The Journal of Open Source Software

pengWann: Descriptors of chemical bonding from Wannier functions

Patrick J. Taylor [1,2] **and Benjamin J. Morgan** [1,2]

1 Department of Chemistry, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom **2** The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot, OX11 0RA, United Kingdom

DOI: 10.21105/joss.07890

Software

- Review
- Repository
- Archive

Editor: Rachel Kurchin

Reviewers:

- @JosePizarro3
- @LIVazquezS

Submitted: 14 February 2025
Published: 18 April 2025

Summary

First-principles calculations of periodic systems typically represent electronic structures through sets of eigenvectors and their corresponding eigenvalues, obtained by diagonalising a many-body Hamiltonian. These eigenvectors, known as Bloch states, molecular orbitals, or crystal orbitals, are generally delocalised across the entire structure, making them difficult to interpret in terms of chemically intuitive concepts such as bonds. To address this, it is common practice to project these extended Bloch states onto a localised basis, enabling the calculation of various local descriptors of chemical bonding or electronic structure. pengwann is a Python package for calculating some of these descriptors by projecting Bloch states onto Wannier functions (Kundu et al., 2021). Wannier functions provide a highly optimised local basis and, when derived from energetically isolated bands, span the same Hilbert space as the canonical Bloch states. The package provides a simple interface to the popular Wannier90 code (Pizzi et al., 2020), making it readily accessible to researchers already familiar with this widely-used tool.

Statement of need

The technique of deriving bonding descriptors from the projection of Bloch states onto a local basis is widespread in materials modelling (Coles et al., 2023; Foppa et al., 2016; Hu et al., 2025; Legein et al., 2024; Lin et al., 2018; Yu et al., 2024; Zhong et al., 2020; Zhu et al., 2025). Key to the success of this method is the choice of local basis functions, which should be able to effectively reproduce the canonical Bloch states when appropriately combined. The ability of a given basis set to accurately represent the original Bloch states is quantified by the spilling factor (Sanchez-Portal et al., 1995):

$$S = \frac{1}{N_b}\frac{1}{N_k}\sum_{nk} 1 - \sum_{\alpha} |\langle \psi_{nk} | \phi_{\alpha} \rangle|^2,$$

where $|\psi_{nk}\rangle$ is a Bloch state, $|\phi_{\alpha}\rangle$ is a localised basis function, n labels bands, k labels k-points, N_b is the total number of bands and N_k is the total number of k-points. The spilling factor takes values between 0 and 1; if the local basis spans the same Hilbert space as the Bloch states, then $S = 0$, while $S = 1$ indicates that the two bases are orthogonal to one another. The most common choice of local basis is to use atomic or pseudo-atomic orbitals (Bunge et al., 1993; Koga et al., 1999; Maintz et al., 2016; Ruggiero et al., 2015), with these parameterised with respect to atomic species but usually not to the specific system being modelled. Because these orbitals are designed to be transferable between materials, they cannot represent the Bloch states of an arbitrary system perfectly: the spilling factor will always be non-zero and some information will be lost during the projection. For many systems, the error introduced by this loss of information is relatively small and so can be safely ignored, but this is not always

the case. To give a pathological example, in electride materials, atom-centred basis functions cannot accurately represent the Bloch states because some of the valence electrons behave like anions and occupy their own distinct space in the structure (Kanno et al., 2021).

pengwann employs a Wannier basis which, when derived from energetically isolated bands, spans the same vector space as the canonical Bloch states. The spilling factor is therefore strictly zero and there is no loss of information in switching from the Bloch basis to the Wannier basis. For Wannier functions derived from bands that are not energetically isolated everywhere in the Brillouin zone, the spilling factor will no longer be strictly zero, but should remain very small, since Wannier functions are calculated by a unitary transformation of the Bloch states (Marzari et al., 2012). Importantly, Wannier functions are not constrained to be atom-centred and can therefore accurately represent the Bloch states of electrides and other such anomalous systems. More generally, even in systems where pre-defined atomic or pseudo-atomic orbital basis sets perform well, a Wannier basis will always give a reduced spilling factor, thereby reducing the corresponding error in all derived descriptors.

pengwann implements the following core features:

- Identification of interatomic and on-site interactions in terms of the Wannier functions associated with each atom
- Parsing of Wannier90 output files
- Parallelised computation of the following descriptors:
  - The Wannier orbital Hamilton population (WOHP)
  - The Wannier orbital bond index (WOBI)
  - The Wannier-projected density of states (pDOS)
  - Orbital and k-resolved implementations of all of the above
- Integration of descriptors to derive:
  - Löwdin-style populations and charges
  - Measures of bond strength and bond order

A full API reference is available as part of the official pengwann documentation at pengwann.readthedocs.io, which also includes further details with regards to the mathematical formalism of Wannier-derived bonding descriptors and a series of tutorial-style examples.

Related software

The LOBSTER code (Maintz et al., 2016; Nelson et al., 2020) implements much of the same functionality as pengwann using basis sets of pre-defined atomic and pseudo-atomic orbitals. LOBSTER offers additional features not directly supported by pengwann, such as generating fatband plots and obtaining localised molecular orbitals via transformation of the projected atomic orbital basis (Müller et al., 2024). We anticipate that most potential users of pengwann will already be familiar with LOBSTER, warranting a brief discussion of the relative advantages and disadvantages of each code.

pengwann achieves lower spilling factors than LOBSTER[1] and readily handles systems when non-atom-centered basis functions are preferable. Moreover, since Wannier functions serve many purposes in addition to the calculation of bonding descriptors, they may already have been calculated for systems of interest, making the use of pengwann particularly efficient in these cases. Conversely, LOBSTER's main advantage lies in its pre-defined atomic and pseudo-atomic basis sets, which can be applied to any system with minimal user input, while obtaining high-quality Wannier functions can be complex due to their strong non-uniqueness (Marzari et al., 2012). However, recent advances such as those of Refs. (Qiao et al., 2023; Vitale et al., 2020) have significantly simplified the computation of well-localised Wannier

[1]For example, even for relatively simple systems such as diamond and rutile, we find that (when using the recommended basis functions) LOBSTER still yields non-zero spilling factors of 9.8×10^{-3} and 9.5×10^{-3} respectively. For the same set of Bloch states, pengwann is able to reduce the spilling factor to $< 1.0 \times 10^{-13}$ in both cases.

functions with appropriate symmetry for chemical bonding analysis, lowering the barrier to their use.

The WOBSTER code (Kundu et al., 2021; Xiao, 2020) implements a subset of the features found in `pengwann`, allowing users to compute the Wannier-projected density of states and the Wannier orbital Hamilton population. `pengwann` provides a broader set of features than WOBSTER and is more performant: `pengwann` leverages `numpy` (Harris et al., 2020) and the built-in `multiprocessing` library to vectorise and parallelise most operations, whereas WOBSTER's loop-based approach is considerably slower. To give a specific example: the Wannier orbital Hamilton population for a given pair of Wannier functions is computed in WOBSTER using a triple nested loop over energies, k-points and bands. The same calculation in `pengwann` is fully vectorised across these three dimensions and furthermore many such computations can be carried out in parallel for different pairs of Wannier functions.

Acknowledgements

The authors thank the Faraday Institution CATMAT project (EP/S003053/1, FIRG016) for financial support and the Michael High-Performance Computing (HPC) facility (FIRG030). P.J.T would like to acknowledge Chengcheng Xiao (author of the WOBSTER code) for inspiring the development of `pengwann` and thanks the University of Bath (indirectly via the EPSRC) for PhD funding. B.J.M. thanks the Royal Society for a fellowship (URF/R/191006).

References

Bunge, C. F., Barrientos, J. A., & Bunge, A. V. (1993). Roothaan-Hartree-Fock ground-state atomic wave functions: Slater-type orbital expansions and expectation values for $Z = 2$–54. *Atomic Data and Nuclear Data Tables*, *53*(1), 113–162. https://doi.org/10.1006/adnd.1993.1003

Coles, S. W., Falkowski, V., Geddes, H. S., Pérez, G. E., Booth, S. G., Squires, A. G., O'Rourke, C., McColl, K., Goodwin, A. L., Cussen, S. A., Clarke, S. J., Islam, M. S., & Morgan, B. J. (2023). Anion-polarisation-directed short-range-order in antiperovskite Li_2FeSO. *Journal of Materials Chemistry A*, *11*(24), 13016–13026. https://doi.org/10.1039/d2ta10037a

Foppa, L., Copéret, C., & Comas-Vives, A. (2016). Increased back-bonding explains step-edge reactivity and particle size effect for CO activation on Ru nanoparticles. *Journal of the American Chemical Society*, *138*(51), 16655–16668. https://doi.org/10.1021/jacs.6b08697

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hu, M., Yang, J., Wang, Y., Xia, J., Gan, Q., Yang, S., Xu, J., Liu, S., Yin, W., Jia, B., Xie, L., Li, H., & He, J. (2025). Helical dislocation-driven plasticity and flexible high-performance thermoelectric generator in α-Mg_3Bi_2 single crystals. *Nature Communications*, *16*(1), 128. https://doi.org/10.1038/s41467-024-55689-7

Kanno, S., Tada, T., Utsumi, T., Nakamura, K., & Hosono, H. (2021). Electronic correlation strength of inorganic electrides from first principles. *The Journal of Physical Chemistry Letters*, *12*(50), 12020–12025. https://doi.org/10.1021/acs.jpclett.1c03637

Koga, T., Kanayama, K., Watanabe, S., & Thakkar, A. J. (1999). Analytical Hartree–Fock wave functions subject to cusp and asymptotic constraints: He to Xe, Li^+ to Cs^+, H^- to I^-. *International Journal of Quantum Chemistry*, *71*(6), 491–497. https://doi.org/10.1002/(SICI)1097-461X(1999)71:6%3C491::AID-QUA6%3E3.0.CO;2-T

Kundu, S., Bhattacharjee, S., Lee, S.-C., & Jain, M. (2021). Population analysis with Wannier orbitals. *The Journal of Chemical Physics*, *154*(10), 104111. https://doi.org/10.1063/5.0032605

Legein, C., Morgan, B. J., Squires, A. G., Body, M., Li, W., Burbano, M., Salanne, M., Charpentier, T., Borkiewicz, O. J., & Dambournet, D. (2024). Correlated anion disorder in heteroanionic cubic $TiOF_2$. *Journal of the American Chemical Society*, *146*(31), 21889–21902. https://doi.org/10.1021/jacs.4c06304

Lin, J., Zhang, S., Guan, W., Yang, G., & Ma, Y. (2018). Gold with +4 and +6 oxidation states in AuF_4 and AuF_6. *Journal of the American Chemical Society*, *140*(30), 9545–9550. https://doi.org/10.1021/jacs.8b04563

Maintz, S., Deringer, V. L., Tchougréeff, A. L., & Dronskowski, R. (2016). LOBSTER: A tool to extract chemical bonding from plane-wave based DFT. *Journal of Computational Chemistry*, *37*(11), 1030–1035. https://doi.org/10.1002/jcc.24300

Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I., & Vanderbilt, D. (2012). Maximally localized Wannier functions: Theory and applications. *Reviews of Modern Physics*, *84*, 1419–1475. https://doi.org/10.1103/RevModPhys.84.1419

Müller, P. C., Schmit, N., Sann, L., Steinberg, S., & Dronskowski, R. (2024). Fragment orbitals extracted from first-principles plane-wave calculations. *Inorganic Chemistry*, *63*(43), 20161–20172. https://doi.org/10.1021/acs.inorgchem.4c01024

Nelson, R., Ertural, C., George, J., Deringer, V. L., Hautier, G., & Dronskowski, R. (2020). LOBSTER: Local orbital projections, atomic charges, and chemical-bonding analysis from projector-augmented-wave-based density-functional theory. *Journal of Computational Chemistry*, *41*(21), 1931–1940. https://doi.org/10.1002/jcc.26353

Pizzi, G., Vitale, V., Arita, R., Blügel, S., Freimuth, F., Géranton, G., Gibertini, M., Gresch, D., Johnson, C., Koretsune, T., Ibañez-Azpiroz, J., Lee, H., Lihm, J.-M., Marchand, D., Marrazzo, A., Mokrousov, Y., Mustafa, J. I., Nohara, Y., Nomura, Y., … Yates, J. R. (2020). Wannier90 as a community code: New features and applications. *Journal of Physics: Condensed Matter*, *32*(16), 165902. https://doi.org/10.1088/1361-648X/ab51ff

Qiao, J., Pizzi, G., & Marzari, N. (2023). Projectability disentanglement for accurate and automated electronic-structure Hamiltonians. *Npj Computational Materials*, *9*(1), 208. https://doi.org/10.1038/s41524-023-01146-w

Ruggiero, M. T., Erba, A., Orlando, R., & Korter, T. M. (2015). Origins of contrasting copper coordination geometries in crystalline copper sulfate pentahydrate. *Physical Chemistry Chemical Physics*, *17*, 31023–31029. https://doi.org/10.1039/C5CP05554G

Sanchez-Portal, D., Artacho, E., & Soler, J. M. (1995). Projection of plane-wave calculations into atomic orbitals. *Solid State Communications*, *95*(10), 685–690. https://doi.org/10.1016/0038-1098(95)00341-X

Vitale, V., Pizzi, G., Marrazzo, A., Yates, J. R., Marzari, N., & Mostofi, A. A. (2020). Automated high-throughput Wannierisation. *Npj Computational Materials*, *6*(1), 66. https://doi.org/10.1038/s41524-020-0312-y

Xiao, C. (2020, November). *WOBSTER*. https://github.com/Chengcheng-Xiao/WOBSTER

Yu, S., Yamauchi, H., Wang, S., Aggarwal, A., Kim, J., Gordiz, K., Huang, B., Xu, H., Zheng, D. J., Wang, X., Iriawan, H., Menga, D., & Shao-Horn, Y. (2024). CO_2-to-methanol electroconversion on a molecular cobalt catalyst facilitated by acidic cations. *Nature Catalysis*, *7*(9), 1000–1009. https://doi.org/10.1038/s41929-024-01197-2

Zhong, C., Liu, B., Ding, J., Liu, X., Zhong, Y., Li, Y., Sun, C., Han, X., Deng, Y., Zhao, N., & Hu, W. (2020). Decoupling electrolytes towards stable and high-energy rechargeable aqueous zinc–manganese dioxide batteries. *Nature Energy*, *5*(6), 440–449.

https://doi.org/10.1038/s41560-020-0584-y

Zhu, T., Ma, L., Duan, X., Deng, S., & Liu, S. (2025). Origin of interstitial doping induced coercive field reduction in ferroelectric hafnia. *Physical Review Letters*, *134*, 056802. https://doi.org/10.1103/PhysRevLett.134.056802