lower and upper diagonal elements are 0.5, other elements are 0. Atoms only interact with nearest neighbours, which means  $T$  is localized. Fig. 1(b) is the eigenvalue spectrum of  $T$ . The SCF calculation can only obtain the lowest few eigenvalues and corresponding eigenvectors, here we assume only eigenvalues smaller than 1.5 is obtained. Fig. 1(c) is the spectrally truncated Hamiltonian reconstructed from SCF eigenvalues and eigenvectors, and Fig. 1(d) is its eigenvalue spectrum. It is clear that the reconstructed  $T$  in Fig. 1(c) is no longer localized. In fact, the localized  $T$  has nonlocal eigenvectors, and the eigenvectors oscillate between positive and negative and cancel each other out away from diagonal. When some eigenvalues are truncated, the remaining corresponding eigenvectors cannot cancel each other out, therefore the reconstructed  $T$  is delocalized.

A key observation is that we can shift down the remaining eigenvalues by 1.5 to localize  $T$ , which is verified in Fig. 1(e) and (f). If we ignore elements whose magnitude smaller than 0.01, the reconstructed  $T$  can be taken as banded matrices with a band width of 20 in Fig. 1(c) and a band width of 3 in Fig. 1(e). Note that the eigenvalue spectrum is discontinuous before shifting and is continuous after shifting. We may further expect that a nonlinear transformation  $f$  which makes the eigen-

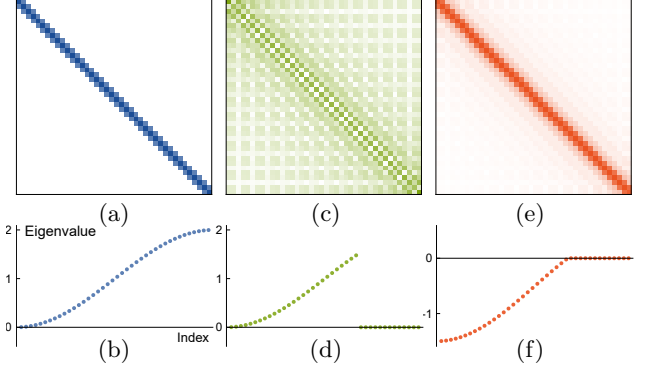


Fig. 1. Transforming eigenvalues of tridiagonal Toeplitz matrix could affect its localization. The matrices generated by different eigenvalue spectra show different localization properties.

value spectrum smoother could generate a more localized Hamiltonian.

To identify such a function  $f$ , the first step is define *decay radius*  $r$  to quantitatively describe the localization of the Hamiltonian with  $\mathbf{k}$ -point sampling

$$r = \sup \left\{ \|\mathbf{R}\|_2 : \max_{ij} |H_{\mathbf{R},ij}| \geq t \right\}, \quad (2)$$

where  $t > 0$  is a threshold. Furthermore, we need a functional  $F$  which gives the upper bound of  $H_{ij}$  after transformation:

$$F[f, |i-j|] = \sup \left\{ |f(H)_{ij}| : H^\dagger = H, \sigma(H) \in [-1, 1] \right\}, \quad (3)$$

where  $\sigma$  is eigenvalue spectrum. Analyzing localization property of a general Hermitian matrix is extremely difficult, but the task is easier for an infinite large system whose Hamiltonian is a sparse matrix without  $\mathbf{k}$ -point sampling. In this case, for the Hamiltonian discretized in real space, Eq. 2 can be simplified to

$$r = \sup \{ |i-j| : |H_{ij}| \geq t \}. \quad (4)$$

For a banded Hermitian matrix  $H$  with eigenvalues in the interval  $[-1, 1]$ , we have (Supplemental Material ??):

$$\begin{aligned} |f(H)_{ij}| &\leq \frac{2}{\pi} \left\| \sum_{l=k+1}^{\infty} \cos l\theta \int_0^\pi f(\cos t) \cos lt \, dt \right\|_{\theta \in [0, \pi]} \\ &= c F[f, k], \end{aligned} \quad (5)$$

where  $i, j$  satisfy  $mk < |i-j| \leq m(k+1)$ ,  $m$  is the bandwidth of  $H$ ,  $c$  is a factor normalizing  $F[f, 0]$  to 1.

With  $F$ , we are able to verify the conjecture. It can be shown that for sufficient large  $|i-j|$  and a  $n$ -differentiable function  $f$ ,  $|f(H)_{ij}|$  decays at least polynomially with a

rate  $\sim |i-j|^{-(n+1)}$  (Supplemental Material ??). The result does not require that  $f(H)$  actually decays polynomially away from the main diagonal, but only that  $f(H)$  is bounded above in a polynomially decaying manner.

Then we study the form of  $f$ . It will be shown later that optimizing  $f$  is a multi-objective problem, and it is hard to decide which form of  $f$  is the best, thus a practical approach is designing a group of  $f$  and comparing their effects. The  $f$  is designed by derivative:

$$f'_{a,n}(x) = \begin{cases} 1 & x - \varepsilon < -a \\ \frac{1}{2} - \frac{\text{erf}(n(\frac{1}{2} + \frac{x-\varepsilon}{a}))}{2\text{erf}(\frac{n}{2})} & -a \leq x - \varepsilon < 0 \\ 0 & x - \varepsilon \geq 0 \end{cases} \quad (6)$$

Here  $\varepsilon$  is the maximum eigenvalue in the SCF calculation,  $\text{erf}(x)$  is error function,  $a \geq 0$  is a constant denoting the width of transition region,  $n$  is to control the smoothness of  $f$ , the larger, the smoother. Without loss of generality, we assume  $\varepsilon = 0$  in the following discussion. The image of  $f'_{a,n}(x)$  is shown in Fig. 2(b), and  $f_{a,n}(x)$  can be obtained by integral, whose formula is provided in Supplemental Material ?? with image shown in Fig. 2(a).

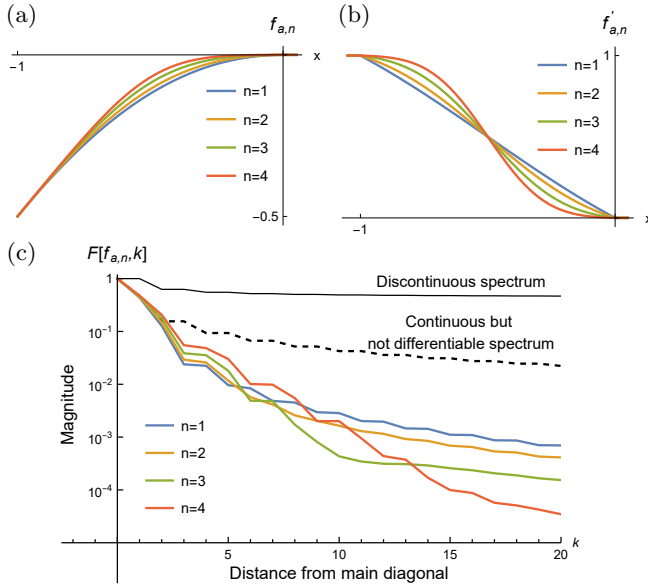


Fig. 2. Transform function and its effect. (a) Transform function  $f_{a,n}(x)$ . (b)  $f'_{a,n}(x)$ . Transition region width  $a = 1$ . (c) Decay properties of the  $m$ -banded Hermitian matrix  $H$  after transformation,  $|f_{a,n}(H)_{ij}| \leq c_{a,n} F[f_{a,n}, k]$ ,  $mk < |i-j| \leq m(k+1)$ ,  $c_{a,n}$  is a factor normalizes  $F[f_{a,n}, 0]$  to 1.

Substituting  $f_{a,n}$  into  $F$  in Eq. 5, we can obtain the results shown in Fig. 2(c). The black solid line is obtained by setting  $\varepsilon = 0.5$  in  $f_{0,n}$  that makes eigenvalue spectrum discontinuous; it does not decay to 0 at all! This phenomenon arises from approximating a discontinuous function. No matter how large the supercell is, there are always some extreme cases that a localized SCF

Hamiltonian have a decay radius proportional to the size of supercell after spectral truncation. The black dashed line is  $F[f_{0,n}, k]$  with a continuous but not differentiable spectrum; it decays fast when  $k \leq 2$ , but slower for larger  $k$ . The colored solid lines are  $F[f_{1,n}, k]$ ; here we only plot the  $a = 1$  case since nonzero  $a$  generate similar images of  $F$  after scaling, but the larger  $a$  makes  $F$  decreases faster (Supplemental Material ??). The colored lines decay much faster than the black dashed line, which means the transform function  $f_{1,n}$  works better than shifting. When  $k$  is sufficient large, larger  $n$  makes  $f_{1,n}(x)$  smoother and makes  $F[f_{1,n}, k]$  decreases faster, which is in consistent with the conjecture. But when  $k$  is small, smaller  $n$  makes  $F[f_{1,n}, k]$  decreases faster.

The decay radius  $r$  can be obtained from Fig. 2(c). If we set threshold  $t = 0.01$ ,  $n = 1$  corresponds to the minimum  $r = 5m$ ; if we set  $t = 0.001$ ,  $n = 3$  generates the minimum  $r = 8m$ , for  $m$  being the bandwidth of  $H$ . There is a trade-off between decay radius and the number of bands in choosing  $a$  and  $n$ . For larger  $a$  and  $n$ ,  $F[f_{a,n}, k]$  decreases faster and have a smaller decay radius, but  $f_{a,n}^{-1}(x)$  is more ill-conditioned near 0 and will introduce more error near the top bands, so that we have to prepare more bands in the SCF calculation. On the contrary, smaller  $a$  and  $n$  make  $F[f_{a,n}, k]$  decrease slower but could reduce the deformation of top bands. Our experience shows that  $f_{a,n}$  performs well for  $n = 3$  and

$$a = 4(\max_{\mathbf{k}}(\varepsilon_{i\mathbf{k}}) - \min_{\mathbf{k}}(\varepsilon_{i\mathbf{k}})), \quad (7)$$

where  $i$  is the band index of the top band.

*Hamiltonian transformation* – With the transformation of eigenvalues, the decay radius  $r_1$  of eigenvalues-truncated Hamiltonian  $H_1$  is reduced to  $r_2$ , as shown in Fig. 3. But the size of  $f(H_1)$  is  $N_r \times N_r$ ,  $N_r$  is the number of real space grids per unit cell, which is too large for interpolation. Projecting Hamiltonian into subspace spanned by  $N_w$  WFs could obtain a  $N_w \times N_w$  Hamiltonian matrix, but the projection will increase decay radius by  $2r_w$ , where  $r_w$  is the decay radius of WFs (proof in Supplemental Material ??). The spreads of MLWFs converge quite slowly with respect to  $\mathbf{k}$ -point density [35], which means  $r_w$  cannot be ignored.

Band structure interpolation does not necessarily require a minimal set of orbitals with a smooth gauge. We can circumvent the construction of MLWFs and reduce the size of the original Hamiltonian matrix by changing large plane-wave basis sets to a relatively small  $\mathbf{k}$ -independent numerical basis sets:

$$\Psi = [\psi_{\mathbf{k}_1}(\mathbf{r}), \dots, \psi_{\mathbf{k}_{N_k}}(\mathbf{r})] = QC. \quad (8)$$

Here  $\psi_{\mathbf{k}_i}(\mathbf{r})$  is a  $N_r \times N_b$  matrix ( $N_b$  is the number of bands) obtained by truncating orbitals from the supercell into the unit cell with  $\mathbf{R} = \mathbf{0}$ ,  $\Psi$  is obtained by concatenating  $\psi_{\mathbf{k}_i}(\mathbf{r})$  with a dimension of  $N_r \times N_b N_k$ ,  $Q$

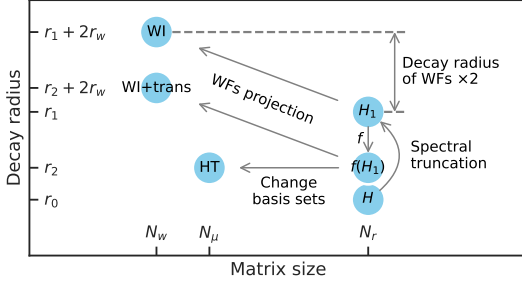


Fig. 3. The sizes and decay radii of different Hamiltonians. For interpolation, smaller matrix size is faster, and smaller decay radius is more accurate.

is basis sets with  $N_\mu$  orthonormal columns, and  $C$  is new orbitals (details in Supplemental Material ??). Through this decomposition, the original  $N_r \times N_r$  Hamiltonian matrix is reduced to  $N_\mu \times N_\mu$  without increasing decay radius. When all valence bands are included, we find that  $N_w, N_\mu, N_r$  along the  $x$ -axis of Fig. 3 typically satisfies  $N_\mu/N_r \approx 0.01 \sim 0.1$ , and  $N_w/N_\mu \approx 0.1$  (details in Supplemental Material ??).  $N_\mu$  can be further reduced if we discard some deep valence bands disentangling with desired bands.

Combining eigenvalue transformation and changing basis sets, we propose Hamiltonian transformation method to calculate band structures, which is outlined in Algorithm 1.

---

**Algorithm 1:** Hamiltonian transformation

---

**Input** : uniform grid  $\{\mathbf{k}\}$ , nonuniform path  $\{\mathbf{q}\}$ , eigenvalues  $\{\varepsilon_{i\mathbf{k}}\}$ , eigenvectors  $\{\psi_{i\mathbf{k}}(\mathbf{r})\}$

**Output:**  $\{\varepsilon_{i\mathbf{q}}\}$

1. Change basis sets

$$\Psi = [\psi_{\mathbf{k}_1}(\mathbf{r}), \dots, \psi_{\mathbf{k}_{N_k}}(\mathbf{r})] = QC$$

2. Transform eigenvalues

$$\tilde{\varepsilon}_{i\mathbf{k}} = f(\varepsilon_{i\mathbf{k}})$$

3. Construct quasi-Hamiltonian

$$M_{\mathbf{k}}(\mu, \nu) = \sum_i \tilde{\varepsilon}_{i\mathbf{k}} C_{i\mathbf{k}\mu} C_{i\mathbf{k}\nu}^*$$

4. Fourier interpolate quasi-Hamiltonian

$$M_{\mathbf{q}}(\mu, \nu) = \frac{1}{N_k} \sum_{\mathbf{k}, \mathbf{R}} M_{\mathbf{k}}(\mu, \nu) e^{i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}}$$

5. Diagonalize quasi-Hamiltonian

$$M_{\mathbf{q}}(\mu, \nu) = \sum_i \tilde{\varepsilon}_{i\mathbf{q}} C_{i\mathbf{q}\mu} C_{i\mathbf{q}\nu}^*$$

6. Recover eigenvalues

$$\varepsilon_{i\mathbf{q}} = f^{-1}(\tilde{\varepsilon}_{i\mathbf{q}})$$


---

*Numerical results* – Numerical tests are performed to compare the accuracy and efficiency of HT and WI, and the results are shown in Fig. 4. The detailed computational parameters are provided in Supplemental Material ??. Here the WI+trans method is a combination of WI and transformation shown in Fig. 3, which transform eigenvalues first, then perform WI and inverse transform eigenvalues.

Fig. 4(a) and (b) are the band structures of silicon and FeB<sub>2</sub> calculated by GW and Perdew-Burke-Ernzerhof

(PBE) [7] functional, respectively. FeB<sub>2</sub> is a topological nontrivial Weyl semimetal [54]. Red points are benchmarks, inteqp is a method in BerkeleyGW [55], which requires the orbitals on fine  $\mathbf{k}$ -point grids that cannot be taken as a general band structure calculation method. In both figures, WI (orange lines) show visible errors, but the results get better with transformation (green lines), and HT band structures (blue lines) are the best, which agree well with red points. The transformation usually causes some errors near top bands, which can be attributed to the entanglement between the top bands and higher missing bands. It can be resolved by setting a bit more bands and discarding them later. The top bands generated by HT have been discarded in Fig. 4(a).

Fig. 4(c) shows the mean absolute error (MAE) of the lowest 8 bands of silicon with PBE functional. The  $n$  comes from transform function  $f_{a,n}$ . The transformation improves the accuracy of WI significantly. HT is more accurate than WI and WI+trans, and its accuracy can be improved systematically by increasing  $N_k$ . These phenomena can be explained by decay radius. In Fig. 3, the decay radii of WI, WI+trans and HT are  $r_1 + 2r_w$ ,  $r_2 + 2r_w$  and  $r_2$ , respectively, which satisfy  $r_1 + 2r_w > r_2 + 2r_w > r_2$ . A smaller decay radius corresponds to a more localized Hamiltonian, thus the interpolation is more accurate. Furthermore, the MAEs of HT in Fig. 4(c) show similar decay patterns as the lines in Fig. 2(c), when  $N_k$  is small, smaller  $n$  has smaller MAE; when  $N_k$  is large, larger  $n$  has smaller MAE. This similarity further verifies our theoretical derivations.

In Fig. 4(d), we build supercells based on the unit cell of Si<sub>8</sub>, change the number of electrons and  $\mathbf{k}$ -points to test the time complexity of HT and WI. Parallel is disabled here. The tested time complexity of HT is  $\mathcal{O}(N_e^{2.4} N_k^{1.1})$ , which is slightly smaller than theoretical value  $\mathcal{O}(N_e^3 N_k \log(N_k))$  (details in Supplemental Material ??), since  $N_e$  is not large enough. WI has a similar scaling to HT, but its prefactor is larger, since the Wannier localization procedure converges very slowly for entangled bands. Though  $N_\mu$  is larger than  $N_w$  in Fig. 3, HT circumvents MLWFs so it does not involve any optimization and is faster.

*Conclusion and Outlook* – Hamiltonian transformation is a parameter-free, functional-independent, fast and accurate band structure calculation method. For band structure calculations, HT can outperform WI using the same number of  $\mathbf{k}$ -points for SCF calculations. The eigenvalue transformation can also be used to improve the accuracy of WI for systems with entangled bands to construct localized orbitals. Our current OpenMP based implementation of Hamiltonian transformation is already applicable to systems with hundreds of electrons. Symmetry-adapted Hamiltonian transformation and applications to the projector augmented wave (PAW) method [56] with nonorthogonal orbitals will be our future works.



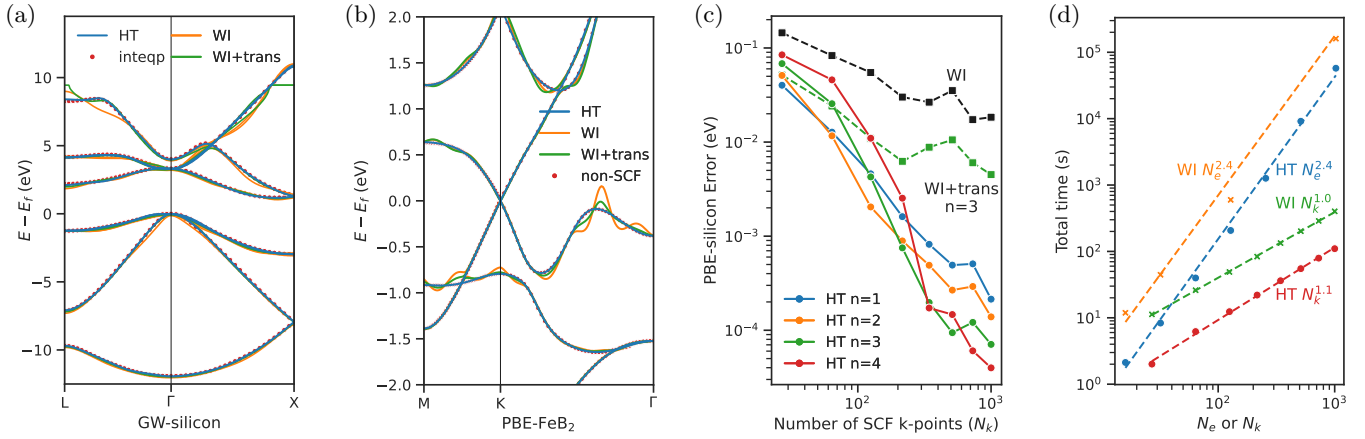


Fig. 4. The accuracy and efficiency of HT and WI. (a) GW band structure of silicon. (b) Band structures of Weyl semimetal FeB<sub>2</sub>. (c) Mean absolute error of the lowest 8 bands of silicon. (d) Time complexity with respect to the number of electrons ( $N_e$ ) and SCF  $\mathbf{k}$ -points ( $N_k$ ); parallel is disabled.

*Acknowledgement* – We thank Anil Damle, Yong Xu and Zhao Liu for valuable comments. This work is partly supported by the National Natural Science Foundation of China (22173093, 21688102, 12271109), the National Key Research and Development Program of China (2016YFA0200604, 2021YFB0300600), the Anhui Provincial Key Research and Development Program (2022a05020052), the Anhui Initiative in Quantum Information Technologies (AHY090400), the CAS Project for Young Scientists in Basic Research (YSBR-005), by the Hefei National Laboratory for Physical Sciences at the Microscale (KF2020003). LL is a Simons Investigator. We thank the Supercomputing Center of Chinese Academy of Sciences, the Supercomputing Center of USTC, the National Supercomputing Center in Wuxi, and Tianjin, Shanghai, and Guangzhou Supercomputing Centers for the computational resources.

\* [whuustc@ustc.edu.cn](mailto:whuustc@ustc.edu.cn)

† [jlyang@ustc.edu.cn](mailto:jlyang@ustc.edu.cn)

- [1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* **136**, B864 (1964).
- [2] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133 (1965).
- [3] D. M. Ceperley and B. J. Alder, Ground state of the electron gas by a stochastic method, *Phys. Rev. Lett.* **45**, 566 (1980).
- [4] J. P. Perdew and A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, *Phys. Rev. B* **23**, 5048 (1981).
- [5] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A* **38**, 3098 (1988).
- [6] C. Lee, W. Yang, and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* **37**, 785 (1988).
- [7] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [8] A. D. Becke, A new mixing of Hartree-Fock and local density-functional theories, *J. Chem. Phys.* **98**, 1372 (1993).
- [9] J. P. Perdew, M. Ernzerhof, and K. Burke, Rationale for mixing exact exchange with density functional approximations, *J. Chem. Phys.* **105**, 9982 (1996).
- [10] J. Heyd, G. E. Scuseria, and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, *J. Chem. Phys.* **118**, 8207 (2003).
- [11] L. Hedin, New method for calculating the one-particle Green's function with application to the electron-gas problem, *Phys. Rev.* **139**, A796 (1965).
- [12] M. S. Hybertsen and S. G. Louie, Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, *Phys. Rev. B* **34**, 5390 (1986).
- [13] F. Aryasetiawan and O. Gunnarsson, The GW method, *Rep. Prog. Phys.* **61**, 237 (1998).
- [14] G. Onida, L. Reining, and A. Rubio, Electronic excitations: density-functional versus many-body Green's-function approaches, *Rev. Mod. Phys.* **74**, 601 (2002).
- [15] The time complexity of Fock exchange term can be reduced to  $\mathcal{O}(N_e^3 N_k^2)$  with a much larger prefactor and a much higher memory consumption by changing the order of summation.
- [16] J. Lu and L. Ying, Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost, *J. Comput. Phys.* **302**, 329 (2015).
- [17] W. Hu, L. Lin, and C. Yang, Interpolative separable density fitting decomposition for accelerating hybrid density functional calculations with applications to defects in silicon, *J. Chem. Theory Comput.* **13**, 5420 (2017).
- [18] K. Dong, W. Hu, and L. Lin, Interpolative separable density fitting through centroidal voronoi tessellation with applications to hybrid functional electronic structure calculations, *J. Chem. Theory Comput.* **14**, 1311 (2018).

- [19] I. Duchemin and X. Blase, Cubic-scaling all-electron GW calculations with a separable density-fitting space–time approach, *J. Chem. Theory Comput.* **17**, 2383 (2021).
- [20] H. Ma, L. Wang, L. Wan, J. Li, X. Qin, J. Liu, W. Hu, L. Lin, C. Yang, and J. Yang, Realizing effective cubic-scaling Coulomb hole plus screened exchange approximation in periodic systems via interpolative separable density fitting with a plane-wave basis set, *J. Phys. Chem. A* (2021).
- [21] F. Henneke, L. Lin, C. Vorwerk, C. Draxl, R. Klein, and C. Yang, Fast optical absorption spectra calculations for periodic solid state systems, *Comm App Math Comp Sci* **15**, 89 (2020).
- [22] K. Wu, X. Qin, W. Hu, and J. Yang, Low-Rank Approximations Accelerated Plane-Wave Hybrid Functional Calculations with k-Point Sampling, *J. Chem. Theory Comput.* **18**, 206 (2021).
- [23] J. L. Whitten, Coulombic potential energy integrals and approximations, *J. Chem. Phys.* **58**, 4496 (1973).
- [24] B. I. Dunlap, J. Connolly, and J. Sabin, On some approximations in applications of  $X \alpha$  theory, *J. Chem. Phys.* **71**, 3396 (1979).
- [25] O. Vahtras, J. Almlöf, and M. Feyereisen, Integral approximations for LCAO-SCF calculations, *Chem. Phys. Lett.* **213**, 514 (1993).
- [26] F. Weigend, A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).
- [27] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions, *New J. Phys.* **14**, 053020 (2012).
- [28] L. Lin, Adaptively compressed exchange operator, *J. Chem. Theory Comput.* **12**, 2242 (2016).
- [29] W. Hu, L. Lin, A. S. Banerjee, E. Vecharynski, and C. Yang, Adaptively compressed exchange operator for large-scale hybrid density functional calculations with applications to the adsorption of water on silicene, *J. Chem. Theory Comput.* **13**, 1188 (2017).
- [30] F. Neese, F. Wennmohs, A. Hansen, and U. Becker, Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree-Fock exchange, *Chem. Phys.* **356**, 98 (2009).
- [31] R. Izsák and F. Neese, An overlap fitted chain of spheres exchange method, *J. Chem. Phys.* **135**, 144105 (2011).
- [32] E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory, *J. Chem. Phys.* **137**, 044103 (2012).
- [33] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, Tensor hypercontraction. II. Least-squares renormalization, *J. Chem. Phys.* **137**, 224106 (2012).
- [34] E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions, *J. Chem. Phys.* **137**, 221101 (2012).
- [35] N. Marzari and D. Vanderbilt, Maximally localized generalized Wannier functions for composite energy bands, *Phys. Rev. B* **56**, 12847 (1997).
- [36] 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).
- [37] G. Pizzi, V. Vitale, R. Arita, S. Blügel, F. Freimuth, G. Géranton, M. Gibertini, D. Gresch, C. Johnson, T. Koretsune, *et al.*, Wannier90 as a community code: new features and applications, *J. Phys.: Condens. Matter* **32**, 165902 (2020).
- [38] J. Jung and A. H. MacDonald, Tight-binding model for graphene  $\pi$ -bands from maximally localized Wannier functions, *Phys. Rev. B* **87**, 195450 (2013).
- [39] K. F. Garrity and K. Choudhary, Database of Wannier tight-binding Hamiltonians using high-throughput density functional theory, *Sci. Data* **8**, 1 (2021).
- [40] 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).
- [41] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Spectral and Fermi surface properties from Wannier interpolation, *Phys. Rev. B* **75**, 195121 (2007).
- [42] C. Wang, X. Liu, L. Kang, B.-L. Gu, Y. Xu, and W. Duan, First-principles calculation of nonlinear optical responses by Wannier interpolation, *Phys. Rev. B* **96**, 115147 (2017).
- [43] J. I. Mustafa, S. Coh, M. L. Cohen, and S. G. Louie, Automated construction of maximally localized Wannier functions: Optimized projection functions method, *Phys. Rev. B* **92**, 165134 (2015).
- [44] A. Damle, L. Lin, and L. Ying, Compressed representation of Kohn–Sham orbitals via selected columns of the density matrix, *J. Chem. Theory Comput.* **11**, 1463 (2015).
- [45] A. Damle, L. Lin, and L. Ying, SCDM-k: Localized orbitals for solids via selected columns of the density matrix, *J. Comput. Phys.* **334**, 1 (2017).
- [46] A. Damle and L. Lin, Disentanglement via entanglement: a unified method for Wannier localization, *Multiscale Model Simul* **16**, 1392 (2018).
- [47] E. Cancès, A. Levitt, G. Panati, and G. Stoltz, Robust determination of maximally-localized Wannier functions, *Phys. Rev. B* **95**, 075114 (2017).
- [48] K. D. Stubbs, A. B. Watson, and J. Lu, Iterated projected position algorithm for constructing exponentially localized generalized wannier functions for periodic and nonperiodic insulators in two dimensions and higher, *Phys. Rev. B* **103**, 075125 (2021).
- [49] A. A. Soluyanov and D. Vanderbilt, Wannier representation of  $Z_2$  topological insulators, *Phys. Rev. B* **83**, 035108 (2011).
- [50] H. D. Cornean, D. Monaco, and S. Teufel, Wannier functions and  $Z_2$  invariants in time-reversal symmetric topological insulators, *Rev. Math. Phys.* **29**, 1730001 (2017).
- [51] I. Souza, N. Marzari, and D. Vanderbilt, Maximally localized wannier functions for entangled energy bands, *Phys. Rev. B* **65**, 035109 (2001).
- [52] A. Damle, A. Levitt, and L. Lin, Variational formulation for Wannier functions with entangled band structure, *Multiscale Model Simul* **17**, 167 (2019).
- [53] S. N. L. Pasquini and L. Reichel, Tridiagonal Toeplitz matrices: Properties and novel applications, *Numer. Linear Algebra Appl.* **30**, 302 (2006).
- [54] H. Zhang, Y. Li, J. Hou, A. Du, and Z. Chen, Dirac state in the  $\text{FeB}_2$  monolayer with graphene-like boron sheet, *Nano Lett.* **16**, 6124 (2016).
- [55] 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 quasi-particle and optical properties of materials and nanostructures, *Comput. Phys. Commun.* **183**, 1269 (2012).
- [56] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B* **59**, 1758 (1999).
- [57] R. Baer and M. Head-Gordon, Sparsity of the density matrix in Kohn-Sham density functional theory and an assessment of linear system-size scaling methods, *Phys. Rev. Lett.* **79**, 3962 (1997).
- [58] M. Benzi, P. Boito, and N. Razouk, Decay properties of spectral projectors with applications to electronic structure, *SIAM Rev Soc Ind Appl Math* **55**, 3 (2013).
- [59] S. Bernstein, *Sur l'ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné*, Vol. 4 (Hayez, imprimeur des académies royales, 1912).
- [60] S. Xiang, X. Chen, and H. Wang, Error bounds for approximation in Chebyshev points, *Numer Math (Heidelb)* **116**, 463 (2010).
- [61] W. R. Inc., Mathematica, Version 12.0, champaign, IL, 2019.
- [62] T. F. Chan, Rank revealing QR factorizations, *Linear Algebra Appl* **88**, 67 (1987).
- [63] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, *et al.*, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [64] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, *et al.*, Advanced capabilities for materials modelling with Quantum ESPRESSO, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [65] P. Giannozzi, O. Basergio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino, *et al.*, Quantum ESPRESSO toward the exascale, *J. Chem. Phys.* **152**, 154105 (2020).
- [66] M. Schlipf and F. Gygi, Optimization algorithm for the generation of ONCV pseudopotentials, *Comput. Phys. Commun.* **196**, 36 (2015).

# Hamiltonian Transformation for Band Structure Calculations Supplemental Material

Kai Wu,<sup>1</sup> Yingzhou Li,<sup>2</sup> Wentiao Wu,<sup>1</sup> Lin Lin,<sup>3,4</sup> Wei Hu,<sup>1,\*</sup> and Jinlong Yang<sup>1,†</sup>

<sup>1</sup>*School of Future Technology, Department of Chemical Physics, and Anhui Center for Applied Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, China*

<sup>2</sup>*School of Mathematical Sciences, Fudan University, Shanghai 200433, China*

<sup>3</sup>*Department of Mathematics, University of California, Berkeley, California 94720, United States*

<sup>4</sup>*Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

## S1. DECAY PROPERTIES OF TRANSFORMATION

The basic idea of analyzing decay properties of a sparse matrix is approximating the transform function  $f$  using polynomials and analyzing the expansion coefficients. Similar ideas were adopted to study the sparsity of density matrix [1, 2].

Consider a  $m$ -banded matrix  $H$  such that:

1.  $H$  is a Hermitian matrix.
2. The eigenvalue spectrum  $\sigma(H)$  is contained in the interval  $[-1, 1]$ .
3. There is an integer  $m \geq 0$ , so that  $H_{ij} = 0$  if  $|i - j| > m$ .

Although  $H$  is limited to banded matrix here, our results in this section can be extended to general sparse matrices as long as we associate  $H$  with a degree-limited and sparsely connected graph [2]. Therefore we can take the  $m$ -banded matrix  $H$  as the SCF Hamiltonian.

We define the  $k$ th best approximation error of a transform function  $f \in C[-1, 1]$  as

$$E_k(f) = \inf \left\{ \max_{-1 \leq x \leq 1} |f(x) - p(x)| : p \in \mathcal{P}_k \right\}, \quad (\text{S1})$$

where  $\mathcal{P}_k$  is the subspace of algebraic polynomials of degree at most  $k$  in  $C[-1, 1]$ . Let  $i, j$  indices satisfy  $mk < |i - j| \leq m(k + 1)$ , for any  $p_k \in \mathcal{P}_k$ , we have

$$\begin{aligned} |f(H)_{ij}| &= |[f(H) - p_k(H)]_{ij}| \\ &\leq \|f(H) - p_k(H)\|_2 = \max_{x \in \sigma(H)} |f(x) - p_k(x)| \\ &\leq \max_{-1 \leq x \leq 1} |f(x) - p_k(x)| \end{aligned} \quad (\text{S2})$$

which means that

$$|f(H)_{ij}| \leq E_k(f). \quad (\text{S3})$$

In Eq. S2 we have used

$$|A_{ij}| \leq \sqrt{\sum_i |A_{ij}|^2} = \|Ae_j\|_2 \leq \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2. \quad (\text{S4})$$

The exact expression of the optimal  $p_k$  is unknown. While we can approximately achieve  $E_k(f)$  using Chebyshev polynomials. Approximation theory guarantees that Chebyshev polynomials are nearly optimal. The error bounds for the Chebyshev series are known for smooth functions [3, 4]. Here we calculate exact error bounds for some specific functions.

\* whuustc@ustc.edu.cn

† jlyang@ustc.edu.cn



The expression of  $f$  in the Chebyshev polynomials basis sets is

$$f(x) = \frac{1}{2}\alpha_0 + \sum_{l=1}^{\infty} \alpha_l T_l(x), \quad (\text{S5})$$

$$\alpha_l = \frac{2}{\pi} \int_0^\pi f(\cos \theta) \cos l\theta d\theta, \quad (\text{S6})$$

where  $T_l(x)$  is the  $l$ th Chebyshev polynomial of the first kind. As a result, the decay properties of  $f(H)$  can be estimated using

$$\begin{aligned} |f(H)_{ij}| &\leq E_k(f) \leq \left\| \sum_{l=k+1}^{\infty} \alpha_l T_l(x) \right\|_{x \in [-1,1]} \\ &= \frac{2}{\pi} \left\| \sum_{l=k+1}^{\infty} \cos l\theta \int_0^\pi f(\cos t) \cos lt dt \right\|_{\theta \in [0,\pi]}. \end{aligned} \quad (\text{S7})$$

In numerical calculations, Eq. S5 is adopted to simplify  $\sum_{k+1}^{\infty}$  to  $\sum_1^k$ .

## S2. ANALYTICAL SOLUTION OF DECAY PROPERTIES

Here we provide an analytical solution of the decay properties of the  $m$ -banded matrix  $H$  with a simple transform function

$$f_n(x) = \begin{cases} -(-x)^n & x \leq 0 \\ 0 & x > 0 \end{cases}, \quad (\text{S8})$$

where  $n \in \mathbb{N}$ .

According to Eq. S6,

$$\alpha_{l,n} = -\frac{2}{\pi} \int_{\frac{\pi}{2}}^\pi (-\cos \theta)^n \cos l\theta d\theta. \quad (\text{S9})$$

The closed-form of  $\alpha_{l,n}$  is quite complicated, but it turns simpler when  $l > n$ . In this case, it is easy to prove

$$\alpha_{l,0} = \frac{2 \sin(\frac{l\pi}{2})}{\pi l}, \quad (\text{S10})$$

$$\alpha_{l,1} = \frac{2 \cos(\frac{l\pi}{2})}{\pi(l^2 - 1)}, \quad (\text{S11})$$

and the recurrence relation

$$\alpha_{l,n} = \frac{-n(n-1)\alpha_{l,n-2}}{(l+n)(l-n)}, \quad (\text{S12})$$

using integration by parts. Thus

$$\alpha_{l,n} = \frac{2n! \sin(\frac{(l+n)\pi}{2})}{\pi \prod_{s=0}^n (l-n+2s)} \quad (l > n). \quad (\text{S13})$$

We have

$$\begin{aligned}
E_k(f_n) &\leq \left\| \sum_{l=k+1}^{\infty} \alpha_{l,n} T_l(x) \right\| \\
&= \frac{2n!}{\pi} \left\| \sum_{l=k-n+1}^{\infty} \frac{\sin(\frac{(l+2n)\pi}{2}) \cos(l+n)\theta}{\prod_{s=0}^n (l+2s)} \right\| \\
&= \frac{2n!}{\pi} \left\| \sum_{l=\lfloor \frac{k-n+1}{2} \rfloor}^{\infty} \frac{\sin(\frac{(2l+2n+1)\pi}{2}) \cos(2l+n+1)\theta}{\prod_{s=0}^n (2l+2s+1)} \right\| \\
&= \frac{2n!}{\pi} \left\| \sum_{l=l_0}^{\infty} \frac{(-1)^{l+n} \cos(2l+n+1)\theta}{\prod_{s=0}^n (2l+2s+1)} \right\|,
\end{aligned} \tag{S14}$$

where  $l_0 = \lfloor \frac{k-n+1}{2} \rfloor$ ,  $\lfloor x \rfloor$  is floor function and  $k \geq n$ .

We can rewrite Eq. S14 using special function:

$$\begin{aligned}
&\frac{2n!}{\pi} \left\| \Re \left[ \sum_{l=l_0}^{\infty} \frac{(-1)^{l+n} e^{i(2l+n+1)\theta}}{\prod_{s=0}^n (2l+2s+1)} \right] \right\| \\
&= \frac{n!}{2^n \pi} \frac{\Gamma(l_0 + \frac{1}{2})}{\Gamma(l_0 + n + \frac{3}{2})} \left\| \Re \left[ e^{i(2l_0+n+1)\theta} F(1, l_0 + \frac{1}{2}, l_0 + n + \frac{3}{2}; -e^{2i\theta}) \right] \right\| \\
&\leq \frac{n! \Gamma(l_0 + \frac{1}{2})}{2^n \pi \Gamma(l_0 + n + \frac{3}{2})} \left\| F(1, l_0 + \frac{1}{2}, l_0 + n + \frac{3}{2}; -e^{2i\theta}) \right\|,
\end{aligned} \tag{S15}$$

where  $\Re$  is real part,  $\Gamma$  is Gamma function, and  $F$  is hypergeometric function defined as

$$F(a, b, c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \tag{S16}$$

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}. \tag{S17}$$

It is clear to see that

$$\left| F(1, l_0 + \frac{1}{2}, l_0 + n + \frac{3}{2}; -e^{2i\theta}) \right| = \left| \sum_{k=0}^{\infty} \frac{(l_0 + \frac{1}{2})_k}{(l_0 + n + \frac{3}{2})_k} (-e^{2i\theta})^k \right| \tag{S18}$$

reaches the maximum when  $\theta = \frac{\pi}{2}$  and all terms of the sum are positive. Thus we have

$$\begin{aligned}
|f_n(H)_{ij}| &\leq E_k(f_n) \\
&\leq \frac{n! \Gamma(l_0 + \frac{1}{2})}{2^n \pi \Gamma(l_0 + n + \frac{3}{2})} F(1, l_0 + \frac{1}{2}, l_0 + n + \frac{3}{2}; 1) \\
&= \frac{n! \Gamma(l_0 + \frac{1}{2})}{2^n \pi \Gamma(l_0 + n + \frac{3}{2})} \frac{2l_0 + 2n + 1}{2n} \\
&= \frac{n!}{\pi n \prod_{s=0}^{n-1} (2l_0 + 2s + 1)} \\
&\leq \frac{(n-1)!}{\pi (k-n+1)^n} \quad (k \geq n > 0).
\end{aligned} \tag{S19}$$

Recall that  $mk < |i-j| \leq m(k+1)$ , for sufficient large  $|i-j|$ ,  $|f_n(H)_{ij}|$  decays at least polynomially with a rate  $\sim |i-j|^{-n}$ . Note that  $f_n$  is  $(n-1)$ -differentiable.

### S3. NUMERICAL SOLUTION OF DECAY PROPERTIES

With the help of numerical methods, we could study a group of more complicated transform functions:

$$f_{a,n}(x) = \begin{cases} x + a/2 & x \leq -a \\ \frac{2a(e^{-\frac{n^2}{4}} - e^{-\frac{n^2(2x+a)^2}{4a^2}}) + (2x+a)(\operatorname{erf}(\frac{n}{2}) - \operatorname{erf}(n(\frac{x}{a} + \frac{1}{2})))}{\sqrt{\pi}n} & -a < x \leq 0 \\ 0 & x > 0 \end{cases}, \quad (\text{S20})$$

where  $\operatorname{erf}(x)$  is error function.

The decay properties are calculated by

$$\begin{aligned} |f_{a,n}(H)_{ij}| &\leq E_k(f_{a,n}) \\ &\leq \frac{2}{\pi} \left\| \sum_{l=k+1}^{\infty} \cos l\theta \int_{\frac{\pi}{2}}^{\pi} f_{a,n}(\cos t) \cos lt \, dt \right\|_{\theta \in [0, \pi]} \\ &= c_{a,n} F[f_{a,n}, k], \end{aligned} \quad (\text{S21})$$

where  $c_{a,n}$  is a factor normalizes  $F[f_{a,n}, 0]$  to 1. Similar to  $f_n$  in Eq. S8,  $f_{a,n}$  also becomes smoother for larger  $n$ . We choose  $a = 1$ ,  $a = \frac{1}{2}$  and  $a = \frac{1}{4}$ , respectively, and the results are plotted in Fig. S1. Calculations are performed using Mathematica [5].

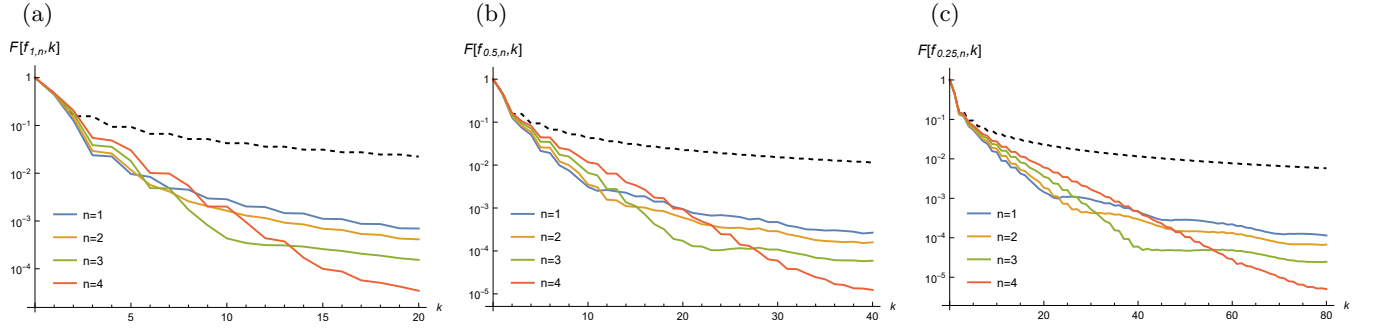


Fig. S1. The decay properties of transformed  $m$ -banded matrix  $H$ ,  $|f_{a,n}(H)_{ij}| \leq c_{a,n} F[f_{a,n}, k]$ ,  $mk < |i - j| \leq m(k + 1)$ ,  $c_{a,n}$  is a factor normalizes  $F[f_{a,n}, 0]$  to 1. (a), (b) and (c) corresponds to  $a = 1$ ,  $a = \frac{1}{2}$  and  $a = \frac{1}{4}$ , respectively. Black dashed lines are  $F[f_{0,n}, k]$ .

In Fig. S1, the black dashed lines are  $F[f_{0,n}, k]$ , and colored solid lines are  $F[f_{a,n}, k]$  with  $a \neq 0$ , we can see that  $H$  decays much faster after transformation. Comparing Fig. S1(a), (b) and (c), their images are similar after scaling, but the larger  $a$  makes  $F[f_{a,n}, k]$  decreases faster. When  $k$  is sufficient large, larger  $n$  makes  $F[f_{a,n}, k]$  smoother and makes  $H$  decreases faster, which is in consistent with the results of the analytical case. However, when  $k$  is small, smaller  $n$  makes  $H$  decreases faster.

### S4. PROJECTION OF WFS INCREASES DECAY RADIUS

All our discussions in this section are in the infinite large supercell instead of unit cell or finite supercell, and only isotropic systems are considered for simplicity. Suppose a periodic system with Born-von Karman boundary condition, its Hamiltonian is

$$H(\mathbf{r}, \mathbf{r}'), \quad \mathbf{r}, \mathbf{r}' \in \Omega^l, \quad (\text{S22})$$

where  $\Omega^l$  is supercell, and

$$\lim_{|\mathbf{r} - \mathbf{r}'| \rightarrow \infty} H(\mathbf{r}, \mathbf{r}') = 0. \quad (\text{S23})$$

In the Wannier interpolation, we cannot utilize  $H$  directly since its multiplications with eigenvectors are quite time consuming. In practice, only  $N_b N_k$  eigenvalues  $\varepsilon_{i\mathbf{k}}$  and eigenvectors  $\psi_{i\mathbf{k}}$  are available, where  $N_b$  is the number of bands,  $N_k$  is the number of SCF  $\mathbf{k}$ -points. The Hamiltonian for Wannier interpolation is

$$H_1(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^{N_b} \sum_{\mathbf{k}=1}^{N_k} \varepsilon_{i\mathbf{k}} \psi_{i\mathbf{k}}(\mathbf{r}) \psi_{i\mathbf{k}}^*(\mathbf{r}'). \quad (\text{S24})$$

We choose a threshold  $t > 0$  and define the decay radius  $r_0$  of  $H(\mathbf{r}, \mathbf{r}')$  as

$$r_0 = \sup \{ |\mathbf{r} - \mathbf{r}'| : |H(\mathbf{r}, \mathbf{r}')| \geq t \}, \quad (\text{S25})$$

and define the decay radius of  $H_1$  as  $r_1$  similarly.

The localization property of  $H_1$  is worse than  $H$ , which means  $r_1 > r_0$ . For the disentangled bands,  $H_1(\mathbf{r}, \mathbf{r}')$  usually decays rapidly, the difference between  $r_0$  and  $r_1$  can be ignored. However, for the entangled bands,  $r_1$  is usually significantly larger than  $r_0$ . Of course, we can define a new objective function that describes the locality of Hamiltonian, then localize Hamiltonian with iterative methods to obtain the “maximally localized Hamiltonian”. However, the size of Hamiltonian is large, and localizing the Hamiltonian is much more expensive than localizing Wannier functions.

Wannier interpolation projects  $H_1$  to the subspace spanned by  $\phi_{i\mathbf{k}}$ :

$$\begin{aligned} (H_{\mathbf{k}}^{\text{wan}})_{mn} &= \sum_{i,j=1}^{N_b} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' V_{\mathbf{k},im}^\dagger \psi_{i\mathbf{k}}^*(\mathbf{r}) H_1(\mathbf{r}, \mathbf{r}') \psi_{j\mathbf{k}}(\mathbf{r}') V_{\mathbf{k},jn} \\ &= \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \phi_{m\mathbf{k}}^*(\mathbf{r}) H_1(\mathbf{r}, \mathbf{r}') \phi_{n\mathbf{k}}(\mathbf{r}') \end{aligned} \quad (\text{S26a})$$

$$= \sum_{i=1}^{N_b} \varepsilon_{i\mathbf{k}} V_{\mathbf{k},im}^\dagger V_{\mathbf{k},in}, \quad (\text{S26b})$$

where  $1 \leq m, n \leq N_w$ ,  $N_w$  ( $\leq N_b$ ) is the number of Wannier bands,  $V_{\mathbf{k},im}$  is a  $k$ -dependent  $N_b \times N_w$  gauge matrix combines disentanglement and maximal localization,  $\phi_{m\mathbf{k}}(\mathbf{r})$  is the Wannier-gauge Bloch states defined as

$$\phi_{m\mathbf{k}}(\mathbf{r}) = \sum_{i=1}^{N_b} \psi_{i\mathbf{k}}(\mathbf{r}) V_{\mathbf{k},im}. \quad (\text{S27})$$

Eq. S26a and S26b are two equivalent expressions of  $H_{\mathbf{k}}^{\text{wan}}$ . Eq. S26a is used in our following analysis, and Eq. S26b is the actual formula adopted in the Wannier interpolation.

Considering the projected Hamiltonian in real space,

$$(H_{\mathbf{R}}^{\text{wan}})_{mn} = \mathcal{F}^{-1}(H_{\mathbf{k}}^{\text{wan}})_{mn} = \frac{1}{\sqrt{N_k}} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \left( \phi_{m(-\mathbf{R})}^*(\mathbf{r}) \otimes \phi_{n\mathbf{R}}(\mathbf{r}') \right) H_1(\mathbf{r}, \mathbf{r}'), \quad (\text{S28})$$

where  $\mathcal{F}^{-1}$  is inverse Fourier transform on the Bravais lattice vector  $\mathbf{R}$  (not on  $\mathbf{r}$ ),  $\otimes$  is convolution along  $\mathbf{R}$ .  $\phi_{m\mathbf{R}}(\mathbf{r})$  is Wannier function which satisfies

$$\phi_{m\mathbf{R}}(\mathbf{r}) = \mathcal{F}^{-1}(\phi_{m\mathbf{k}}(\mathbf{r})) = \frac{1}{\sqrt{N_k}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}} \phi_{m\mathbf{k}}(\mathbf{r}), \quad \mathbf{r} \in \Omega^l. \quad (\text{S29})$$

The last equality of Eq. S28 can be proven by

$$\begin{aligned} & \frac{1}{\sqrt{N_k}} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \left( \phi_{m(-\mathbf{R})}^*(\mathbf{r}) \otimes \phi_{n\mathbf{R}}(\mathbf{r}') \right) H_1(\mathbf{r}, \mathbf{r}') \\ &= \frac{1}{\sqrt{N_k}} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \sum_{\mathbf{R}_1} \phi_{m(-\mathbf{R}_1)}^*(\mathbf{r}) \phi_{n(\mathbf{R}-\mathbf{R}_1)}(\mathbf{r}') H_1(\mathbf{r}, \mathbf{r}') \\ &= \frac{1}{N_k^{3/2}} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \sum_{\mathbf{R}_1, \mathbf{k}_1, \mathbf{k}} \phi_{m\mathbf{k}_1}^*(\mathbf{r}) e^{i\mathbf{k}_1 \cdot \mathbf{R}_1} \phi_{n\mathbf{k}}(\mathbf{r}') e^{i\mathbf{k} \cdot (\mathbf{R}-\mathbf{R}_1)} H_1(\mathbf{r}, \mathbf{r}') \\ &= \frac{1}{\sqrt{N_k}} \sum_{\mathbf{k}} \int_{\Omega^l \times \Omega^l} d\mathbf{r} d\mathbf{r}' \phi_{m\mathbf{k}}^*(\mathbf{r}) H_1(\mathbf{r}, \mathbf{r}') \phi_{n\mathbf{k}}(\mathbf{r}') e^{i\mathbf{k} \cdot \mathbf{R}} \\ &= \mathcal{F}^{-1}(H_{\mathbf{k}}^{\text{wan}})_{mn}. \end{aligned} \quad (\text{S30})$$

We define the decay radius  $r_w$  of Wannier function  $\phi_{m\mathbf{R}}(\mathbf{r})$  as

$$r_w = \sup \{|\mathbf{r}| : |\phi_{m\mathbf{0}}(\mathbf{r})| \geq t\}, \quad (\text{S31})$$

and define the truncated  $\tilde{H}_1$  and  $\tilde{\phi}$  as

$$\tilde{H}_1(\mathbf{r}, \mathbf{r}') = \begin{cases} H_1(\mathbf{r}, \mathbf{r}') & |\mathbf{r} - \mathbf{r}'| \leq r_1 \\ 0 & |\mathbf{r} - \mathbf{r}'| > r_1 \end{cases}, \quad (\text{S32})$$

$$\tilde{\phi}(\mathbf{r}) = \begin{cases} \phi(\mathbf{r}) & |\mathbf{r}| \leq r_w \\ 0 & |\mathbf{r}| > r_w \end{cases}. \quad (\text{S33})$$

Thus

$$(H_{\mathbf{R}}^{\text{wan}})_{mn} = (\tilde{H}_{\mathbf{R}}^{\text{wan}})_{mn} + \mathcal{O}(t) = \frac{1}{\sqrt{N_k}} \int_{\Omega' \times \Omega'} d\mathbf{r} d\mathbf{r}' \left( \tilde{\phi}_{m(-\mathbf{R})}^*(\mathbf{r}) \otimes \tilde{\phi}_{n\mathbf{R}}(\mathbf{r}') \right) \tilde{H}_1(\mathbf{r}, \mathbf{r}') + \mathcal{O}(t). \quad (\text{S34})$$

$\tilde{H}_{\mathbf{R}}^{\text{wan}} \neq 0$  requires

$$\begin{cases} |\mathbf{r} - \mathbf{r}'| \leq r_1 \\ |\mathbf{r} + \mathbf{R}_1| \leq r_w \\ |\mathbf{r}' - \mathbf{R} + \mathbf{R}_1| \leq r_w, \end{cases} \quad (\text{S35})$$

which means that

$$|\mathbf{R}| \leq r_1 + 2r_w. \quad (\text{S36})$$

Strictly speaking, we only provide an upper bound here and cannot guarantee the upper bound is always reached. The convolution of two functions may have a smaller decay radius if their spectra have non-intersecting supports. However, the supports of Hamiltonian and Wannier functions in reciprocal space usually take up the whole cut-off sphere and coincide with each other. We can expect  $H_{\mathbf{R}}^{\text{wan}}$  have a decay radius of  $r_1 + 2r_w$  in most cases.

## S5. CHANGING BASIS SETS

To reduce the size of basis sets, we expand the original orbitals  $\psi_{i\mathbf{k}}$  in the auxiliary basis  $Q_{\mu}(\mathbf{r})$ :

$$\psi_{i\mathbf{k}}(\mathbf{r}) = \sum_{\mu=1}^{N_{\mu}} Q_{\mu}(\mathbf{r}) C_{i\mathbf{k}\mu}. \quad (\text{S37})$$

Here  $\psi_{i\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{i\mathbf{k}}(\mathbf{r})$ , but  $\mathbf{r}$  is restricted to the unit cell with  $\mathbf{R} = \mathbf{0}$ . Noting that  $Q_{\mu}(\mathbf{r})$  is independent of  $\mathbf{k}$ , orbitals at all  $\mathbf{k}$ -points share the same auxiliary basis, and changing to this basis sets will not affect the decay radius of Hamiltonian.

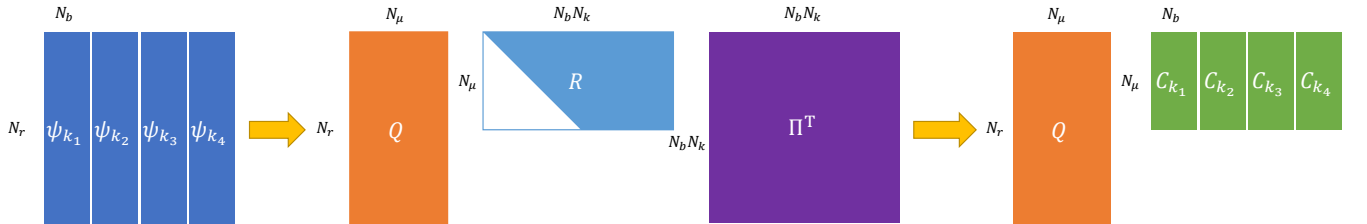


Fig. S2. Graphical representation of changing basis sets through QRCP. Equivalent to Eq. S38.

Changing basis sets is performed by QR factorization with column pivoting (QRCP) [6]:

$$\Psi \Pi = Q R \Leftrightarrow \Psi = Q R \Pi^T = Q C. \quad (\text{S38})$$



Fig. S2 is the graphical representation of Eq. S38, where  $\Psi$  is a  $N_r \times N_k N_b$  matrix obtained by concatenating all  $\psi_{i\mathbf{k}}$  along the  $N_b$  dimension,  $Q$  is a  $N_r \times N_\mu$  matrix whose columns are orthonormal,  $R$  is an upper triangular matrix, and  $\Pi$  is a permutation matrix.  $\Pi$  permutes columns of  $\psi$  such that the magnitudes of the diagonal elements of  $R$  are in a nonincreasing ordering. The magnitudes of diagonal elements of  $R$  indicate the importance of the corresponding rows of  $R$  and columns of  $Q$ . We denote  $N_\mu$  as the number of diagonal elements of  $R$  larger than a threshold.  $N_\mu$  is the numerical rank of space spanned by the lowest  $N_b$  orbitals of all  $\mathbf{k}$ -points; it will tend to be a constant multiple of  $N_b$  even if  $N_k$  approaches infinity.

The Hamiltonian can be written as

$$H_{\mathbf{k}}(\mathbf{r}, \mathbf{r}') = \sum_{\mu\nu} Q_\mu(\mathbf{r}) M_{\mathbf{k}}(\mu, \nu) Q_\nu^*(\mathbf{r}'), \quad (\text{S39})$$

$$M_{\mathbf{k}}(\mu, \nu) = \sum_{i=1}^{N_b} \varepsilon_{i\mathbf{k}} C_{i\mathbf{k}\mu} C_{i\mathbf{k}\nu}^*. \quad (\text{S40})$$

Considering the Fourier interpolation of Eq. S39,

$$\begin{aligned} H_{\mathbf{q}}(\mathbf{r}, \mathbf{r}') &= \frac{1}{N_k} \sum_{\mathbf{k}, \mathbf{R}} H_{\mathbf{k}}(\mathbf{r}, \mathbf{r}') e^{i(\mathbf{k}-\mathbf{q})\mathbf{R}} \\ &= \sum_{\mu\nu} Q_\mu(\mathbf{r}) \left( \frac{1}{N_k} \sum_{\mathbf{k}, \mathbf{R}} M_{\mathbf{k}}(\mu, \nu) e^{i(\mathbf{k}-\mathbf{q})\mathbf{R}} \right) Q_\nu^*(\mathbf{r}') \\ &= \sum_{\mu\nu} Q_\mu(\mathbf{r}) M_{\mathbf{q}}(\mu, \nu) Q_\nu^*(\mathbf{r}'). \end{aligned} \quad (\text{S41})$$

We can see the Fourier interpolation of a  $N_r \times N_r$  matrix  $H_{\mathbf{k}}(\mathbf{r}, \mathbf{r}')$  is simplified to that of a  $N_\mu \times N_\mu$  matrix  $M_{\mathbf{k}}(\mu, \nu)$ .

When the projector augmented wave method (PAW) or ultrasoft pseudopotentials are adopted, orbitals are nonorthogonal. Hamiltonian transformation method is still available as long as we replace  $\psi_{\mathbf{k}}$  with  $S^{\frac{1}{2}}\psi_{\mathbf{k}}$ , where  $S$  is the overlap matrix independent of orbitals and  $\mathbf{k}$ -points.

## S6. THE SIZE OF $N_\mu$

We perform some tests to study the relationship between  $N_\mu$ ,  $N_r$ ,  $N_b$  and  $N_w$ . The system is silicon, generalized gradient approximation of Perdew-Burke-Ernzerhof (GGA-PBE) [7] functional is adopted, cutoff energy is 100 Ry.  $n = 3$  for transform function. The SCF FFT grid is  $48 \times 48 \times 48$ , in HT we adopt a coarse grid for orbitals, which is  $25 \times 25 \times 25$ , so  $N_r = 15625$ . Here we do not perform WI, but the mean absolute error (MAE) considers the lowest 8 bands, it is equivalent to  $N_w = 8$ .

We sort the diagonal values of  $R$  in QRCP according to their absolute values, assuming the largest as 1, and ignore those smaller than  $\epsilon_t$ . The relationship between  $\epsilon_t$ ,  $N_\mu$ ,  $N_k$  and MAE is shown in Fig. S3. If  $\epsilon_t$  is fixed,  $N_\mu$  tends to a constant when  $N_k$  is large enough. When  $N_k = 27$ ,  $N_\mu \approx 90$  could reach the maximum accuracy; when  $N_k = 1000$ ,  $N_\mu \approx 225$  could reach the maximum accuracy. In this case, we have  $N_\mu \approx 0.01N_r$  and  $N_\mu \approx 20N_w$ . In the QRCP procedure we consider 14 bands, i.e.  $N_b = 14$ . If  $N_b$  is reduced to 10,  $N_\mu$  can be smaller.  $N_\mu/N_r$  is also dependent of cutoff energy, the higher cutoff energy corresponds to smaller  $N_\mu/N_r$ . From our experience,  $N_\mu/N_r \approx 0.01 \sim 0.1$  in most cases.

## S7. COMPUTATIONAL METHODS

Our Hamiltonian transformation method is implemented in Quantum ESPRESSO [8–10], DFT calculations are performed by Quantum ESPRESSO. The quasi-particle energy in the GW level is calculated by BerkeleyGW [11, 12]. Wannier interpolations are performed by Wannier90 [13].

We adopt optimized norm-conserving Vanderbilt (ONCV) [14] pseudopotential in all calculations. Selected columns of the density matrix (SCDM) method [15–17] is used for Wannier interpolation if not specified otherwise. In Fig. ??(a), the unit cell of silicon contains two silicon atoms,  $\mathbf{k}$ -point mesh is  $5 \times 5 \times 5$ , cutoff energy is 25 Ry, and we use  $\text{sp}^3$  projection for Wannier interpolation. In Fig. ??(b),  $\mathbf{k}$ -point mesh is  $11 \times 11 \times 1$ , cutoff energy is 50 Ry, SCDM- $\mu$  is 2, SCDM- $\sigma$  is 1. In Fig. ??(c), generalized gradient approximation of Perdew-Burke-Ernzerhof

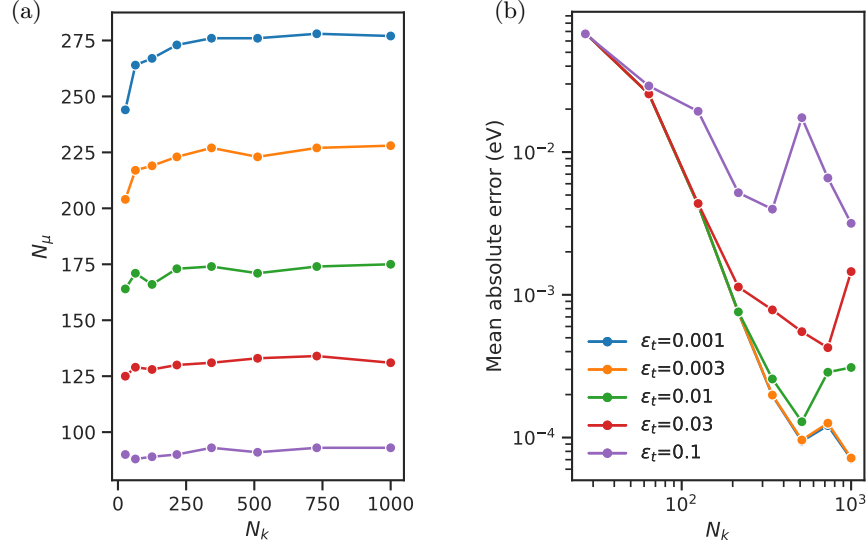


Fig. S3. The relationship between  $\epsilon_t$ ,  $N_\mu$ ,  $N_k$  and MAE.

(GGA-PBE) [7] functional is adopted, cutoff energy is 100 Ry, SCDM- $\mu$  is 10, SCDM- $\sigma$  is 2. The mean absolute error (MAE) calculation considers  $\mathbf{k}$ -points between  $\Gamma$  and X. In Fig. ??(d), GGA-PBE functional is adopted, cutoff energy is 10 Ry.

## S8. TIME COMPLEXITY

The theoretical time complexity of Hamiltonian transformation is shown in Table S1. Its computational bottleneck lies in randomized QRCP and iterative diagonalization, which have time complexities of  $\mathcal{O}(N_\mu^2(N_r + N_b N_k))$  and  $\mathcal{O}(N_\mu^2 N_b N_q)$ , respectively. At present, randomized QRCP, NUFFT, and iterative diagonalization are not yet implemented in our code. They are temporarily replaced by QRCP, matrix multiplication, and direct diagonalization, respectively.

If we assume  $N_r, N_\mu, N_b$  proportion to the number of electrons  $N_e$ , and  $N_q$  is constant, the total time complexity of Hamiltonian transformation is  $\mathcal{O}(N_e^3 N_k \log(N_k))$ . Here the  $\log(N_k)$  term comes from fast Fourier transform, which has a small preconstant and is neglectable in most cases.

Table S1. The theoretical time complexity of Hamiltonian transformation.  $N_r$  is the number of real space grids,  $N_\mu$  is the size of new basis sets,  $N_k$  is the number of SCF  $\mathbf{k}$ -points,  $N_b$  and  $N_q$  are the number of bands and  $\mathbf{k}$ -points in the band structure calculation.

Operation	Algorithm	Time complexity
Change basis sets	Randomized QRCP	$\mathcal{O}(N_\mu^2(N_r + N_b N_k))$
Build Hamiltonian	Matrix multiplication	$\mathcal{O}(N_\mu^2 N_b N_k)$
Fourier interpolation	Fast Fourier transform (FFT)	$\mathcal{O}(N_\mu^2 N_k \log(N_k))$
	Nonuniform FFT (NUFFT) or butterfly factorization	$\mathcal{O}(N_\mu^2 N_q \log(N_q))$
Diagonalization	Iterative diagonalization	$\mathcal{O}(N_\mu^2 N_b N_q)$

[1] R. Baer and M. Head-Gordon, Sparsity of the density matrix in Kohn-Sham density functional theory and an assessment of linear system-size scaling methods, Phys. Rev. Lett. **79**, 3962 (1997).

- [2] M. Benzi, P. Boito, and N. Razouk, Decay properties of spectral projectors with applications to electronic structure, *SIAM Rev Soc Ind Appl Math* **55**, 3 (2013).
- [3] S. Bernstein, *Sur l'ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné*, Vol. 4 (Hayez, imprimeur des académies royales, 1912).
- [4] S. Xiang, X. Chen, and H. Wang, Error bounds for approximation in Chebyshev points, *Numer Math (Heidelb)* **116**, 463 (2010).
- [5] W. R. Inc., Mathematica, Version 12.0, champaign, IL, 2019.
- [6] T. F. Chan, Rank revealing QR factorizations, *Linear Algebra Appl* **88**, 67 (1987).
- [7] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [8] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, *et al.*, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [9] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, *et al.*, Advanced capabilities for materials modelling with Quantum ESPRESSO, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [10] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino, *et al.*, Quantum ESPRESSO toward the exascale, *J. Chem. Phys.* **152**, 154105 (2020).
- [11] M. S. Hybertsen and S. G. Louie, Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, *Phys. Rev. B* **34**, 5390 (1986).
- [12] 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).
- [13] G. Pizzi, V. Vitale, R. Arita, S. Blügel, F. Freimuth, G. Géranton, M. Gibertini, D. Gresch, C. Johnson, T. Koretsune, *et al.*, Wannier90 as a community code: new features and applications, *J. Phys.: Condens. Matter* **32**, 165902 (2020).
- [14] M. Schlipf and F. Gygi, Optimization algorithm for the generation of ONCV pseudopotentials, *Comput. Phys. Commun.* **196**, 36 (2015).
- [15] A. Damle, L. Lin, and L. Ying, Compressed representation of Kohn–Sham orbitals via selected columns of the density matrix, *J. Chem. Theory Comput.* **11**, 1463 (2015).
- [16] A. Damle, L. Lin, and L. Ying, SCDM-k: Localized orbitals for solids via selected columns of the density matrix, *J. Comput. Phys.* **334**, 1 (2017).
- [17] A. Damle and L. Lin, Disentanglement via entanglement: a unified method for Wannier localization, *Multiscale Model Simul* **16**, 1392 (2018).