

Bridging Crystal Structure and Material Properties via Bond-Centric Descriptors

Jian-Feng Zhang,^{1,*} Ze-Feng Gao,^{2,3} Xiao-Qi Han,^{2,3} Bo Zhan,⁴ Dingshun Lv,⁵
Miao Gao,⁶ Kai Liu,^{2,3,†} Xinguo Ren,^{4,‡} Zhong-Yi Lu,^{2,3,§} and Tao Xiang^{4,7,¶}

¹*Center for High Pressure Science and Technology Advanced Research, Beijing 100193, China.*

²*School of Physics and Beijing Key Laboratory of Opto-electronic Functional Materials & Micro-nano Devices, Renmin University of China, Beijing 100872, China*

³*Key Laboratory of Quantum State Construction and Manipulation (Ministry of Education), Renmin University of China, Beijing 100872, China*

⁴*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*

⁵*Field Quantum Research Institute, Beijing 100084, China*

⁶*Department of Physics, School of Physical Science and Technology, Ningbo University, Zhejiang 315211, China*

⁷*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

(Dated: July 28, 2026)

Although chemical bonding is the fundamental mechanistic bridge connecting atomic structure to macroscopic material properties, current data-driven materials science largely treats it as an implicit “black box”. Existing machine-learning (ML) models rely predominantly on geometric coordinates, forcing them to infer complex quantum-mechanical relationships from scratch. This lack of intermediate physical features limits model interpretability and generalizability, particularly when training data are scarce. To solve this problem, we introduce **MattKeyBond**, a bond-centric materials database that explicitly maps the local electronic landscape and bonding interactions of materials. Building on this, we propose Bonding Attractivity (BA), a novel element-specific descriptor that quantifies the intrinsic capability of atoms to form covalent networks. By providing precomputed, energy-resolved bonding descriptors, **MattKeyBond** transforms the implicit “black box” into a set of physically interpretable features. This strategy relieves ML models of the burden of deducing physical laws from pure geometry, enabling accurate predictions even with limited data and seamlessly integrating electronic structure theory into modern AI workflows.

I. INTRODUCTION

Crystal structure dictates the physical properties of materials, yet it is chemical bonding that mechanistically links atomic arrangement to electronic structure and macroscopic behavior. In solids, each atom interacts with a local environment determined by neighboring atoms and the resulting crystal field. Through field-induced charge redistribution and orbital hybridization, inter-atomic bonding yields an inherent electronic structure strictly dependent on atomic arrangement. Atom- and bond-resolved analyses are therefore essential for establishing mechanistic connections between structure and properties, and for enabling the rapid screening and design of materials.

However, despite the emergence of a rich ecosystem of materials databases, current resources generally lack this crucial information. Existing platforms primarily capture structural geometry and global scalar properties. General-purpose databases, such as the Materials Project (MP)[1], Atomly[2], OQMD[3, 4], AFLOW[5], and NOMAD[6], aggregate high-throughput calculations to provide structures, formation energies, and electronic

band structures. Similarly, domain-targeted databases like C2DB[7, 8], JARVIS-DFT[9], CoRE MOF[10], and the Open Catalyst Project (OC20/OC22)[11, 12] focus on specific material classes. Despite their breadth, the absence of physics-based intermediate features limits the effectiveness of current data-driven materials science. Models are therefore forced to rely predominantly on geometric information [13–16], effectively treating the chemical bond as an implicit “black box” (top panel of Figure 1). This compels machine-learning models to implicitly relearn complex quantum-mechanical relationships from geometry alone, which limits their interpretability and generalizability, particularly in complex systems where experimental data are scarce, such as superconductors.

While recent efforts have attempted to incorporate electronic features by learning continuous charge densities or DFT Hamiltonians [17, 18], a comprehensive, bond-centric database remains absent. To bridge this gap and decode the underlying interactions, we introduce **MattKeyBond**, a bond-centric materials database constructed from high-throughput first-principles calculations. Leveraging the Closest Wannier Functions (CWF)[19] method and integral crystal orbital Hamilton population (ICOHP) analyses[20–22], **MattKeyBond** transcends conventional structural geometry to map the electronic landscape into a real-space, bond-resolved representation. Through its high-fidelity, atom-pair-resolved features, including charge transfer, orbital Hamiltonians, bond energy, and bond-order-related density matrices, **MattKeyBond** enables systematic, searchable, and com-

* jianfeng.zhang@hpstar.ac.cn

† kliu@ruc.edu.cn

‡ renxg@iphy.ac.cn

§ zlu@ruc.edu.cn

¶ txiang@iphy.ac.cn

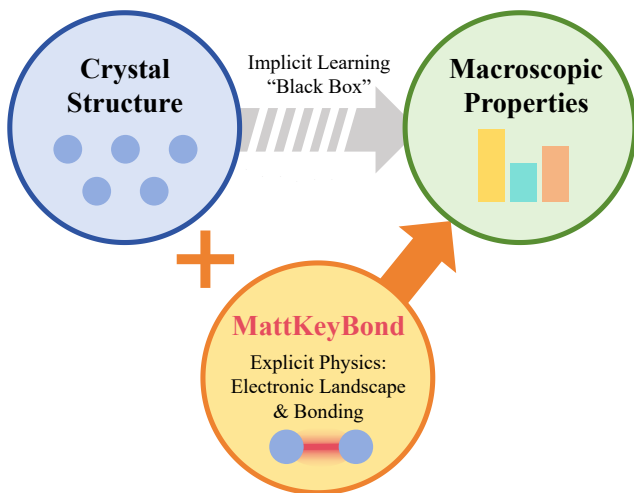


FIG. 1. MattKeyBond: Enhancing implicit learning with interpretable electronic and bonding descriptors.

parable bonding analysis across 36,377 inorganic compounds in the current release.

Building on this extensive dataset, we further propose **Bonding Attractivity (BA)**, a novel element-specific descriptor designed to quantify the intrinsic capability of atoms to form covalent networks. While traditional electronegativity and its modern extensions[23, 24] primarily characterize the tendency for charge transfer (ionicity), they do not fully capture the energetic contributions of shared electrons (covalency). BA complements this classical concept by explicitly measuring the strength of orbital hybridization that binds atoms together. Leveraging 3.6 million bond records, we have successfully parametrized BA for elements from hydrogen ($Z = 1$) to bismuth ($Z = 83$), quantifying its dependence on bond length and valence state. As a human-readable descriptor, this metric provides an intuitive way to understand the intrinsic bonding ability of elements in covalent systems.

Crucially, we position MattKeyBond and BA not merely as static data, but as physically interpretable intermediate features that connect atomic coordinates to macroscopic behavior. Unlike purely structural inputs, these descriptors add an explicit energy dimension derived from rigorous electronic structure theory. For the “AI for Science” community, this work represents a clear strategy: by providing precomputed quantum insights, we relieve machine-learning models of the burden of learning physical laws from scratch. MattKeyBond thus serves as a foundational resource, integrating electronic structure theory into modern AI workflows to accelerate bonding-guided materials discovery.

II. DESCRIPTOR OF CHEMICAL BOND

Fundamentally, material synthesizability is governed by energy minimization, a process intrinsically driven by the formation of chemical bonds. The release of energy serves as a quantitative measure of bond strength, typically described by metrics such as cohesive energy, bond energy, formation energy, or energy above the hull (depending on the different reference used). Consequently, beyond specific electronic or mechanical properties, a numerical descriptor of chemical bond strength offers a direct metric for energetic stability. Below, we briefly outline the formation of chemical bonds, their descriptors, and their relationship to bond energy.

As illustrated in Figure 2, consider isolated atoms A and B . Their atomic orbitals, denoted as $|a_{A/B}\rangle$, exhibit element-specific spherical harmonic distributions and energy levels (omitting orbital indices for simplicity). Taking a simple diatomic molecule as an example, as atoms A and B approach each other, their orbitals undergo overlap, renormalization, and hybridization under the influence of the local field, resulting in new molecular orbitals $|n\rangle$:

$$|n\rangle = C_A^n |\tilde{a}_A\rangle + C_B^n |\tilde{a}_B\rangle. \quad (1)$$

Here, $|\tilde{a}_{A/B}\rangle$ represents the renormalized atomic orbital in the local environment, and $C_{A/B}^n$ denotes the combination coefficients. Within the framework of density functional theory (DFT)[25, 26], this process is described by the redistribution of charge density and self-consistent field. In the single-particle representation, this redistribution is reflected in the reduced density matrix (RDM) of $\langle \tilde{a}_A | \hat{D} | \tilde{a}_B \rangle$. As illustrated in the third row of Fig. 2, along the bond between A and B , the RDM is reconstructed in two primary ways: one is the hybridization between atomic orbitals (orange arrows), which provides the cohesive force between the bonding atoms; the other is the charge transfer from one atom to another (blue arrows), resulting in their respective valence states x . These two aspects represent two major pathways for energy release during bond formation.

Within the DFT framework[25, 26], the total energy of a system can be decomposed into a band-structure term, a double counting correction to the Hartree and exchange-correlation energies, and the ion-ion Ewald energy:

$$E_{\text{tot}} = E_{\text{band}} - E_{\text{dc}} + E_{\text{Ewald}}. \quad (2)$$

The E_{Ewald} term arises from ion-ion Coulomb repulsion, which generally increases as bonding atoms approach each other. E_{dc} strongly depends on the charge density derived from the eigenstates $|n\rangle$ of E_{band} , and generally follows a trend similar to that of E_{band} due to the self-consistent nature of the calculation. Consequently, during chemical bond formation, the energy gain primarily originates from the first term, E_{band} , which sums the energies ε_n of all occupied single-particle eigenstates $|n\rangle$.

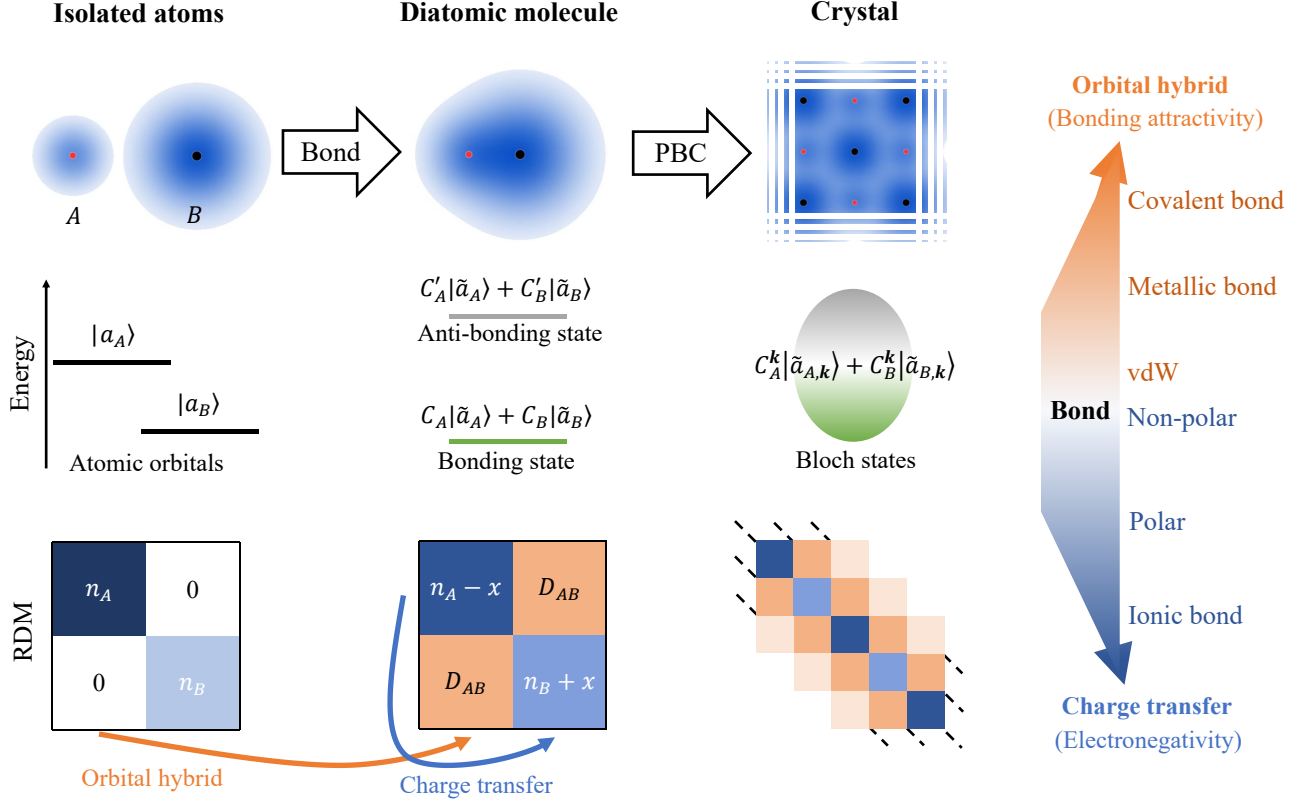


FIG. 2. The evolution of electronic structure from isolated atomic orbitals to hybridized molecular orbitals and finally to Bloch states in a periodic crystal. The reconstruction of the reduced density matrix (RDM) highlights two distinct energetic pathways driving bond formation: charge transfer (represented by blue arrows), and orbital hybridization (represented by orange arrows). The last column lists a conceptual classification of chemical bonds based on these two primary dimensions.

Within the subspace of the atomic-orbital basis, E_{band} can be further separated into an on-site term and an inter-site hybridization term:

$$E_{\text{band}} = E_{\text{onsite}} + E_{\text{hybrid}}, \quad (3)$$

which respectively correspond to the two major processes in bond formation: charge transfer and orbital hybridization.

Furthermore, the respective contributions from E_{onsite} and E_{hybrid} provide a basis for classifying chemical bonds. As illustrated in the last column of Fig. 2, based on the tendency for charge transfer, bonds can be described as non-polar (no transfer), polar (partial transfer), or ionic (full transfer). From the perspective of orbital hybridization strength, one can roughly distinguish between the weakest van der Waals (vdW) interactions, weak metallic bonds, and the strongest covalent bonds.

For any energy ε_n contributing to E_{band} , COHP analysis resolves it into pairwise contributions[20–22]:

$$\varepsilon_n = \sum_{A,B} \text{COHP}_{A,B}^n = \text{Re} \sum_{A,B} C_A^{m*} C_B^n \langle \tilde{a}_A | \hat{H} | \tilde{a}_B \rangle. \quad (4)$$

The sign of COHP_{AB}^n characterizes the nature of the interaction: $\text{COHP} < 0$ indicates a bonding state, while

$\text{COHP} > 0$ indicates an anti-bonding state between atoms A and B. Additionally, based on the magnetic quantum number m of $|a_{A/B}\rangle$ (aligning the A - B bond along the z -axis), they can hybridize into strong σ bonds ($m = 0$), intermediate π bonds ($m = \pm 1$), or weaker δ bonds ($m = \pm 2$).

Summing over all occupied states, the integral of COHP (ICOHP) yields:

$$\begin{aligned} E_{\text{onsite}} &= \sum_A \text{ICOHP}_{AA}, \\ E_{\text{hybrid}} &= \sum_{A \neq B} \text{ICOHP}_{AB}, \end{aligned} \quad (5)$$

which enables bond-resolved energetics for every inter-atomic pair (A, B) in complex structures. This energy integration also simplifies the calculation of ICOHP_{AB} :

$$\text{ICOHP}_{AB} = \langle \tilde{a}_B | \hat{D} | \tilde{a}_A \rangle \langle \tilde{a}_A | \hat{H} | \tilde{a}_B \rangle \quad (6)$$

Generally, $\text{ICOHP} < 0$ indicates a stabilizing interaction. For crystal systems (third column of Fig. 2) with periodic boundary conditions (PBC), the combination coefficients in Eq. 1 acquire a Bloch phase factor in the momentum- $\mathbf{k}$ representation. While discrete molecular orbital levels

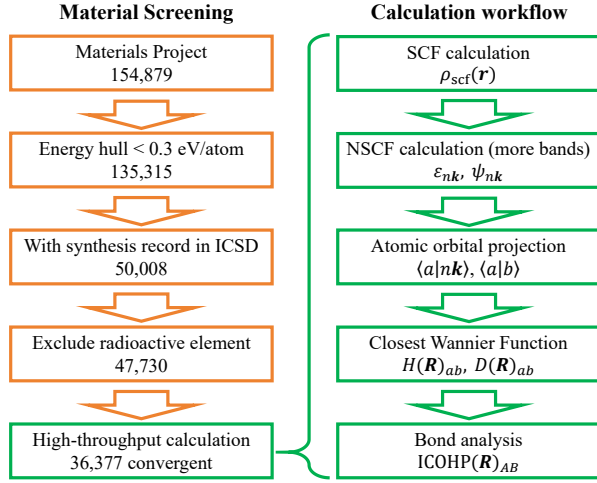


FIG. 3. Materials-screening criteria and the high-throughput calculation/analysis workflow.

evolve into continuous band structures, the physical basis of the chemical bond and the analytical tools of ICOHP remain identical to those used in molecular systems. See the Appendix B and C for more details regarding PBC and multi-orbital systems.

Although ICOHP does not account for all contributions to the total bond energy, it provides a direct measure of the energy associated with orbital hybridization between specific atomic pairs. Furthermore, this hybridization energy is intrinsically tied to inter-atomic binding forces and thus informs a range of material properties, including hardness, inter-atomic force constants (phonon stiffness), and, in metal systems, electron-phonon coupling strength, electronic transport, and, superconductivity[27]. Therefore, in this work, we adopt ICOHP as a primary descriptor of bond strength for the construction of the database.

III. HIGH-THROUGHPUT WORKFLOW

In this work, we first retrieved all crystal structures available in the MP database[1], comprising 154,879 entries as of October 2025. The detailed screening criteria and the computational workflow are summarized in Fig. 3. To ensure energetic stability, we retained compounds with an energy above the hull of less than 0.3 eV/atom, resulting in 135,315 stable or metastable structures. To further guarantee synthesizability, we restricted the set to compounds with experimental records in the Inorganic Crystal Structure Database (ICSD)[28], yielding 50,008 materials. We additionally excluded systems containing radioactive elements, resulting in 47,730 candidates for high-throughput calculations. At present, within the constraints of the allotted wall time and convergence criteria, 36,377 materials have been successfully computed and are included in this initial release.

For each material, the calculations and analyses proceed in five steps:

1. Self-consistent field (SCF) calculation. We performed DFT SCF calculations using the Quantum ESPRESSO (QE) package [29] to obtain the electronic ground state and charge density. Additional settings (functional[30], pseudopotentials[31], $\mathbf{k}$ -point meshes, smearing, and convergence thresholds) are provided in Appendix A.

2. Non-self-consistent field (NSCF) calculation. For subsequent Wannier downfolding calculations, we carried out NSCF calculations with an increased number of bands. The total number of bands was empirically set to

$$\text{nbnd} = N_{\text{atomorb}} + 5N_{\text{atom}}.$$

Here, N_{atomorb} is the total number of atomic valence orbitals included in the projection and N_{atom} is the number of atoms in the primitive cell (see TABLE I in the Appendix). Both SCF and NSCF steps were executed with `pw.x` in QE[29].

3. Atomic-orbital projections. Using the `projwfc.x` module of QE[29], we computed (i) the projection matrix between atomic orbitals and Kohn-Sham eigenstates, $\langle a | n \rangle$, and (ii) the overlap integrals between neighboring atomic orbitals, $\langle a | b \rangle$. Here $|a\rangle$ and $|b\rangle$ denote atomic-like orbitals, and $|n\rangle$ denotes a Kohn-Sham eigenstate obtained from the DFT calculations above.

4. Closest Wannier Function (CWF) downfolding. To capture the orbital renormalization in the crystal field, we constructed CWFs using atomic orbitals as guiding functions to downfold the plane-wave Kohn-Sham subspace into a compact Wannier basis[19]. This procedure yields a CWF-based tight-binding Hamiltonian and reduced density matrix (RDM) for each system. Algorithmic details and numerical settings are provided in Appendix B.

5. ICOHP-based bond analysis. We analyzed bonding using the crystal orbital Hamilton population (COHP) and its energy integral (ICOHP) for all inter-atomic pairs with real-space distances shorter than 6 Å. The resulting bond-resolved energetics were used to quantify bond strength across all distinct bonded pairs[20–22]. Additional details are provided in Appendix C.

To demonstrate the microscopic insights provided by MattKeyBond, we present a representative analysis of the nearest carbon-carbon bond in graphene in Figure 4. As shown in panel (d), the band structure obtained from our CWF downfolding (purple circles) accurately reproduces the DFT band structure (gray lines) around the Fermi level. Beyond global electronic structure, MattKeyBond provides the local inter-atomic Hamiltonian in panel (b) and the RDM in panel (c) for every orbital pair. These matrices allow for advanced post-processing, such as decomposing the total interaction into independent bonding channels via the singular value decomposition (SVD) of the inter-atomic RDM block. As shown

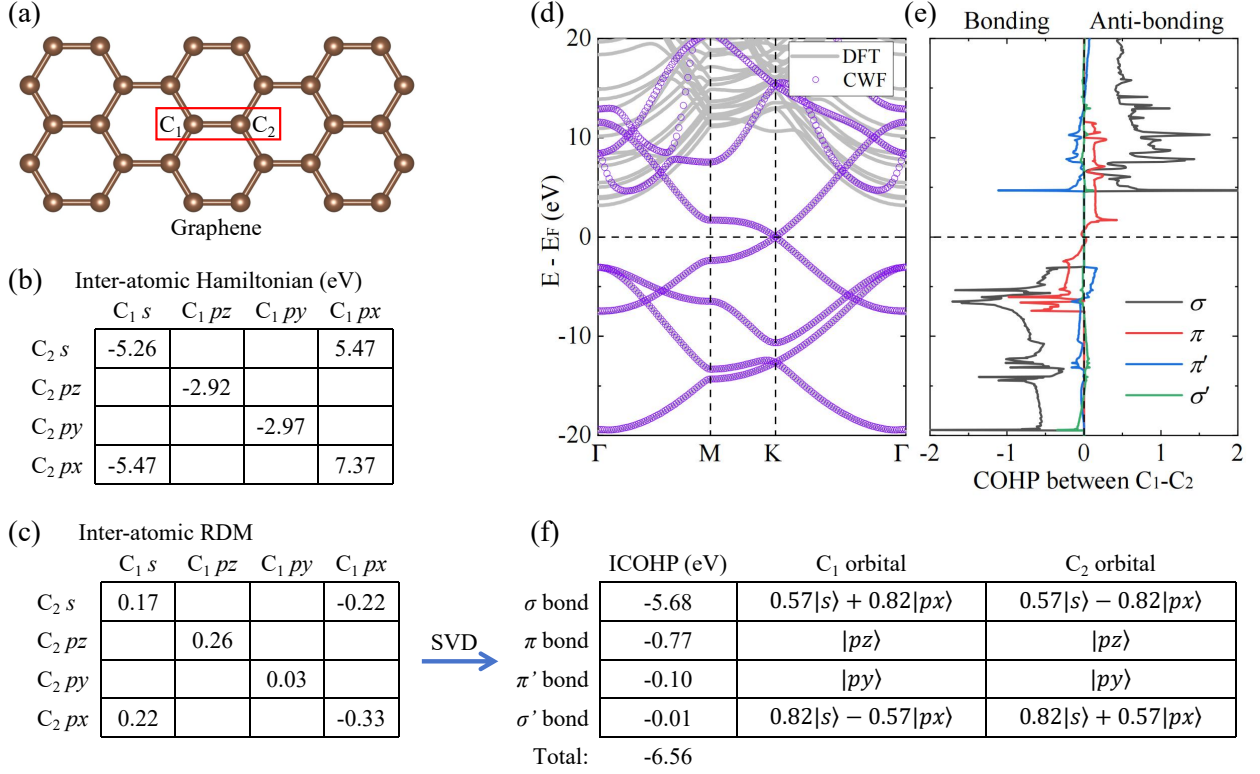


FIG. 4. Detailed bonding analysis of the nearest C-C bond in graphene as a representative example. (a) Crystal structure of graphene with the target C_1 - C_2 atomic pair highlighted. (b) The real-space inter-atomic Hamiltonian matrix and (c) RDM for the C_1 - C_2 pair in the atomic orbital basis (s, p_z, p_y, p_x). (d) Comparison between the electronic band structure calculated by DFT (gray lines) and that obtained from our CWF model (purple circles). (e) Bond-resolved COHP curves in energy space, clearly distinguishing the contributions and strengths from σ (black), π (red), π' (blue), and σ' (green). (f) Quantitative decomposition of the bond strength via the SVD of RDM.

in panel (f), this analysis can automatically identify the strong σ bond (ICOHP = -5.68 eV) formed by sp^2 -like hybrids ($\sqrt{1/3}|s\rangle \pm \sqrt{2/3}|p_x\rangle$) and the characteristic π bond (ICOHP = -0.77 eV) formed by vertical p_z orbitals. The remaining two are the in-plane π' bond formed by p_y orbitals and the σ' bond resulting from the tail-to-tail overlap of $|s\rangle$ and $|p_x\rangle$ orbitals. Together with their COHP distributions in energy space shown in panel (e), these analyses allow users to go beyond total bond energies and dissect the specific orbital mechanisms driving material properties (e.g. high- T_c superconductivity in MgB_2 , driven by metallized B-B σ bonds [27]).

IV. BONDING ATTRACTIVITY

Element-specific bonding trends are traditionally summarized by electronegativity (EN)[23, 24], such as that represented by the Pauling scale, which was inferred primarily from experimental bond energies across compounds. EN captures an element's tendency to attract electrons. Electrons tend to transfer from elements with low EN to those with high EN. Inspired by the definition of Pauling EN, using the 3,665,789 bond records

(ICOHP values) currently available across 36,377 materials in MattKeyBond, we introduce Bonding Attractivity (BA), denoted as $\eta_A(R, x_A)$. Unlike EN, which primarily characterizes the tendency for charge transfer (ionicity), BA complements this classical concept by quantifying the intrinsic capability of an atom to bind with neighbors via covalent-like orbital hybridization (covalency). As discussed previously, orbital hybridization plays a decisive role in many physical properties, including inter-atomic force constants, atomic dimerization, and even electron-phonon coupling strength, which can influence electronic transport and superconductivity. Therefore, in addition to providing new insights into structure-bonding relationships, the introduction of BA also enables interpretable property prediction and supports data-driven discovery of functional materials.

In complex crystal environments, bond strength is influenced by numerous factors, including element identity, bond type, hybridized orbital character, valence state, bond length, and crystal-field effects. To obtain a simple, human-readable, yet tractable descriptor, we neglect the dependence on orbital character and local crystal field but retain bond length R and valence state x_A as two dominant variables controlling the BA of atom A .

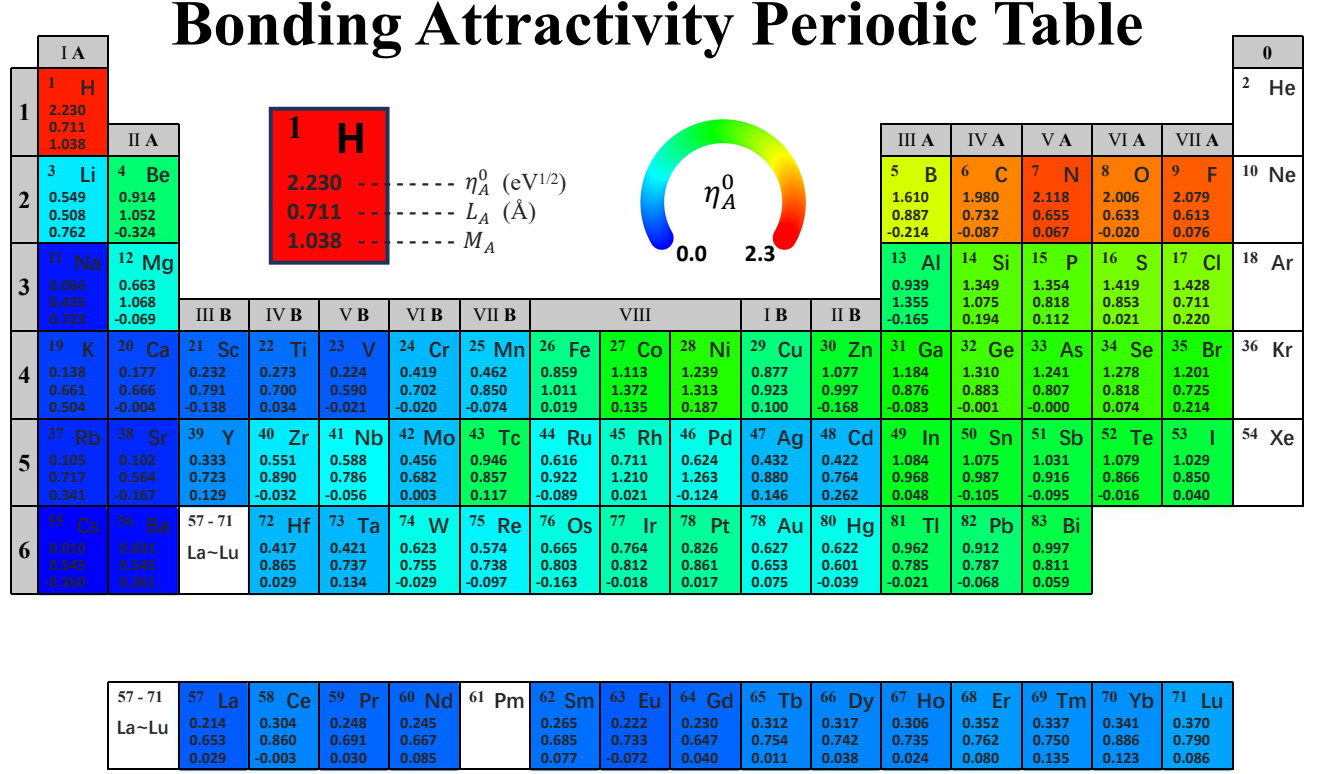


FIG. 5. Periodic table of Bonding Attractivity. The three numbers in each element block are, respectively, the intrinsic bonding attractivity η_A^0 , the characteristic decay length L_A , and the valence-state modulation factor M_A . The background color encodes η_A^0 .

Analogous to Pauling's treatment of bond energies[23], we postulate that the ICOHP between two bonded atoms A and B can be expressed as the product of their respective BAs:

$$-\text{ICOHP}_{AB}(R, x_A, x_B) = \eta_A(R, x_A) \eta_B(R, x_B). \quad (7)$$

Here $\eta(R, x) > 0$ has units of eV^{1/2}. To capture bond-length decay and valence-state modulation, we adopt the exponential form

$$\eta_A(R, x_A) = \eta_A^0 \exp[-(R - 2r_A)/L_A + M_A x_A], \quad (8)$$

with the reference value

$$\eta_A^0 = \eta_A(2r_A, 0). \quad (9)$$

Here η_A^0 is an element-specific baseline BA; r_A is the covalent radius of element A [32], serving as a reference for typical bond lengths; L_A (in Å) is the characteristic decay length of BA with respect to R ; and M_A (dimensionless) quantifies the modulation by the valence state. Consequently, the triplet (η_A^0, L_A, M_A) fully parameterizes the BA for element A . As implied by Eq. (9), $(\eta_A^0)^2$ corresponds to the $-\text{ICOHP}$ of an A - A bond at $R = 2r_A$ and $x_A = 0$, so we treat η_A^0 as the fundamental BA quantity for element A .

Using 3,665,789 bond records (ICOHP values) spanning diverse element combinations, bond lengths, and

valence states collected from 36,377 materials in MatKeyBond, we performed a least-squares fit of Eqs. (7) and (8). The quality of the fit is shown in Figs. 7~10 of the Appendix. Fitted parameters from H ($Z = 1$) to Bi ($Z = 83$) are summarized in the periodic-table map of Fig. 5. In each element block, the three numbers report η_A^0 , L_A , and M_A , and the background color encodes η_A^0 from 0.0 (blue) to 2.3 (red).

Overall, the distribution of η_A^0 shows broad similarity to classic Pauling EN [23], with the following trends:

1. Elements with valence $2p$ orbitals (B~F) exhibit high η_A^0 , consistent with their propensity to form strong covalent bonds.
2. Elements on the left side of the table: alkali metals, alkaline-earth metals, lanthanides, and many group III~VII B elements show relatively low η_A^0 , aligning with their common cationic roles in compounds.
3. Heavier p -block elements ($n > 2$) and numerous late transition metals generally exhibit intermediate η_A^0 values.

A notable difference from the Pauling electronegativity scale is that the largest η_A^0 is obtained for hydrogen, followed by typical $2p$ elements. In contrast, fluorine has the highest EN on the Pauling scale. This

discrepancy reflects the distinct physics captured: BA measures orbital-hybridization strength, whereas EN includes more contributions from charge-transfer energetics. Fluorine’s large EN mainly arises from the charge-transfer term rather than hybridization energy, favoring ionic bonding with comparatively weaker covalent binding. Hydrogen, conversely, readily forms strong bonds via hybridization across diverse chemical environments, a feature consistent with its behavior in hydrogen storage materials.

In addition to the baseline η_A^0 , Eq. (8) introduces bond-length and valence-state dependencies through L_A and M_A (see Fig. 6 for element-resolved maps). The decay length L_A governs how rapidly BA weakens with increasing R . A small L_A manifests in two typical ways: it indicates either a tendency to form short, stiff bonds with large force constants (characteristic of $2p$ elements like C, N, O, and F), or a steep rise in bonding attractivity as atoms approach each other (seen in reactive alkali, alkaline-earth, and halogen elements). In both cases, hybridization-driven bonding is favored primarily with nearest neighbors. Conversely, as shown in Fig. 6, only five elements (Al, Co, Ni, Rh, and Pd) exhibit conspicuously large L_A values, suggesting that their chemical bonding capabilities remain relatively stable across varying local environments.

The valence modulation factor M_A displays a sign oscillation across the periodic table. In Fig. 6, the sign of M_A is encoded by the background color (red for positive, blue for negative). A positive M_A implies that removing electrons (higher oxidation state) enhances BA, whereas a negative M_A indicates that adding electrons strengthens BA. Classical EN arguments suggest that gaining electrons should reduce an atom’s electron-attracting tendency, pointing toward positive M_A . This behavior indeed appears for many simple alkali and halogen elements with a single half-filled valence orbital. However, for elements in the middle of the table with multi-orbital valence manifolds, adding an electron may create a new unpaired or half-filled orbital and thereby enhance the hybridization with other atoms, yielding negative M_A . Consequently, depending on detailed valence-electron counts and orbital occupations, M_A exhibits oscillatory trends within a period.

We emphasize that the current calculations and analyses are performed exclusively within a nonmagnetic framework. For systems with robust magnetic order or high Neel/Curie temperatures, the present BA formulation may yield suboptimal descriptions of bonding energetics. Furthermore, in addition to magnetism, the local environment and specific orbital characters can play decisive roles in ICOHP. For elements with a single dominant valence orbital (e.g., many chalcogens and halogens), Eq. 7 captures ICOHP accurately. However, for elements with multiple bonding orbitals (e.g., boron), stronger environmental and orbital dependencies introduce additional complexity, as evidenced by the parity plots comparing BA-fitted and DFT-calculated $-ICOHP$

values in Figs. 7~10 of the Appendix. Despite these limitations, our simplified BA formulation successfully captures the essential trends of atomic bonding across the periodic table. This validates BA as a compact, intuitive, and interpretable descriptor that compresses complex electronic structure information into a human-readable format.

V. SUMMARY

In summary, we have developed MattKeyBond, a bond-centric materials database constructed through high-throughput first-principles calculations, Closest Wannier Function (CWF) downfolding, and integral crystal orbital Hamilton population (ICOHP) analysis. Currently encompassing 36,377 inorganic compounds and over 3.6 million bond records, this database provides a high-fidelity, real-space mapping of local electronic landscapes and bonding interactions. By moving beyond traditional geometric coordinates and global scalar properties, MattKeyBond extracts explicit, energy-resolved intermediate features at the atom-bond resolution, establishing a mechanistic bridge between atomic arrangements and macroscopic material properties. The dataset presented in this paper is openly available at [ScienceDB](#).

Building upon this extensive dataset, we introduced Bonding Attractivity (BA), denoted as $\eta_A(R, x_A)$, a novel element-specific descriptor parameterized by a baseline attractivity, a characteristic decay length, and a valence-state modulation factor. While classical electronegativity primarily characterizes the tendency for ionic charge transfer, BA serves as a crucial complementary metric that quantifies the intrinsic capability of atoms to form bonds via covalent orbital hybridization. By distilling complex quantum-mechanical information into an intuitive, human-readable format, BA provides a tractable and physically interpretable metric for bond strength across diverse chemical environments.

The introduction of MattKeyBond and BA represents a significant advancement for the “AI for Science” community. These precomputed, physics-based descriptors provide machine-learning models with explicit electronic structure information, relieving them of the burden of implicitly reconstructing complex quantum-mechanical relationships from geometric information alone. This integration fundamentally enhances model interpretability and generalizability, particularly in scenarios where experimental training data are scarce, thereby enabling more robust and accurate property predictions. Furthermore, as generative AI and foundation models continue to evolve, embedding these physics-rich, bond-resolved descriptors into modern workflows holds immense potential for inverse materials design.

Moving forward, we plan to continuously expand MattKeyBond by incorporating newly predicted and potentially synthesizable crystal structures. Additionally, future iterations of the database will explicitly account

for advanced physical interactions, including spin-orbit coupling and magnetism, to provide an even more comprehensive and rigorous description of chemical bonding. We anticipate that this foundational resource will significantly accelerate the data-driven discovery and mechanistic understanding of next-generation functional materials, such as unconventional superconductors, advanced catalysts, and novel energy storage systems.

ACKNOWLEDGMENTS

This work was supported by a project funded by the China Postdoctoral Science Foundation (No. 2022M723355), Chinese funding administered through HPSTAR, the National Natural Science Foundation of China (12488201), and the National Key Research and Development Project of China (2021ZD0301800, 2022YFA1403103).

-
- [1] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, *APL Materials* **1**, 011002 (2013).
 - [2] S. M. Miao Liu, Atomly.net materials database and its application in inorganic chemistry, *SCIENTIA SINICA Chimica* **53**, 19 (2023).
 - [3] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd), *JOM* **65**, 1501 (2013).
 - [4] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies, *npj Computational Materials* **1**, 15010 (2015).
 - [5] H. Eckert, S. Divilov, M. J. Mehl, D. Hicks, A. C. Zettl, M. Esters, X. Campilongo, and S. Curtarolo, The aflow library of crystallographic prototypes: Part 4, *Computational Materials Science* **240**, 112988 (2024).
 - [6] Nomad: A distributed web-based platform for managing materials science research data, *Journal of Open Source Software* **8**, 5388 (2023).
 - [7] S. Haastруп, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. Jørgen Mortensen, T. Olsen, and K. S. Thygesen, The computational 2d materials database: high-throughput modeling and discovery of atomically thin crystals, *2D Materials* **5**, 042002 (2018).
 - [8] M. N. Gjerding, A. Taghizadeh, A. Rasmussen, S. Ali, F. Bertoldo, T. Deilmann, N. R. Knøsgaard, M. Kruse, A. H. Larsen, S. Manti, T. G. Pedersen, U. Petralanda, T. Skovhus, M. K. Svendsen, J. J. Mortensen, T. Olsen, and K. S. Thygesen, Recent progress of the computational 2d materials database (c2db), *2D Materials* **8**, 044002 (2021).
 - [9] K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Bicchi, A. R. Hight Walker, Z. Trautt, J. Hatrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe, and F. Tavazza, The joint automated repository for various integrated simulations (jarvis) for data-driven materials design, *npj Computational Materials* **6**, 173 (2020).
 - [10] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Hanczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr, Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: Core mof 2019, *Journal of Chemical & Engineering Data* **64**, 5985 (2019), <https://doi.org/10.1021/acs.jced.9b00835>.
 - [11] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, Open catalyst 2020 (oc20) dataset and community challenges, *ACS Catalysis* **11**, 6059 (2021), <https://doi.org/10.1021/acscatal.0c04525>.
 - [12] R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi, and C. L. Zitnick, The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts, *ACS Catalysis* **13**, 3066 (2023), <https://doi.org/10.1021/acscatal.2c05426>.
 - [13] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* **120**, 145301 (2018).
 - [14] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chemistry of Materials* **31**, 3564 (2019).
 - [15] K. Choudhary and B. DeCost, Atomistic line graph neural network for improved materials property predictions, *npj Computational Materials* **7**, 185 (2021).
 - [16] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nature Computational Science* **2**, 718 (2022).
 - [17] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation, *Nature Computational Science* **2**, 367 (2022).
 - [18] T. Lv, Z. Zhong, Y. Liang, F. Li, J. Huang, and R. Zheng, Deep charge: Deep learning model of electron density from a one-shot density functional theory calculation, *Phys. Rev. B* **108**, 235159 (2023).
 - [19] T. Ozaki, Closest wannier functions to a given set of localized orbitals, *Phys. Rev. B* **110**, 125115 (2024).
 - [20] R. Dronskowski and P. E. Blochl, Crystal orbital hamilton populations (cohpp): energy-resolved visualization of chemical bonding in solids based on density-functional calculations, *The Journal of Physical Chemistry* **97**, 8617

- (1993).
- [21] V. L. Deringer, A. L. Tchougréeff, and R. Dronskowski, Crystal orbital hamilton population (cohp) analysis as projected from plane-wave basis sets, *J. Phys. Chem. A* **115**, 5461 (2011).
 - [22] S. Maintz, V. L. Deringer, A. L. Tchougréeff, and R. Dronskowski, Analytic projection from plane-wave and paw wavefunctions and application to chemical-bonding analysis in solids, *J. Comput. Chem.* **34**, 2557 (2013).
 - [23] L. Pauling, The nature of the chemical bond. application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules, *Journal of the American Chemical Society* **53**, 1367 (1931).
 - [24] M. Rahm, T. Zeng, and R. Hoffmann, Electronegativity seen as the ground-state average valence electron binding energy, *Journal of the American Chemical Society* **141**, 342 (2019).
 - [25] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* **136**, B864 (1964).
 - [26] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133 (1965).
 - [27] J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, and J. Akimitsu, Superconductivity at 39 K in magnesium diboride, *Nature* **410**, 63 (2001).
 - [28] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme, Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features, *J. Appl. Cryst.* **52**, 918 (2019).
 - [29] QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
 - [30] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
 - [31] A. Dal Corso, Pseudopotentials periodic table: From h to pu, *Computational Materials Science* **95**, 337 (2014).
 - [32] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, Covalent radii revisited, *Dalton Trans.* , 2832 (2008).
 - [33] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, An updated version of wannier90: A tool for obtaining maximally-localised wannier functions, *Comput. Phys. Commun.* **185**, 2309 (2014).

Appendix A: First-principles calculation

The electronic structures and atomic-orbital projections of more than 50,000 materials were investigated in a high-throughput manner based on the density functional theory (DFT) [25, 26] calculations as implemented in the Quantum ESPRESSO (QE) package [29]. The interactions between electrons and nuclei were described by ultrasoft pseudopotentials as implemented in the PSLIBRARY [31]. The generalized gradient approximation (GGA) of Perdew-Burke-Ernzerhof (PBE) [30] type was adopted for the exchange-correlation functional. For elements from Z=1 (H) to Z=83 (Bi) present in our dataset,

their valence orbital configurations and the number of atomic orbitals used in the Wannier downfolding are listed in TABLE I. The wave-function and charge-density cutoffs for the plane-wave basis were set to 1.3 times the suggested values by PSLIBRARY. A Gamma-centered $\mathbf{k}$ -point mesh with a grid spacing of $0.2 \text{ \AA}^{-1}$ was adopted for the Brillouin zone (BZ) sampling. The Gaussian smearing method with a width of 0.004 Ry was employed for the Fermi surface broadening. The lattice constants and internal atomic positions were taken directly from the Materials Project database. All calculations were performed within a nonmagnetic framework.

Appendix B: Closest Wannier Function

In plane-wave-based DFT calculations, the atomic orbitals used for post-processing and projection analysis are typically incomplete and non-orthogonal:

$$\sum_{ab} \langle n | a \rangle S_{ab}^{-1} \langle b | n \rangle \neq 1 \quad (\text{B1})$$

$$S_{ab} = \langle a | b \rangle \neq \delta_{ab}.$$

Here $|n\rangle$ is the calculated Kohn–Sham eigenstate in a plane-wave basis. $|a\rangle$ and $|b\rangle$ are two atomic orbitals from two neighboring atoms, and S_{ab} is their orbital overlap integral. $\hat{S}^{-1}$ is the inverse of the overlap matrix $\hat{S}$. Notably, these incomplete and non-orthogonal atomic orbitals pose significant challenges for bonding analysis. Although several strategies have been proposed to address this issue in orthogonal bases (like Lowdin orthogonalization or directly cutting the orbital wave-function in real-space), the incomplete atomic-orbital basis still cannot adequately describe all electronic states of interest, especially for electronic states around or above the Fermi level. An ideal strategy is to construct a set of complete and orthogonal Wannier Functions. Nowadays, the most widely used orbital-downfolding method is based on maximally localized Wannier functions (MLWFs)[33]. However, the construction of MLWFs is highly sensitive to parameter choices. It inevitably involves manual parameter adjustment and is therefore not well suited for high-throughput calculations. In this work, for constructing a set of ideal Wannier functions to represent the original atomic orbital, we adopted a recently developed Closest Wannier Function (CWF) method[19]. CWFs are characterized by weak parameter dependence and require no human intervention. Unlike MLWF construction, which requires iterative calculations to optimize an artificial spread function Ω , CWFs are obtained through a single-step singular value decomposition of the projection matrix $\hat{A}$:

$$\hat{A} = \hat{U} \hat{\Sigma} \hat{V}^\dagger. \quad (\text{B2})$$

Here $\hat{A}$ is a non-square matrix whose elements A_{an} reflect the projection amplitudes between atomic orbital $|a\rangle$ and

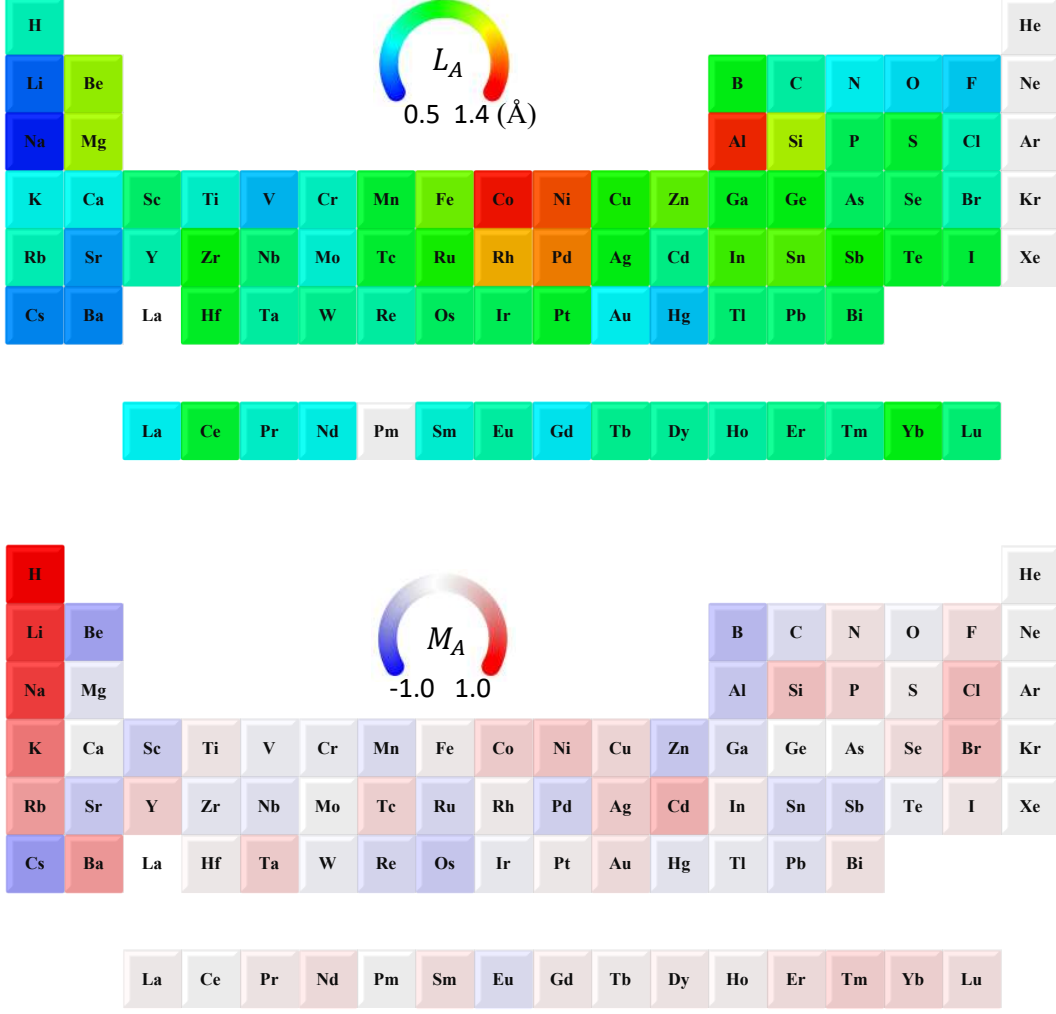


FIG. 6. Visualization of the characteristic decay length L_A and the valence-state modulation factor M_A for Bonding Attractivity.

the eigen- Kohn-Sham state $|n\rangle$ from DFT calculation. In general, $\hat{A}$ is not unitary. But we can further define a semi-unitary matrix $\hat{B}$ using:

$$\hat{B} = \hat{U}\hat{V}^\dagger. \quad (\text{B3})$$

Utilizing the half-unitarity of $\hat{B}$, we can therefore construct a set of orthogonal and complete Wannier function:

$$|a_{\text{CWF}}\rangle = \sum_n |n\rangle B_{an}. \quad (\text{B4})$$

As proved in Ref. [19], this set of $|a_{\text{CWF}}\rangle$ has the minimal deviation from the original guiding functions defined by $\hat{A}$, and is therefore called the set of Closest WFs.

The only parameter dependence of the CWF construction lies in the definition of the guiding matrix $\hat{A}$. In

earlier works, this was defined by the atomic orbital projection matrix $\langle a | n \rangle$ with a weight of w_n :

$$A_{an} = \langle a | n \rangle w_n. \quad (\text{B5})$$

The role of w_n is to emphasize the energy window of interest, similar to the frozen windows used in MLWF. Here we let

$$w_n = \frac{1}{e^{\frac{\varepsilon_n - \varepsilon_f - \Delta}{\Delta}} + 1} + o \quad (\text{B6})$$

to ensure that all occupied states can be well described by CWFs. Δ is a broadening parameter to control the energy window above the Fermi level. o is a small finite quantity introduced to ensure that all singular values of $\hat{A}$ are non-zero. In this work, we set $\Delta = 2.0$ eV and $o = 0.01$.

In addition to w_n , in the definition of guiding matrix $\hat{A}$, we also introduced the dual basis. Because of the nonorthogonality of the original set of atomic orbitals $|a\rangle$, we can define a set of corresponding dual basis functions, $|a'\rangle$:

$$|a'\rangle = \sum_b |b\rangle S_{ba}^{-1}. \quad (\text{B7})$$

In mathematical terms, the dual orbital $|a'\rangle$ describes the unique dual vector associated with $|a\rangle$ and is therefore orthogonal to the other atomic orbitals:

$$\langle a' | b \rangle = \delta_{ab}. \quad (\text{B8})$$

Here we further introduced two projection matrices $\hat{C}$ and $\hat{D}$ for the original atomic orbital $|a\rangle$ and its dual orbital $|a'\rangle$, respectively:

$$\begin{aligned} C_{an} &= \langle a | n \rangle \\ D_{an} &= \langle a' | n \rangle. \end{aligned} \quad (\text{B9})$$

In practice, the types and number of orbitals included in the CWF downfolding can significantly influence the bonding analysis. Typically, certain outer-shell orbitals contribute minimally to atomic bonding because they are nearly unoccupied, such as the Mg $3p$ orbitals or the Ti $4p$ orbitals. However, due to their spatial extension, they can easily overlap with other atomic orbitals and capture electronic states that do not physically belong to them, like the core electronic state. To avoid undue dependence on these insignificant outer-shell orbitals, we used a composite strategy when constructing the $\hat{A}$ matrix:

$$A_{an} = [f_n D_{an} + (1 - f_n) C_{an}] w_n. \quad (\text{B10})$$

Here f_n is the Fermi distribution of the eigenstate $|n\rangle$. As shown by this expression, we used the projection matrix of the dual orbital, D_{an} , for occupied states. This can effectively reduce the influence of outer-shell orbitals on the bond analysis of occupied states.

In the MLWF downfolding calculation[33], one needs to optimize a local indicator Ω . This involves the overlap calculation between the wave-functions of different $\mathbf{k}$ points: $\langle m\mathbf{k} | e^{-i\mathbf{b}\cdot\mathbf{r}} | n\mathbf{k} + \mathbf{b} \rangle$. But for CWFs, the calculations at different $\mathbf{k}$ -points are fully independent. We just need to calculate the CWF transition matrix $\hat{B}(\mathbf{k})$ at each $\mathbf{k}$ -point. The final real-space tight-binding model and RDM in the CWF basis are obtained by Fourier transformation of $H(\mathbf{k})_{ab}$ and $D(\mathbf{k})_{ab}$, respectively:

$$\begin{aligned} H(\mathbf{R})_{ab} &= \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} H(\mathbf{k})_{ab} \\ H(\mathbf{k})_{ab} &= \sum_n \varepsilon_{n\mathbf{k}} B_{an}(\mathbf{k}) B_{bn}^*(\mathbf{k}) \end{aligned} \quad (\text{B11})$$

$$\begin{aligned} D(\mathbf{R})_{ab} &= \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} D(\mathbf{k})_{ab} \\ D(\mathbf{k})_{ab} &= \sum_n f_{n\mathbf{k}} B_{an}(\mathbf{k}) B_{bn}^*(\mathbf{k}) \end{aligned} \quad (\text{B12})$$

Notably, because CWFs are constructed to be as close as possible to the guiding functions, they are not necessarily well localized at the same time like the MLWF. Thus, within a finite $\mathbf{k}$ -mesh or under Born-von Karman (BvK) boundary conditions, sometimes the long-distance orbital hopping cannot fully decay to zero. Usually, this may slightly influence the effectiveness of Wannier interpolation. However, for the short-distance bonds between adjacent atoms of interest, their orbital hybridization can still be well described.

Appendix C: Bond analysis of ICOHP

In this work, we employed the CWF-based integrated crystal orbital Hamilton population (ICOHP) method to quantitatively describe atomic bond strength [20–22]. Crystal Orbital Hamilton Population (COHP) is a computational tool used in quantum chemistry and solid-state physics to analyze the electronic structure of molecules and solids. Within the DFT framework, ICOHP represents the contribution of orbital hybridization and helps elucidate the covalent character of chemical bonds in materials.

After CWF downfolding, we can rewrite the electronic eigenwavefunctions $|n\mathbf{k}\rangle$ in the CWF basis:

$$|n\mathbf{k}\rangle = \sum_{Aa} C_{Aa}^{n\mathbf{k}} |Aa\mathbf{k}\rangle, \quad (\text{C1})$$

Here $|Aa\mathbf{k}\rangle$ denotes the CWF of atomic orbital a on atom A in $\mathbf{k}$ -space, and $C_{Aa}^{n\mathbf{k}}$ is the corresponding combination coefficient. The band structure $\varepsilon_{n\mathbf{k}}$ can therefore be written as:

$$\begin{aligned} \varepsilon_{n\mathbf{k}} &= \langle n\mathbf{k} | \hat{H} | n\mathbf{k} \rangle = \sum_{Aa, Bb} C_{Bb}^{n\mathbf{k}*} C_{Aa}^{n\mathbf{k}} \langle Bb\mathbf{k} | \hat{H} | Aa\mathbf{k} \rangle \\ &= \sum_{\mathbf{R}Aa, \mathbf{R}'Bb} C_{Bb}^{n\mathbf{k}*} C_{Aa}^{n\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}'-\mathbf{R})} \langle \mathbf{R}'Bb | \hat{H} | \mathbf{R}Aa \rangle. \end{aligned} \quad (\text{C2})$$

Here we employed a Fourier transform $|Aa\mathbf{k}\rangle = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}Aa\rangle$ for the bond analysis between specific atomic pairs in real-space. In the summation of Eq. C2, the contribution from the orbital pair $|\mathbf{R}Aa\rangle$ and $|\mathbf{R}'Bb\rangle$ provides the corresponding COHP in the $|n\mathbf{k}\rangle$ state:

$$\text{COHP}_{\mathbf{R}Aa, \mathbf{R}'Bb}^{n\mathbf{k}} = C_{Bb}^{n\mathbf{k}*} C_{Aa}^{n\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}'-\mathbf{R})} \langle \mathbf{R}'Bb | \hat{H} | \mathbf{R}Aa \rangle. \quad (\text{C3})$$

In general, one is often more concerned with its distribution in energy space:

$$\text{COHP}_{\mathbf{R}Aa, \mathbf{R}'Bb}(\varepsilon) = \sum_{n\mathbf{k}} \delta(\varepsilon - \varepsilon_{n\mathbf{k}}) \text{COHP}_{\mathbf{R}Aa, \mathbf{R}'Bb}^{n\mathbf{k}} \quad (\text{C4})$$

The ICOHP between the orbitals $|\mathbf{R}Aa\rangle$ and $|\mathbf{R}'Bb\rangle$ is defined by integrating $\text{COHP}_{\mathbf{R}Aa, \mathbf{R}'Bb}(\varepsilon)$ below the

Fermi level:

$$\begin{aligned} \text{ICOHP}_{\mathbf{R}Aa,\mathbf{R}'Bb} &= \int_{-\infty}^{\varepsilon_f} d\varepsilon \text{COHP}_{\mathbf{R}Aa,\mathbf{R}'Bb}(\varepsilon) \\ &= \langle \mathbf{R}Aa | \hat{D} | \mathbf{R}'Bb \rangle \langle \mathbf{R}'Bb | \hat{H} | \mathbf{R}Aa \rangle \end{aligned} \quad (\text{C5})$$

Here we introduce the reduced density matrix (RDM) operator $\hat{D}$:

$$\hat{D} = \sum_{n\mathbf{k}} |n\mathbf{k}\rangle f_{n\mathbf{k}} \langle n\mathbf{k}|, \quad (\text{C6})$$

whose off-diagonal terms also carry information about bond order. $f_{n\mathbf{k}}$ is the Fermi–Dirac occupation with a finite-temperature broadening of 0.2 eV. The total ICOHP for the atomic pair $(\mathbf{R}A, \mathbf{R}'B)$ requires an additional summation of their hybrid orbitals:

$$\text{ICOHP}_{\mathbf{R}A,\mathbf{R}'B} = \sum_{a \in A, b \in B} \text{ICOHP}_{\mathbf{R}Aa,\mathbf{R}'Bb} \quad (\text{C7})$$

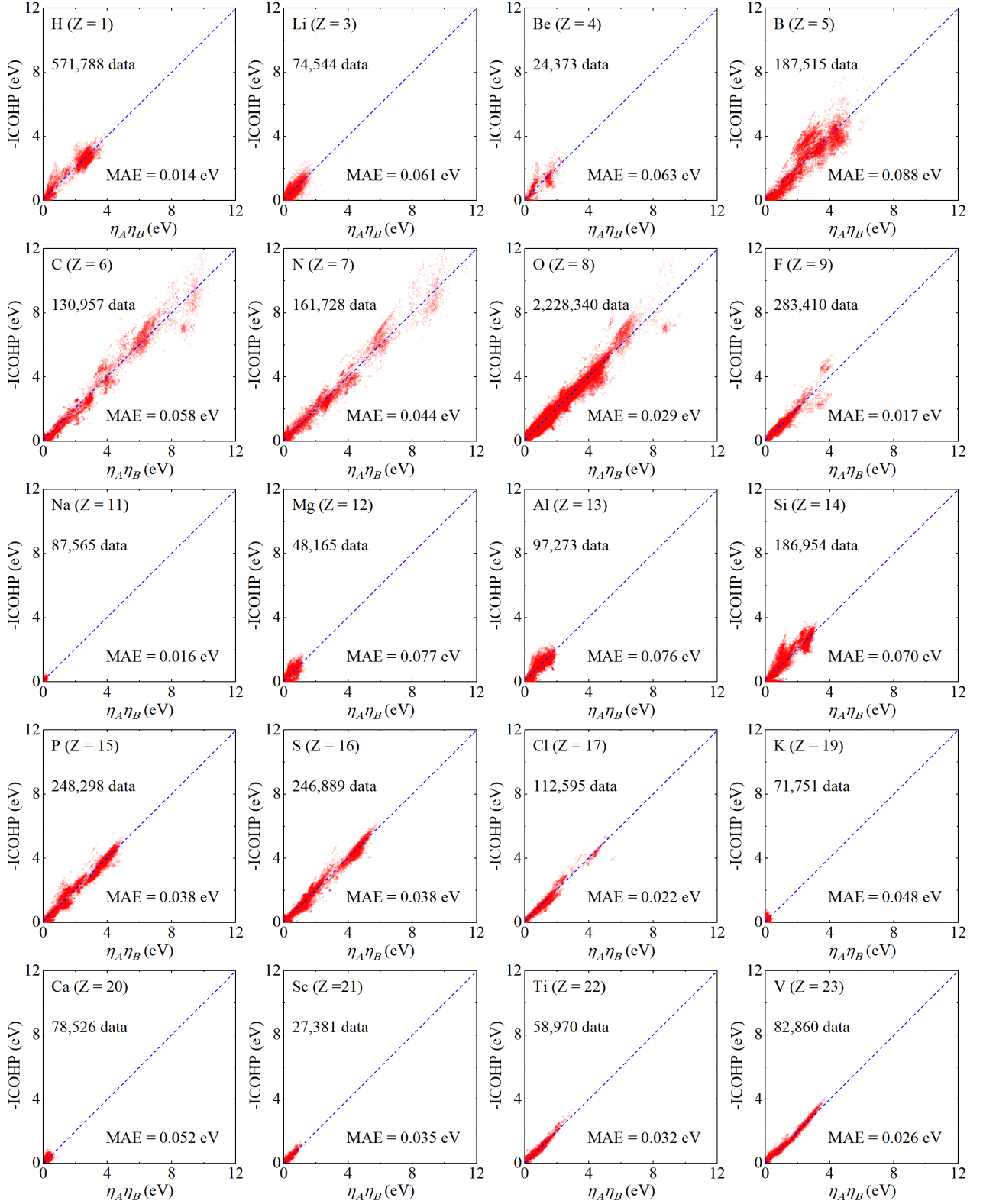


FIG. 7. Parity plots of BA-predicted ICOHP versus DFT-calculated ICOHP, resolved by the central element, from $Z = 1$ to $Z = 23$. Each panel aggregates all $A - B$ bonds that include the labeled element, with the diagonal indicating perfect agreement. Numbers inside each panel denote the sample size used for that element's bonds.

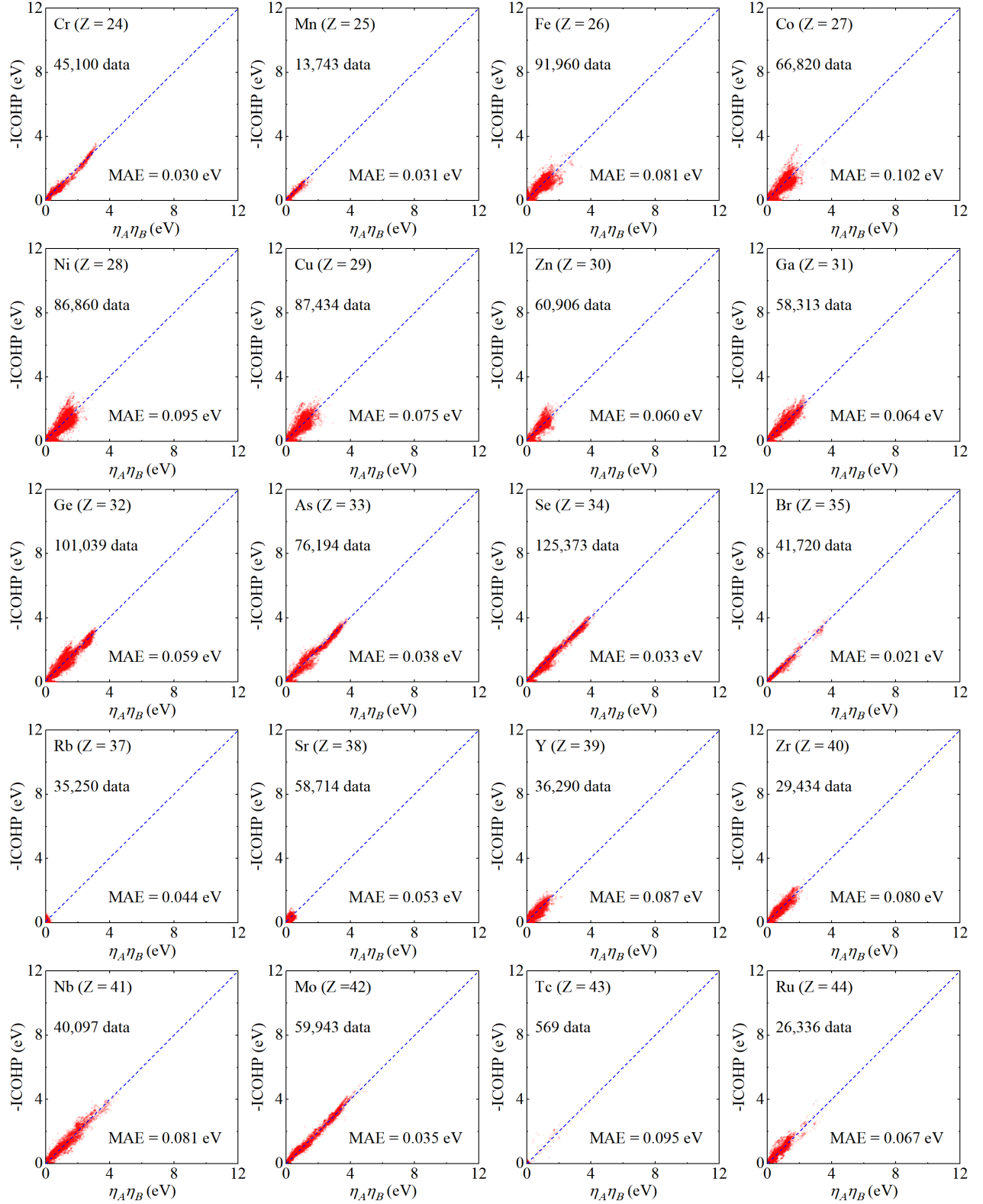


FIG. 8. Parity plots of BA-predicted ICOHP versus DFT-calculated ICOHP, resolved by the central element, from $Z = 24$ to $Z = 44$. Each panel aggregates all $A - B$ bonds that include the labeled element, with the diagonal indicating perfect agreement. Numbers inside each panel denote the sample size used for that element's bonds.

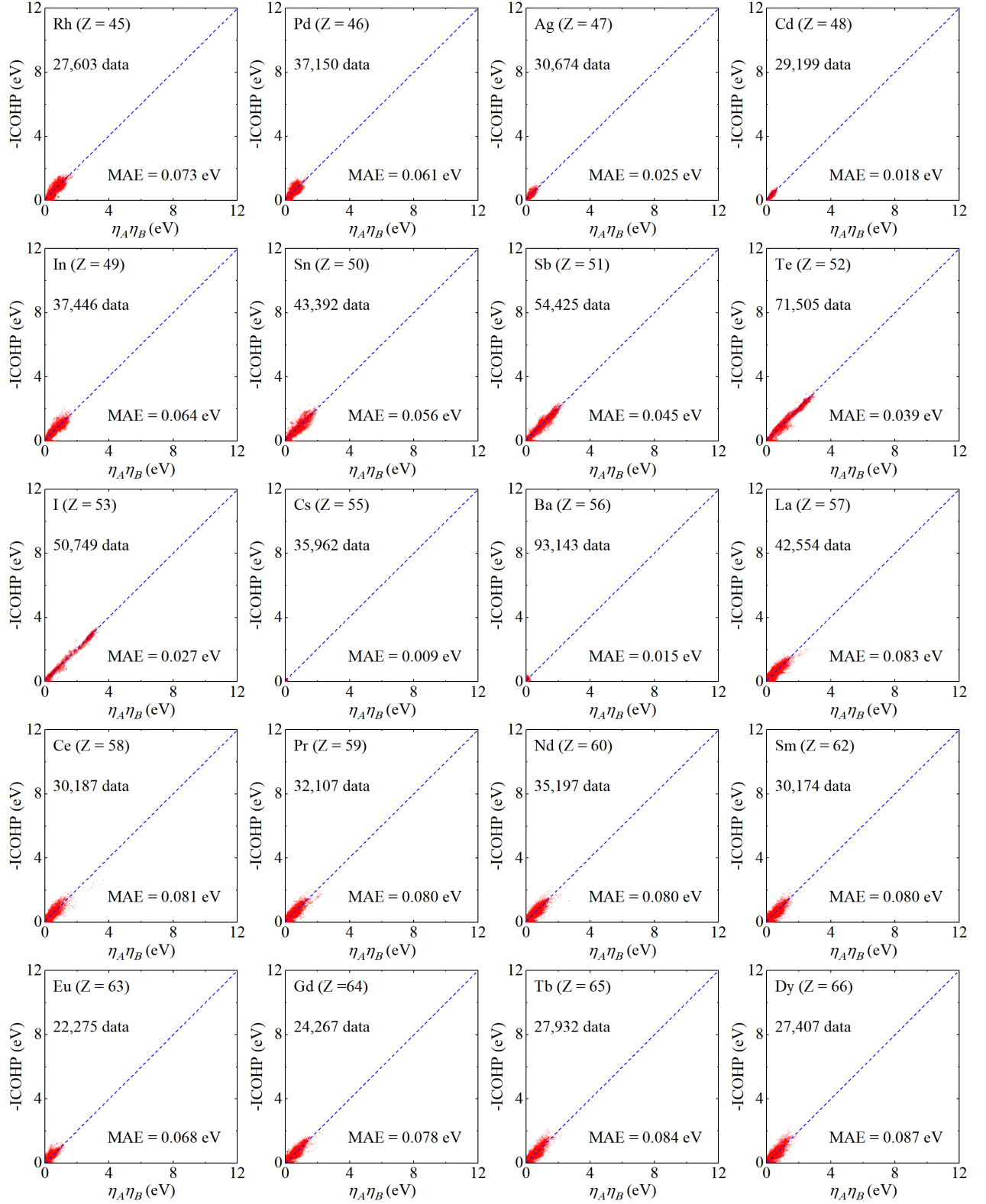


FIG. 9. Parity plots of BA-predicted ICOHP versus DFT-calculated ICOHP, resolved by the central element, from $Z = 45$ to $Z = 66$. Each panel aggregates all $A - B$ bonds that include the labeled element, with the diagonal indicating perfect agreement. Numbers inside each panel denote the sample size used for that element's bonds.

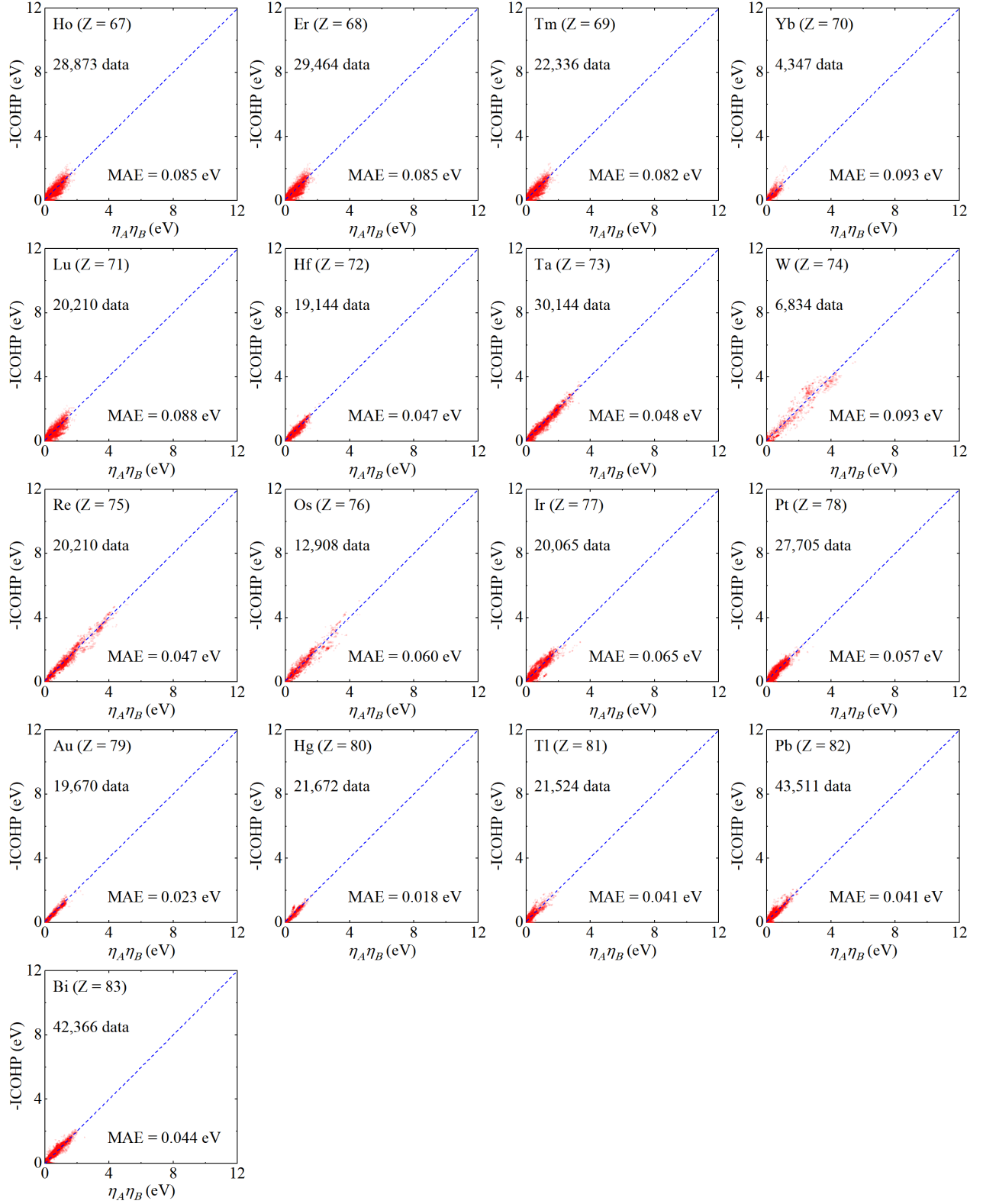


FIG. 10. Parity plots of BA-predicted ICOHP versus DFT-calculated ICOHP, resolved by the central element, from $Z = 67$ to $Z = 83$. Each panel aggregates all $A - B$ bonds that include the labeled element, with the diagonal indicating perfect agreement. Numbers inside each panel denote the sample size used for that element's bonds.

TABLE I. List of the valence orbital configurations, orbital numbers [31], and covalent radii [32] of all elements used in our calculations and analysis.

Element	Orbital Conf.	Orbital Num.	r_A (Å)	Element	Orbital Conf.	Orbital Num.	r_A (Å)
H	1s	1	0.31	Tc	4s4p5s4d5p	13	1.47
He	1s	1	0.28	Ru	4s4p5s4d	10	1.46
Li	1s2s2p	5	1.28	Rh	4s4p5s4d	10	1.42
Be	2s2p	4	0.96	Pd	4s4p5s4d	10	1.39
B	2s2p	4	0.84	Ag	4s4p5s4d	10	1.45
C	2s2p	4	0.76	Cd	4s4p5s4d	10	1.44
N	2s2p	4	0.71	In	4d5s5p	9	1.42
O	2s2p	4	0.66	Sn	4d5s5p	9	1.39
F	2s2p	4	0.57	Sb	5s5p	4	1.39
Ne	2s2p	4	0.58	Te	5s5p	4	1.38
Na	2s2p3s	5	1.66	I	5s5p	4	1.39
Mg	2s2p3s3p	8	1.41	Xe	4d5s5p	9	1.40
Al	3s3p	4	1.21	Cs	5s5p6s	5	2.44
Si	3s3p	4	1.11	Ba	5s5p6s	5	2.15
P	3s3p	4	1.07	La	5s5p6s5d6p4f	20	2.07
S	3s3p	4	1.05	Ce	5s5p6s5d6p	13	2.04
Cl	3s3p	4	1.02	Pr	5s5p6s5d6p	13	2.03
Ar	3s3p	4	1.06	Nd	5s5p6s5d6p	13	2.01
K	3s3p4s4p	8	2.03	Pm	5s5p6s5d6p	13	1.99
Ca	3s3p4s4p	8	1.76	Sm	5s5p6s5d6p	13	1.98
Sc	3s3p4s3d	10	1.70	Eu	5s5p6s5d6p	13	1.98
Ti	3s3p4s3d	10	1.60	Gd	5s5p6s5d6p	13	1.96
V	3s3p4s3d	10	1.53	Tb	5s5p6s5d6p	13	1.94
Cr	3s3p4s3d	10	1.39	Dy	5s5p6s5d6p	13	1.92
Mn	3s3p4s3d	10	1.39	Ho	5s5p6s5d6p	13	1.92
Fe	4s3d4p	9	1.32	Er	5s5p6s5d6p	13	1.89
Co	4s3d4p	9	1.26	Tm	5s5p6s5d6p	13	1.90
Ni	4s3d4p	9	1.24	Yb	5s5p6s5d6p	13	1.87
Cu	4s3d4p	9	1.32	Lu	5s5p6s5d6p	13	1.87
Zn	4s3d4p	9	1.22	Hf	5s5p6s5d	10	1.75
Ga	4s3d4p	9	1.22	Ta	5s5p6s5d	10	1.70
Ge	4s4p	4	1.20	W	5s5p6s5d6p	13	1.62
As	4s4p	4	1.19	Re	5s5p6s5d	10	1.51
Se	4s4p	4	1.20	Os	5s5p6s5d	10	1.44
Br	4s4p	4	1.20	Ir	5s5p6s5d	10	1.41
Kr	4s3d4p	9	1.16	Pt	5s5p6s5d	10	1.36
Rb	4s4p5s5p	8	2.20	Au	5s5p6s5d	10	1.36
Sr	4s4p5s5p	8	1.95	Hg	5s5p6s5d	10	1.32
Y	4s4p5s4d5p	13	1.90	Tl	5d6p6s	9	1.45
Zr	4s4p5s4d5p	13	1.75	Pb	5d6p6s	9	1.46
Nb	4s4p5s4d5p	13	1.64	Bi	5d6p6s	9	1.48
Mo	4s4p5s4d	10	1.54				