Gauge-independent Wannier parametrization for the *ab-initio* simulations of sub-μm-wide nanoribbons

Guan-Hao Peng, Wen-Teng Yang, Chun-Jui Huang, and Jhen-Dong Lin
Department of Electrophysics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

Shun-Jen Cheng*
Department of Electrophysics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan;
Institute of Physics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan;
and Center for Theoretical and Computational Physics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

(Received 12 January 2026; accepted 22 April 2026; published 11 May 2026)

Simulation of mesoscopic nanostructures is a central challenge in condensed matter physics and device applications. First-principles methods provide accurate electronic structures but are computationally too demanding to simulate mesoscopic nanostructures. Though empirical band theories are efficient for the simulation of nanostructures, their validities are limited by the empirical parametrization that usually lacks the consideration of microscopic wavefunctions and do not hold the transferability. In this work, we propose a methodology that bridges these approaches, achieving first-principles-level reliability with computational efficiency through a tight-binding framework. Our approach starts with Wannier tight-binding (WTB) parameters from small nanostructures, which serve as training data for fitting. To remove the gauge freedom of Wannier functions that obscures size- and geometry-dependent parameter trends, we construct gauge-independent (GI) bases and transform the WTB model into a gauge-independent WTB (GI-WTB) model. This enables robust parameter fitting and prediction of parameter variations, yielding the parametrized GI-WTB (P-GI-WTB) model. Applied to MoS_2 armchair-edge nanoribbons, the P-GI-WTB model shows excellent agreement with first-principles results and enables reliable simulations of sub-μm-wide nanoribbons. This framework offers a scalable approach for predicting the electronic properties of realistic nanostructures at a computational cost far lower than that of conventional first-principles methods.

DOI: 10.1103/zh4w-q4vm

I. INTRODUCTION

Over the past few decades, *ab initio* electronic structure calculations based on density functional theory (DFT) have advanced rapidly, driven by the growth of high performance computing (HPC) capabilities [1–8] and the emergence of modern machine-learning methods [9–18], including deep-learning techniques built on neural network architectures. Simulations of mesoscopic nanostructures, which require extremely large simulation cells because translational symmetry is reduced or broken, have thus become feasible with these developments. With exascale HPC platforms, state-of-the-art electronic excited state GW calculations based on DFT can now simulate systems containing over 1000 atoms [6,7] and even more than ten thousand atoms [8]. In parallel, deep-learning models trained on DFT data for small systems provide an efficient and practical route for predicting the properties of much larger structures, including carbon nanotubes and twisted bilayer graphenes [9,10,12].

Despite these advances, the limitations of these approaches are also evident. Exascale HPC calculations require exceptionally large computational resources, making them impractical for most users. Deep-learning approaches, while powerful, function largely as black boxes, making it difficult to interpret how electronic properties evolve as the geometry of a nanostructure changes. Alternatively, a practical and physically transparent strategy is to parametrize the DFT electronic structure from bulk or small systems and then predict the properties of large systems by fitting these parameters to the geometry and size of the nanostructure, thereby revealing how the electronic structure evolves with geometry. Since DFT calculations for bulk and small nanostructures can now be carried out at modest computational cost using well-established packages such as VASP [19] and Quantum Espresso [20,21], this parametrization-based strategy provides an efficient and accessible alternative for mesoscale simulations.

Building on the idea of parametrized models, it is useful to recall that empirical parametrizations have long provided a practical and efficient route for simulating large nanostructures. Representative approaches include $\boldsymbol{k}\cdot\boldsymbol{p}$ methods [22,23] and empirical tight-binding (ETB) models [24–30].

*Contact author: sjcheng@nycu.edu.tw

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

These models rely on a reduced set of fitted parameters chosen to reproduce selected band features, and they are numerically far less expensive than DFT. However, there are several issues that limit the usefulness and validity of empirical band theories in realistic simulations of nanostructures.

Because the set of fitting equations for the limited number of parameters in empirical band theory forms an overdetermined linear system, the parameters obtained for a specific material may vary substantially depending on the fitting procedure and algorithm. This nonuniqueness arises from the lack of microscopic wavefunction information, which limits the validity and applicability of such models across diverse nanostructures. Furthermore, although empirical band theories can reproduce DFT-calculated bands accurately in certain regions of the Brillouin zone (BZ), particularly near high-symmetry points, they often fail to capture the full complexity of the band structure across the entire BZ. This limitation poses significant challenges for studies of material properties, such as spectra of exciton complexes and angle-resolved photoemission spectra [31–33], which require accurate band information throughout the entire BZ [34–38].

Going beyond the limitations of empirical band theories, one can use localized Wannier functions, transformed from the DFT-calculated Bloch states of a material, as the basis set to construct the Wannier tight-binding (WTB) model, which is essentially equivalent to the Kohn-Sham (KS) Hamiltonian with the functional potential obtained from the converged self-consistent DFT calculations. Under the gauge freedom of Wannier functions [39–41], a chosen gauge transformation matrix (a $\boldsymbol{k}$-dependent unitary matrix) can be defined to transform the KS Hamiltonian matrix from the representation of Bloch states into Wannier functions. In this manner, the WTB parameters are deterministic with respect to the basis set of transformed Wannier functions, and the calculated band structures in the WTB scheme almost perfectly reproduce the DFT-calculated results. Practically, one can employ the package WANNIER90 [42–44] to establish the WTB model, which provides maximally localised Wannier functions (MLWFs) as well as all Wannier function-based tight-binding parameters. The high accuracy and physical transparency of the WTB approach make it particularly suitable for a wide range of applications beyond band structure calculations, including electron transport simulations and excited-state properties of materials [30,45,46].

Although the parametrization of a DFT-based WTB scheme is deterministic, its implementation still relies on the numerical feasibility of DFT calculations, whose applications are typically limited to only bulk materials or small nanostructures. Similarly, recently developed neural-network-based parametrization methods for tight-binding (TB) models of nanoribbons (NRs) [15] and defective systems [16,18] also require prior DFT calculations and therefore face the same limitation. Attempts to extend the application of the parameters obtained from the DFT-based results of bulk or specific small nanostructures to the TB models for larger nanostructures are doomed to fail, because the charge density in different nanostructures that are essential in DFT certainly differ from bulk and redistribute with the variation of geometric structures [47–50]. This fundamental limitation causes the nontransferability of TB parameters. In principle, this problem could be addressed by taking Wannier function-based parameters obtained for bulk systems or small nanostructures as training data and then using fitting to predict the WTB parameters for larger nanostructures. However, there exists another issue in the DFT-based WTB scheme that arises from the gauge freedom in Wannier function transformations [40,41]. This gauge freedom introduces arbitrariness in the unitary transformation matrix, leading to the nonuniqueness of the transformed Wannier functions and, consequently, making it infeasible to directly apply fitting techniques. Because of this nonuniqueness, the Wannier function-based parameters obtained from WTB models for different nanostructures are essentially uncorrelated and typically exhibit a scattered distribution with respect to the geometric variables of the nanostructures, thereby impeding the use of fitting methods.

In this work, we present a theoretical methodology for constructing a first-principles-based WTB theory applicable to large nanostructures at the sub-μm scale, beyond the computational reach of conventional DFT simulations. To enable data fitting, we remove the gauge freedom in Wannierization by performing a unitary transformation of the WTB Hamiltonian matrix, converting the basis of gauge-dependent Wannier functions into a specific set of atomic-orbital-like functions. Thus, the resulting WTB Hamiltonian matrix acquires gauge-independent (GI) parameters, which we refer to as the gauge-independent Wannier tight-binding (GI-WTB) model. By properly selecting these atomic-orbital-like functions, the GI-WTB parameters of different nanostructures form a smooth and systematic training dataset for fitting, exhibiting clear trends with respect to the geometric variables of the nanostructures. The smooth and systematic behavior of the training data obtained from the GI-WTB model facilitates classical extrapolation using simple fitting functions and also benefits machine-learning applications by reducing noise extrinsic to the problem, thereby improving accuracy and avoiding overfitting to local fluctuations. Based on DFT-derived GI-WTB parameters for bulk and small nanostructures, fitting techniques can be employed to predict physically valid GI-WTB models for nanostructures of arbitrary sizes, which we name as parametrized gauge-independent Wannier tight-binding (P-GI-WTB) models.

As a test nanosystem, we apply the developed P-GI-WTB methodology to monolayer transition-metal dichalcogenide (TMD) NRs with armchair edges [51–54] and calculate their electronic band structures for ribbon widths up to 200 nm. Using this approach, we perform a systematic DFT-based investigation of the width dependence of the energies and wavefunctions of TMD NRs over a broad range of ribbon widths, from a few nanometers to the sub-μm regime. From the calculated subband structures, we find that the energy gap of an NR rapidly converges to a constant value as the ribbon width increases. Analysis of the wavefunctions further reveals that low-lying conduction edge states near the band gap remain spectrally localized with only minor changes as the width increases, while states with mixed bulk-edge character will redshift toward the band gap and concentrate into a smaller spectral window as the ribbon width increase. High-lying conduction edge states far from the band gap are spectrally

broadly distributed and overlap with bulklike states spatially distributed in the central part of narrow NRs, while in wide NRs they concentrate into a narrow energy window, distinctly separated from the bulk states.

This paper is organized as follows. Section II presents the fundamental theory of the WTB model, the basis transformation framework for constructing the GI-WTB model, and the parameter-fitting strategy used to develop the P-GI-WTB model. In Sec. III, we apply the theoretical framework outlined in Sec. II to monolayer TMD NRs. This includes a statistical analysis of the hopping parameters as functions of the geometric variables of nanostructures, identification of the best-fit two-dimensional (2D) surfaces for these datasets, and construction of the P-GI-WTB model to predict the energy bands and wavefunctions of wide-width monolayer TMD NRs. Finally, Sec. IV summarizes the key findings of this work.

II. THEORY

A. Kohn-Sham equation

The DFT establishes a one-to-one correspondence between the ground-state number density $\rho(\boldsymbol{r})$ of an N_e-electron system and its N_e-electron Hamiltonian. The KS equation implements DFT by introducing a set of KS orbitals $\{\psi_j(\boldsymbol{r})\}$, which determine the density via $\rho(\boldsymbol{r}) = \sum_{j=1}^{N_e} |\psi_j(\boldsymbol{r})|^2$ [55]. Thus, the interacting N_e-electron problem is recast into a single-particle KS equation for $\psi_j(\boldsymbol{r})$, based on the KS Hamiltonian $H_{\rm KS}$, which consists of the kinetic energy and ρ-dependent effective potentials (Hartree and exchange-correlation terms), and can be solved numerically in a self-consistent manner. For crystalline solids, the KS orbitals can be expressed in Bloch form as $\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \langle \boldsymbol{r}|\psi_{n,\boldsymbol{k}}\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\, u_{n,\boldsymbol{k}}(\boldsymbol{r})$, where $u_{n,\boldsymbol{k}}(\boldsymbol{r})$ has the same periodicities as the crystal lattice, and the states are labeled by the band index n and Bloch wavevector $\boldsymbol{k}$. The corresponding KS equation reads

$$H_{\rm KS}|\psi_{n,\boldsymbol{k}}\rangle = \epsilon_{n,\boldsymbol{k}}|\psi_{n,\boldsymbol{k}}\rangle, \tag{1}$$

where $|\psi_{n,\boldsymbol{k}}\rangle$ denotes the KS orbital state, and $\epsilon_{n,\boldsymbol{k}}$ is the eigenenergy of the KS orbital.

B. Linear combination of atomic orbitals method

In the linear combination of atomic orbitals (LCAO) method, the KS orbital (a Bloch state) is expanded as

$$|\psi_{n,\boldsymbol{k}}\rangle = \sum_i C_i^{(n)}(\boldsymbol{k})|\phi_{i,\boldsymbol{k}}\rangle, \tag{2}$$

in terms of the Bloch sum basis $\{|\phi_{i,\boldsymbol{k}}\rangle\}$, defined by

$$|\phi_{i,\boldsymbol{k}}\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}|W_{i,\boldsymbol{R}}\rangle, \tag{3}$$

where N is the total number of unit cells determined by periodic boundary conditions, $C_i^{(n)}(\boldsymbol{k})$ are the expansion coefficients and $\langle \boldsymbol{r}|W_{i,\boldsymbol{R}}\rangle = W_i(\boldsymbol{r}-\boldsymbol{R})$ denotes an atomic-orbital-like function localized around an atom in the unit cell at position $\boldsymbol{R}$, and $\{|W_{i,\boldsymbol{R}}\rangle\}$ is considered to form an orthonormal basis set, which guarantees the orthonormality of the Bloch sum states. In the LCAO scheme, $i \to \{I, \alpha, s\}$ is a composite index specifying the Ith atom at $\boldsymbol{\tau}_I$ in the unit cell, the atomic orbital α, and the electron spin s for $|W_{i,\boldsymbol{R}}\rangle$. The Bloch sum basis of Eq. (3) satisfies the Bloch theorem, as does the Bloch state of Eq. (2).

Substituting Eq. (2) into Eq. (1) and using the orthonormality relations, $\langle W_{i,\boldsymbol{R}}|W_{j,\boldsymbol{R}'}\rangle = \delta_{i,j}\delta_{\boldsymbol{R},\boldsymbol{R}'}$, one obtains the eigenvalue equation

$$\sum_j H_{i,j}(\boldsymbol{k})\, C_j^{(n)}(\boldsymbol{k}) = \epsilon_{n,\boldsymbol{k}}\, C_i^{(n)}(\boldsymbol{k}), \tag{4}$$

which is essentially equivalent to the KS equation evaluated with the converged self-consistent potential obtained at the final iteration but reformulated in the LCAO basis, in terms of TB matrix elements $H_{i,j}(\boldsymbol{k}) \equiv \langle \phi_{i,\boldsymbol{k}}|H_{\rm KS}|\phi_{j,\boldsymbol{k}}\rangle$. From Eq. (3), these matrix elements can be written as

$$H_{i,j}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, t_{i,j}(\boldsymbol{R}), \tag{5}$$

where the on-site ($\boldsymbol{R} = \boldsymbol{0}$ and $\boldsymbol{\tau}_I = \boldsymbol{\tau}_J$) and hopping ($\boldsymbol{R} \neq \boldsymbol{0}$ or $\boldsymbol{\tau}_I \neq \boldsymbol{\tau}_J$) parameters are defined by $t_{i,j}(\boldsymbol{R}) = \langle W_{i,\boldsymbol{0}}|H_{\rm KS}|W_{j,\boldsymbol{R}}\rangle$. Depending on the hopping distance, $|(\boldsymbol{R}+\boldsymbol{\tau}_J) - \boldsymbol{\tau}_I|$, the parameters $\{t_{i,j}(\boldsymbol{R})\}$ are classified as first-, second-, third-nearest neighbors, and so forth.

In the TB theory, the eigenvalues $\epsilon_{n,\boldsymbol{k}}$ and corresponding eigenvectors $C_i^{(n)}(\boldsymbol{k})$ can be obtained by means of standard diagonalization of $H(\boldsymbol{k})$. In the LCAO method, usually only a small number of Bloch sum basis states are needed to achieve satisfactory convergence of the numerically solved eigenenergies. In this work, we include five d orbitals from each transition-metal atom and three p orbitals from each chalcogen atom for TMD NRs. While diagonalizing the reduced-size matrix in Eq. (5) is nowadays technically trivial using mature standard diagonalization packages such as LAPACK [56], the critical challenge in TB theory lies in its universal validity and transferability, that is, how to find out the parameters that are physically reasonable and generally valid.

1. Empirical tight-binding scheme

A common approach to determine the parameters $t_{i,j}(\boldsymbol{R})$ in Eq. (5) is by fitting the simulated band structure in the parametrized TB model to either experimental data or first-principles-calculated data. In the former case, the number of measurable quantities, such as band gaps and effective masses, is usually limited, leading to an underdetermined system. In the latter case, the number of parameters is far smaller than the data available from continuous energy bands, resulting in an overdetermined system in which the fitted parameters depend sensitively on the chosen dataset and fitting procedure. In both cases, the fitted parameters are not deterministic and hold the uniqueness. We refer to a TB model constructed in this way as an empirical tight-binding (ETB) model. The limitations of ETB models stem from the absence of a proper consideration of complex wavefunctions in the fitting process.

2. Wannier tight-binding scheme

In the WTB scheme, the parameters

$$t^{\lambda}_{I\alpha,J\beta}(\boldsymbol{R}) = \left\langle W^{\lambda}_{I\alpha,\boldsymbol{0}}\middle|H^{\lambda}_{\rm KS}\middle|W^{\lambda}_{J\beta,\boldsymbol{R}}\right\rangle \tag{6}$$

in Eq. (5) are evaluated from the atom-site-localized Wannier functions $\{|W^{\lambda}_{i,\boldsymbol{R}}\rangle\} = \{|W^{\lambda}_{I\alpha,\boldsymbol{R}}\rangle\}$. The index λ distinguishes different geometric structures of a common material, such as bulk systems and various nanostructures. To keep our focus on the geometric dependence of the parameters, we neglect spin-orbit coupling and disregard the electron spin s throughout this work. Wannier functions are obtained from DFT-calculated Bloch states via

$$|W^{\lambda}_{i,\boldsymbol{R}}\rangle \equiv \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}|\phi^{\lambda}_{i,\boldsymbol{k}}\rangle = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\sum_{n=1}^{N_\lambda} U^{\lambda(\boldsymbol{k})}_{n,i}|\psi^{\lambda}_{n,\boldsymbol{k}}\rangle, \tag{7}$$

where the Bloch sum basis states

$$|\phi^{\lambda}_{i,\boldsymbol{k}}\rangle = \sum_{n=1}^{N_\lambda} U^{\lambda(\boldsymbol{k})}_{n,i}|\psi^{\lambda}_{n,\boldsymbol{k}}\rangle \tag{8}$$

are obtained from the Bloch states via a $\boldsymbol{k}$-dependent unitary transformation $U^{\lambda(\boldsymbol{k})}$ [40]. Here, N_λ is the number of bands used to construct the sub-Hilbert space for system λ. Since $U^{\lambda(\boldsymbol{k})}$ generalizes the notion of a rotation in Euclidean space, $|\phi^{\lambda}_{i,\boldsymbol{k}}\rangle$ is also referred to as a rotated Bloch state [41,44].

The transformation matrix $U^{\lambda(\boldsymbol{k})}$ is typically determined through an iterative Wannierization procedure, as implemented in the WANNIER90 package [42–44]. Starting from an initial guess of $U^{\lambda(\boldsymbol{k})}$, often obtained by orbital projection [40,44], the procedure iteratively optimizes $U^{\lambda(\boldsymbol{k})}$ to minimize the spread functional of the Wannier functions. The resulting Wannier functions are known as MLWFs [40]. In this construction, the basis index i encodes both the position and the symmetry of the projecting orbital, so that Wannier functions effectively act as atomic orbitals centered on atomic sites.

Unlike ETB models, the Hamiltonian matrix in WTB model is essentially equivalent to the KS Hamiltonian evaluated with the converged self-consistent potential and, in principle, can reproduce the band structures directly from DFT [see Fig. S2(a) in the Supplemental Material [57] for a comparison of DFT and WTB band structures].

3. Nontransferability of parameters

However, both ETB and WTB models lack universal validity due to the nontransferability of parameters. A set of TB parameters that works well for a bulk material is often not applicable to nanostructures of the same material. For example, Fig. S2(b) in the Supplemental Material [57] shows the TB band structure of MoS_2 NRs obtained using the WTB parameters extracted from the 2D-bulk MoS_2 monolayer in Fig. S2(a), illustrating a clear failure of transferability. In nanostructures, valid TB parameters naturally differ from those of the bulk because Bloch states are influenced not only by intrinsic material properties but also by extrinsic factors such as geometry and size. In the next section, we introduce a fitting strategy to predict parameter variations as the geometry and size of the nanostructure change. Although the gauge freedom of Wannier functions can obscure the geometric dependence of WTB parameters, our proposed theoretical framework removes this gauge ambiguity and produces reliable geometry-dependent parametrizations.

C. Fitting of WTB theory for nanostructures

Systematically varying the nanostructure geometry allows one to derive fitting functions that explicitly capture the geometric dependence of parameters within the WTB model. These functions can then be used to predict parameters for larger-scale nanostructures by means of data-fitting procedures. In this section, we introduce a parameter-fitting scheme designed to represent the geometric dependence of WTB parameters, facilitating the calculation of electronic structures for realistically sized nanomaterials that are typically beyond the reach of direct DFT simulations.

1. Gauge freedom in the transformation of wannier functions

To reveal the geometric dependence of WTB parameters, we introduce a set of geometric variables $\{g^{\lambda}_{\ell}\}$ to characterize the structural features of a given nanosystem λ. The WTB parameters incorporating these variables are expressed as

$$t^{\lambda}_{I\alpha,J\beta}(\boldsymbol{R}) = t^{\lambda}_{I\alpha,J\beta}\big(\boldsymbol{R};\{g^{\lambda}_{\ell}\}\big). \tag{9}$$

The explicit definitions of $\{g^{\lambda}_{\ell}\}$ for the NRs considered in this work are provided in Sec. III B. For these NRs, $\{g^{\lambda}_{\ell}\}$ indicates the ribbon width and the position of an atomic chain relative to the ribbon edge.

At first glance, Eq. (9), which explicitly depends on $\{g^{\lambda}_{\ell}\}$, may appear adequate for capturing the geometric dependence of WTB parameters. However, the gauge freedom inherent in Wannier functions complicates the determination of an explicit geometric form for Eq. (9). Since Wannier functions are constructed from Bloch sum states obtained via a unitary transformation $U^{\lambda(\boldsymbol{k})}$ of DFT-calculated KS orbitals [see Eqs. (7) and (8)], the resulting WTB parameters therefore depend on the chosen $U^{\lambda(\boldsymbol{k})}$. In principle, $U^{\lambda(\boldsymbol{k})}$ can take any unitary form, subject only to the translational invariance condition $U^{\lambda(\boldsymbol{k}+\boldsymbol{G})} = U^{\lambda(\boldsymbol{k})}$, where $\boldsymbol{G}$ is a reciprocal lattice vector [41]. This gauge freedom introduces arbitrariness into the Wannier functions, ruins clear trends of WTB parameters with respect to $\{g^{\lambda}_{\ell}\}$, and hinders the fitting extensions of WTB theory for nanostructures.

To remove the influence of gauge freedom in Wannier functions, we introduce a GI basis set $\mathcal{S}^{\lambda,\mathrm{GI}} = \{|W^{\lambda,\mathrm{GI}}_{I\alpha,\boldsymbol{R}}\rangle\}$ for each system λ, where $|W^{\lambda,\mathrm{GI}}_{I\alpha,\boldsymbol{R}}\rangle$ serves as a atomic-orbital-like basis. Properly chosen GI basis functions are considered to span almost the same vector space as the Wannier functions in $\mathcal{S}^{\lambda} = \{|W^{\lambda}_{I\alpha,\boldsymbol{R}}\rangle\}$. Using this GI basis set, one can perform a basis transformation (see the next section) on Eq. (6) to obtain the new WTB parameters,

$$t^{\lambda,\mathrm{GI}}_{I\alpha,J\beta}(\boldsymbol{R}) = \langle W^{\lambda,\mathrm{GI}}_{I\alpha,\boldsymbol{0}}|H^{\lambda}_{\mathrm{KS}}|W^{\lambda,\mathrm{GI}}_{J\beta,\boldsymbol{R}}\rangle. \tag{10}$$

We refer to the new WTB model with parameters defined by Eq. (10) as the GI-WTB model. Incorporating geometric variables as in Eq. (9), we can express the GI-WTB parameters in Eq. (10) as

$$t^{\lambda,\mathrm{GI}}_{I\alpha,J\beta}(\boldsymbol{R}) = t^{\lambda,\mathrm{GI}}_{I\alpha,J\beta}\big(\boldsymbol{R};\{g^{\lambda}_{\ell}\}\big), \tag{11}$$

being a function of $\{g_\ell^\lambda\}$. The remaining task to accomplish a parametrizable WTB model is to find out the explicit form for Eq. (11).

In general, GI basis sets $\mathcal{S}^{\lambda,\mathrm{GI}}$ can be constructed from functions defined independently of the global phase of the KS orbitals, which is the origin of the gauge freedom of Wannier functions through $U^{\lambda(\boldsymbol{k})}$ in Eq. (7). To ensure that the transformation from the Wannier basis set $\mathcal{S}^\lambda$ to the GI basis set $\mathcal{S}^{\lambda,\mathrm{GI}}$ remains unitary, these GI basis functions should have good orthonormality and span the similar vector space as the original Wannier functions.

Here, we present a practical computational strategy for generating GI basis sets. We consider a sequence of nanostructured systems ordered by increasing size, labeled $\lambda_1, \lambda_2, \ldots, \lambda_{N_\mathrm{td}}$, where N_td denotes the number of systems included in the training data set. The smallest system λ_1 is chosen sufficiently large so that both the electronic states localized around the nanostructure (edge, surface, or defect states) and the bulklike states far from the nanostructure are insensitive to system size. In this way, the Wannier functions localized near the nanostructure in λ_1, together with Wannier functions in the bulk system, can be used as building blocks to construct GI basis sets. Specifically, we construct the GI basis set $\mathcal{S}^{\lambda_1,\mathrm{GI}}$ by using the Wannier functions localized near the nanostructure in λ_1 as basis functions centered around the nanostructure, and the Wannier functions extracted from the bulk system as basis functions centered in regions far from the nanostructure (bulk region). The GI basis set for the next system, $\mathcal{S}^{\lambda_2,\mathrm{GI}}$, is then obtained by appending additional Wannier functions extracted from the bulk system, used as basis functions centered in the bulk region, to $\mathcal{S}^{\lambda_1,\mathrm{GI}}$, and repeating this procedure yields nested basis sets satisfying

$$\mathcal{S}^{\lambda_1,\mathrm{GI}} \in \mathcal{S}^{\lambda_2,\mathrm{GI}} \in \cdots \in \mathcal{S}^{\lambda_{N_\mathrm{td}},\mathrm{GI}}. \tag{12}$$

In this construction, the GI basis sets are systematically related across different systems. Although a formal proof is not provided, the resulting GI basis sets exhibit good orthonormality and span the similar vector space as the Wannier functions to high accuracy for the MoS_2 NRs considered in this work.

Once the parameters exhibit a clear trend with respect to $\{g_\ell^\lambda\}$, we fit Eq. (11) using $t^{\lambda,\mathrm{P\text{-}GI}}_{I\alpha,J\beta}(\boldsymbol{R};\{g_\ell^\lambda\})$ and perform the replacement

$$t^{\lambda,\mathrm{GI}}_{I\alpha,J\beta}\big(\boldsymbol{R};\{g_\ell^\lambda\}\big) \rightarrow t^{\lambda,\mathrm{P\text{-}GI}}_{I\alpha,J\beta}\big(\boldsymbol{R};\{g_\ell^\lambda\}\big) \tag{13}$$

to construct a WTB model capable of predicting the electronic structure of large-scale nanostructures. We refer to the WTB model with parameters defined by Eq. (13) as the P-GI-WTB model.

D. Basis transformation

To formulate the basis transformation theory, we have to first define the vector space under discussion. In general, we can define the vector space of system λ as the one spanned by the KS orbitals in the selected N_λ bands, where the corresponding Bloch wavevector $\boldsymbol{k}$ are sampled on an N-point grid determined by periodic boundary conditions (PBCs). Within this space, the completeness relation is given by $1^\lambda = \sum_{n=1}^{N_\lambda}\sum_{\boldsymbol{k}} |\psi^\lambda_{n,\boldsymbol{k}}\rangle\langle\psi^\lambda_{n,\boldsymbol{k}}|$, where 1^λ is the identity operator for system λ. Using Eqs. (8) and (7), it follows that

$$1^\lambda = \sum_{i=1}^{N_\lambda}\sum_{\boldsymbol{k}} |\phi^\lambda_{i,\boldsymbol{k}}\rangle\langle\phi^\lambda_{i,\boldsymbol{k}}| = \sum_{i=1}^{N_\lambda}\sum_{\boldsymbol{R}} |W^\lambda_{i,\boldsymbol{R}}\rangle\langle W^\lambda_{i,\boldsymbol{R}}|, \tag{14}$$

indicating that Bloch sum states and Wannier states span the same vector space as the KS orbitals.

Using Eq. (14), the atomic-orbital-like basis in the set $\mathcal{S}^{\lambda,\mathrm{GI}}$, introduced in the previous section, can be expanded as

$$|W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}\rangle = \sum_{j=1}^{N_\lambda}\sum_{\boldsymbol{R}'} S^\lambda_{j,i}(\boldsymbol{R}-\boldsymbol{R}')|W^\lambda_{j,\boldsymbol{R}'}\rangle, \tag{15}$$

where the basis transformation matrix is defined as

$$S^\lambda_{j,i}(\boldsymbol{R}) = \langle W^\lambda_{j,\boldsymbol{0}}|W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}\rangle = \int_{V^\lambda} d^3\boldsymbol{r}\; W^{\lambda *}_{j,\boldsymbol{0}}(\boldsymbol{r})\, W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}(\boldsymbol{r}). \tag{16}$$

Here, the integration is carried out over the region V^λ defined by the PBCs, with volume $N|(\boldsymbol{a}_1\times\boldsymbol{a}_2)\cdot\boldsymbol{a}_3|$. The function $W^\lambda_{j,\boldsymbol{0}}(\boldsymbol{r}) = \langle\boldsymbol{r}|W^\lambda_{j,\boldsymbol{0}}\rangle$ is generated using the postprocessing tool WANNIER90 [43] and is localized near the atomic center in the home cell at $\boldsymbol{R}=\boldsymbol{0}$. The explicit form of $W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}(\boldsymbol{r}) = \langle\boldsymbol{r}|W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}\rangle$ depends on the system being studied. In later discussions on TMD NRs, we will construct the set $\{W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{R}}(\boldsymbol{r})\}$ that properly combines the Wannier functions of a narrow-width TMD NR and those of the 2D-bulk TMD monolayer.

Since the integrand in Eq. (16) is nonzero only in the region where the two functions overlap, using a uniform $\boldsymbol{r}$ grid would waste significant computational resources in areas where the integrand vanishes. To improve efficiency, we adopt a global adaptive strategy with nonuniform $\boldsymbol{r}$ grids, which significantly reduces the number of integration points and accelerates the computation.

Based on Eq. (15), the parameters in the GI-WTB model can be evaluated as

$$t^{\lambda,\mathrm{GI}}(\boldsymbol{R}) = \sum_{\boldsymbol{R}'}\sum_{\boldsymbol{R}''} S^{\lambda\dagger}(\boldsymbol{R}')\, t^\lambda(\boldsymbol{R}'')\, S^\lambda(\boldsymbol{R}-(\boldsymbol{R}''-\boldsymbol{R}')), \tag{17}$$

where $t^{\lambda,\mathrm{GI}}_{i,j}(\boldsymbol{R}) = \langle W^{\lambda,\mathrm{GI}}_{i,\boldsymbol{0}}|H^\lambda_\mathrm{KS}|W^{\lambda,\mathrm{GI}}_{j,\boldsymbol{R}}\rangle$ and $t^\lambda_{i,j}(\boldsymbol{R}) = \langle W^\lambda_{i,\boldsymbol{0}}|H^\lambda_\mathrm{KS}|W^\lambda_{j,\boldsymbol{R}}\rangle$.

III. RESULTS AND DISCUSSIONS

In this work, we choose monolayer MoS_2 armchair-edge NRs to demonstrate the parameter-fitting scheme proposed in Sec. II C.

A. Monolayer MoS_2 armchair-edge NRs

Figure 1(a) illustrates the atomic structure of a monolayer MoS_2 armchair-edge NR, where the width is characterized by the number of atomic chains, N_a. For convenience, we refer to this nanostructure as an N_a-NR. The lattice of an N_a-NR is described by the lattice vector $\boldsymbol{R} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$, where $\boldsymbol{a}_i$ are the primitive lattice vectors, and n_i are integers constrained by the PBCs. The vector $\boldsymbol{a}_2 = a_2\hat{\boldsymbol{y}}$ lies along the armchair direction, with $a_2 = 5.52$ Å, and determines the translational periodicity of an N_a-NR. The vectors $\boldsymbol{a}_1 = a_1\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_3 = a_3\hat{\boldsymbol{z}}$ lie along the zigzag direction and the out-of-plane direction, respectively. To prevent interactions between

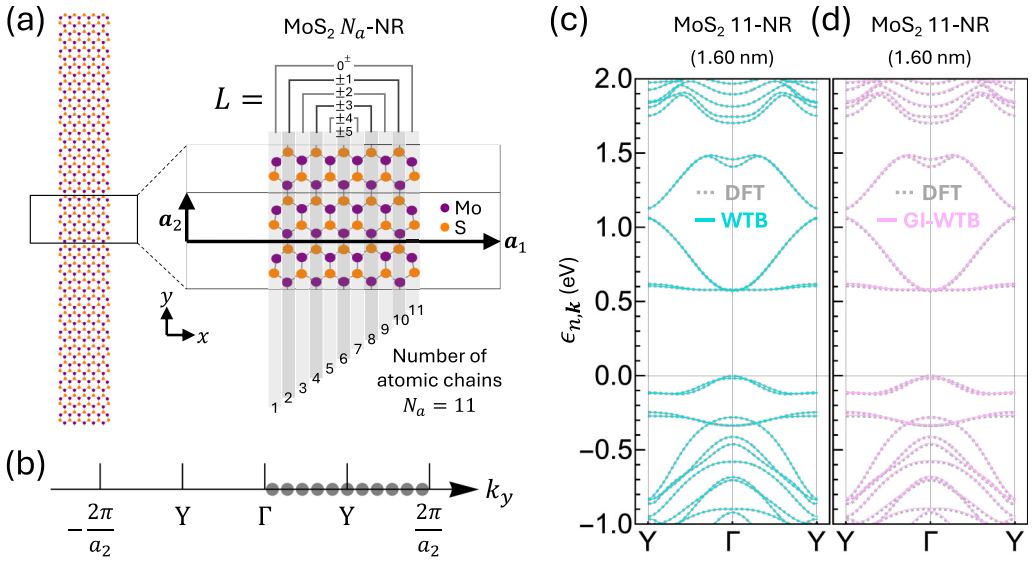


FIG. 1. (a) Top-down view of the structure-relaxed monolayer MoS_2 armchair-edge NR, where the lattice translational symmetry is defined by $\boldsymbol{a}_2 = a_2\hat{\boldsymbol{y}}$. Mo and S atoms are depicted in purple and orange, respectively. The integer N_a denotes the total number of atomic chains, which characterizes the ribbon width, while the L index is an integer that indicates the position of each atomic chain along $\boldsymbol{a}_2$. The positive (negative) L index indicates the position of the atomic chain on the right (left) of an NR. Here, we refer to this ribbon as N_a-NR. (b) The first Brillouin zone of the N_a-NR, where the gray points indicate the mesh used for both the DFT calculations and Wannierization. (c) Calculated band structures of the monolayer MoS_2 11-NR with a width of 1.60 nm, obtained from DFT (gray) and the WTB model (cyan). (d) Calculated band structures of the same nanoribbon obtained from the GI-WTB model (pink). In all cases, the bands are aligned by shifting the valence band maximum to zero.

periodic images in DFT calculations, vacuum layers of thickness 2 and 1.6 nm are introduced along $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, respectively. Following the convention $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{i,j}$, Fig. 1(b) shows the first BZ, which is defined by the reciprocal lattice vector $\boldsymbol{b}_2 = (2\pi/a_2)\hat{\boldsymbol{y}}$.

In this work, the DFT band structures of monolayer MoS_2 N_a-NRs are calculated using QUANTUM ESPRESSO [20,21], employing the generalized gradient approximation with the Perdew-Burke-Ernzerhof exchange-correlation functional [58]. The plane-wave basis cutoff energy is set to 1088 eV, and the $\boldsymbol{k}$ mesh is sampled using a $1 \times 11 \times 1$ Monkhorst-Pack grid, represented by the gray points in Fig. 1(b). Before structure relaxation, the atomic structures of monolayer MoS_2 N_a-NRs are initialized with a Mo-S bond length of 2.42 Å and an out-of-plane S-S distance of 3.13 Å. Structural relaxation and self-consistent calculations are considered converged when the energy difference between consecutive iterations falls below 9.5×10^{-4} and 9.5×10^{-6} eV, respectively. Figure 1(a) presents the relaxed atomic structure of a monolayer MoS_2 11-NR, while the corresponding DFT band structure is shown as gray dashed lines in Figs. 1(c) and 1(d).

Following the first-principles calculations, the DFT results are transformed into the WTB model using the postprocessing tool WANNIER90 [43], which converts the Bloch states from a plane-wave representation into a Wannier representation (see Sec. II B 2). For MoS_2 NRs, a disentanglement procedure [41,44,59] is performed to construct an optimal subspace for Wannierization. Specifically, we extract an optimal subspace of dimension $11N_a$ from the DFT bands using an outer energy window, where this dimension is determined by the initial guess of the rectangular unitary matrix constructed by projecting the KS orbitals within the outer window onto trial orbitals consisting of the five d orbitals of the Mo atom and the three p orbitals of the top and bottom S atoms. To ensure that the electronic states near the band gap, which are of primary interest, remain faithful to the original KS orbitals, we employ an inner frozen window that encloses the target bands around the band gap. Based on the Bloch states within the resulting optimal subspace, we then perform Wannierization according to Eq. (7), where the unitary transformation matrix $U^{\lambda(\boldsymbol{k})}$ is obtained by projecting these Bloch states onto the same set of trial orbitals used in the disentanglement procedure. Figure 1(c) shows the WTB band structure of the MoS_2 11-NR, obtained by diagonalizing the Hamiltonian matrix in Eq. (5), demonstrating excellent agreement with the DFT results.

To ensure that the Wannier functions used in the subsequent calculation of the basis transformation matrix in Eq. (16) are orthonormal, we employ norm-conserving (NC) pseudopotentials in our DFT calculations. As a technical remark, we compare the absolute values of the inner product matrix $|\langle W^{\mathrm{2D}}_{i,\boldsymbol{0}}|W^{\mathrm{2D}}_{j,\boldsymbol{0}}\rangle|$ for Wannier functions constructed from DFT calculations using NC and projector-augmented-wave (PAW) pseudopotentials for 2D-bulk MoS_2, as shown in Fig. S1 of the Supplemental Material [57]. Although the DFT bands obtained with NC and PAW pseudopotentials are nearly identical [see Fig. S1(a) and (b)], the corresponding inner product matrices of the Wannier functions [Fig. S1(c) and (d)] differ significantly. For NC pseudopotentials, the maximum off-diagonal elements of $|\langle W^{\mathrm{2D}}_{i,\boldsymbol{0}}|W^{\mathrm{2D}}_{j,\boldsymbol{0}}\rangle|$ are on the order of 10^{-8}, whereas for PAW pseudopotentials they reach values on the order of 10^{-2}. This comparison indicates that Wannier

functions obtained from NC pseudopotentials exhibit superior orthonormality compared to those obtained from PAW pseudopotentials.

For comparison, we also construct TB models for MoS_2 N_a-NRs using parameters taken from the WTB model of 2D-bulk MoS_2 (see Sec. S.II of the Supplemental Material [57]). The resulting band structure, shown in Fig. S2(b), exhibits apparent discrepancies with both the DFT and WTB results. The discrepancy highlights the nontransferability of TB parameters, which arises because the structure-dependent charge density variation is disregarded. For instance, the charge redistribution effects associated with edge formation are entirely neglected when bulk parameters are directly applied to NRs [47].

Although the WTB model is highly accurate, it relies on prior DFT calculations whose feasibility is restricted by the high demand for computational resources. To overcome this bottleneck, we propose the parameter-fitting scheme introduced in Sec. II C. In this approach, the WTB model is first transformed into the GI-WTB model to remove gauge freedom in Wannier functions. The resulting GI-WTB parameters are then used to construct a training dataset within the geometric variable space g_ℓ^λ of MoS_2 N_a-NRs. Fitting these parameters yields the P-GI-WTB model, which enables the prediction of the TB parameters for very wide NRs.

B. The GI-WTB model for monolayer MoS_2 N_a-NRs

For a series of studies on NRs with different widths, λ can be denoted as N_a-NR, and the geometric variables introduced in Eq. (9) are $g_1^\lambda = N_a$, representing the ribbon width, and $g_2^\lambda = L \in \mathbb{Z}$, representing the position of an atomic chain in the NR relative to the ribbon edge, as illustrated in Fig. 1(a). In this work, the edge region of an NR is defined as the atomic chains with $|L| \leqslant 3$, while the bulk region consists of atomic chains with $|L| > 3$.

The atomic-orbital-like basis $|W_{I\alpha,\boldsymbol{R}}^{\lambda=N_a\text{-NR,GI}}\rangle$ within the set $\mathcal{S}^{\lambda=N_a\text{-NR,GI}}$ is constructed by combining the basis of the Wannier functions from the monolayer MoS_2 11-NR and 2D-bulk MoS_2 monolayer ($\lambda = 2D$). For the edge region of N_a-NRs, the basis is obtained by shifting $|W_{I\alpha,\boldsymbol{R}}^{\lambda=11\text{-NR}}\rangle$ from the edge region of monolayer MoS_2 11-NR. For the bulk region, it is obtained by shifting $|W_{I\alpha,\boldsymbol{R}=\boldsymbol{0}}^{\lambda=2D}\rangle$ from the home cell of 2D-bulk MoS_2 monolayer. This procedure yields GI basis sets satisfying Eq. (12), i.e., $\mathcal{S}^{\lambda_1,\text{GI}} \in \mathcal{S}^{\lambda_2,\text{GI}} \in \cdots \in \mathcal{S}^{\lambda_5,\text{GI}}$, with $\lambda_\mu = (11 + 2(\mu - 1))$-NR for $1 \leqslant \mu \leqslant 5$. Further details are provided in Sec. S.III of the Supplemental Material [57].

With the constructed $\mathcal{S}^{N_a\text{-NR,GI}}$, Eqs. (16) and (17) are used to numerically evaluate the basis transformation matrix $S^{N_a\text{-NR}}(\boldsymbol{R})$ and GI-WTB parameters $t_{I\alpha,J\beta}^{N_a\text{-NR,GI}}(\boldsymbol{R})$. Substituting $t_{I\alpha,J\beta}^{N_a\text{-NR,GI}}(\boldsymbol{R})$ into Eq. (5) yields the Hamiltonian matrix of the GI-WTB model for MoS_2 N_a-NRs, which can then be diagonalized to obtain the corresponding eigenvalues and eigenvectors [see Eq. (4)].

From a physical perspective, it is reasonable to assume that the charge distribution of the edge states of NRs extends only over a limited distance from the edges and remains almost invariant as the ribbon is sufficiently wide. For NRs with $N_a \geqslant 11$, the edge-state charge distribution is expected to remain stable and localized within the region $|L| \leqslant 3$. Likewise, for bulk states, the charge distribution is expected to localize near the ribbon center for sufficiently wide NRs and can be effectively described by the Wannier functions of the 2D-bulk system, which are localized within the region $|L| > 3$.

Numerical tests validate the adoption of the combined basis set. By diagonalizing the Hamiltonian matrix of the GI-WTB model constructed using $\mathcal{S}^{N_a\text{-NR,GI}}$, the resulting band structures, shown as pink lines in Fig. 1(d), exhibit excellent agreement with the DFT results. Further consistent results between the GI-WTB model and DFT for $N_a > 11$ are provided in Sec. S.IV of the Supplemental Material [57]. Although we have not analytically proven that $\mathcal{S}^{N_a\text{-NR,GI}}$ spans the same space as $\mathcal{S}^{N_a\text{-NR}}$, the numerical results clearly demonstrate its suitability and reliability.

As a technical remark, the iteration steps in the Wannierization process are crucial to our basis transformation theory. In WANNIER90 [43], the center and profile of Wannier functions evolve during the iteration process to minimize the spread functional. Since the overlap between two Wannier functions can change significantly due to minor adjustments in their profiles and centers, numerous iterations can introduce unforeseen changes in the basis transformation matrix defined in Eq. (16), leading to instability. To address this, we adopt a *one-shot* Wannierization procedure for both $W_{j,\boldsymbol{0}}^{N_a\text{-NR}}(\boldsymbol{r})$ and the Wannier functions used to construct $W_{i,\boldsymbol{R}}^{N_a\text{-NR,GI}}(\boldsymbol{r})$ for evaluating $S_{j,i}^{N_a\text{-NR}}(\boldsymbol{R})$ in Eq. (16). In this approach, the matrix $U^{\lambda(\boldsymbol{k})}$ is determined in a single step using the orbital projection method [40,44]. The resulting Wannier functions closely preserve the intended profiles and centers specified by the projection orbitals. By shifting these *well-behaved* Wannier functions to the atomic sites of the N_a-NRs, our numerical tests confirm that the basis transformation results are stable and reliable.

C. Parametrization for monolayer MoS_2 N_a-NRs

To demonstrate the advantages of the GI-WTB model, we analyze parameters in the geometric variable space spanned by $\{g_1^\lambda, g_2^\lambda\} = \{N_a, L\}$. In Fig. 2(a), we present parameters from the WTB model, $t_{I\alpha,J\beta}^{N_a\text{-NR}}(\boldsymbol{R}; N_a, L)$, and from the GI-WTB model, $t_{I\alpha,J\beta}^{N_a\text{-NR,GI}}(\boldsymbol{R}; N_a, L)$, for $I = J = 3$, $\alpha = \beta = d_{z^2}$, $\boldsymbol{R} = \boldsymbol{0}$, $L = 0$, and $N_a = 11, 13, 15, 17, 19$. These correspond to the on-site energies of the Mo d_{z^2} orbital at the edge of NRs. The GI-WTB model was constructed by transforming the basis sets of the corresponding one-shot WTB models into the GI basis sets introduced and validated in the previous section.

As seen in Fig. 2(a), the dataset for the GI-WTB model exhibits a better monotonic trend, as compared to that of the WTB model. A more illustrative comparison is shown in Fig. 2(b), which presents the hopping parameters from the WTB and GI-WTB models for the case with $I = 3$, $\alpha = d_{z^2}$, $J = 6$, $\beta = d_{x^2-y^2}$, $\boldsymbol{R} = \boldsymbol{0}$, $L = 0$, and $N_a = 11, 13, 15, 17, 19$. The WTB model shows a scattered geometric dependence, reflecting the influence of gauge freedom. In contrast, the GI-WTB model produces a smooth and consistent trend, demonstrating its gauge-independent nature.

In our fitting procedure, the data of GI-WTB parameters in Eq. (11) are first categorized by the orbital indices (α and β) and tunneling vector $\boldsymbol{d} \equiv (\boldsymbol{\tau}_J + \boldsymbol{R}) - \boldsymbol{\tau}_I$. For example,

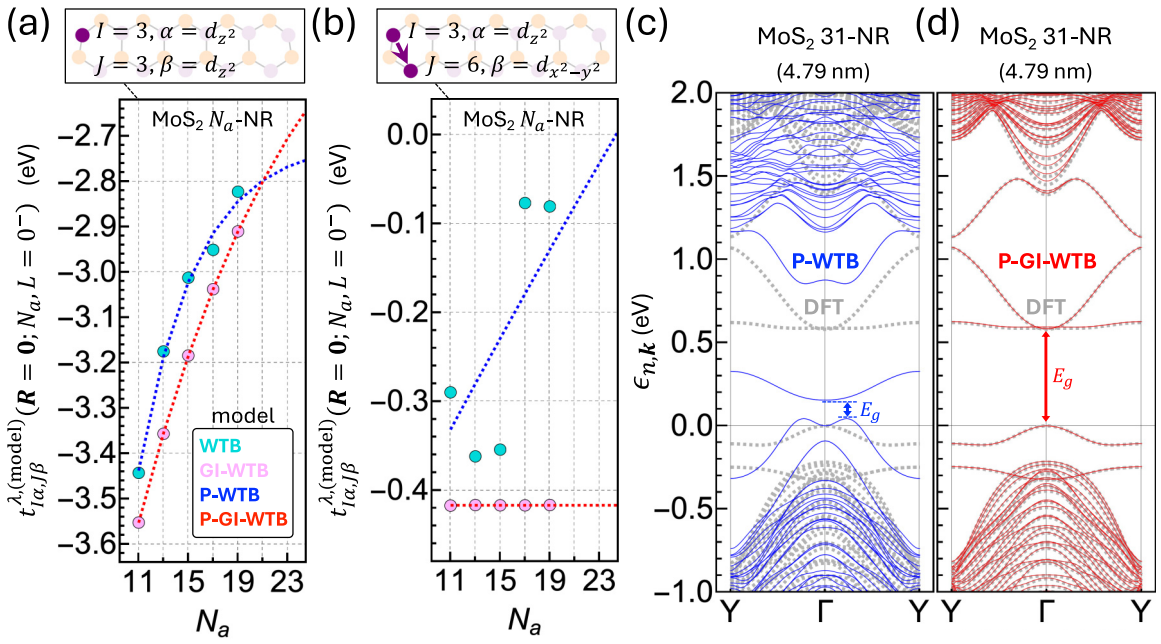


FIG. 2. (a) N_a dependence of the on-site energy for the d_{z^2} orbital at the atom indicated in the top schematic (illustrated using 11-NR). (b) N_a dependence of the hopping parameter between the d_{z^2} and $d_{x^2-y^2}$ orbitals, as denoted by the arrow in the top schematic (again shown using 11-NR). The cyan and pink data points are obtained from the WTB and GI-WTB models, respectively. The blue dashed line for the P-WTB model is fitted to the cyan data points, while the red dashed line for the P-GI-WTB model is fitted to the pink data points. (c) Band structure of monolayer MoS_2 31-NR from the P-WTB model (blue) compared with DFT (gray). (d) Band structure of monolayer MoS_2 31-NR from the P-GI-WTB model (red) compared with DFT (gray). Double-headed arrows indicate the energy band gap, E_g.

Fig. 2(a) belongs to the category $\{\alpha = d_{z^2}, \beta = d_{z^2}, \boldsymbol{d} = \boldsymbol{0}\}$, while Fig. 2(b) belongs to $\{\alpha = d_{z^2}, \beta = d_{x^2-y^2}, \boldsymbol{d} = \boldsymbol{d}_0\}$, with $\boldsymbol{d}_0$ indicated by the purple arrow in the top schematic of Fig. 2(b). Within each category, parameters are organized into a training dataset on the two-dimensional geometric variable space spanned by $g^\lambda_1 = N_a$ and $g^\lambda_2 = L$. For instance, Fig. 2(a) shows only the subset at $L = 0^-$. To build the complete dataset, the same plotting procedure as in Fig. 2(a) is repeated for all other L values ($0^+, \pm1, \pm2, \ldots$). By combining these plots, we obtain the full training dataset over the N_a-L plane for the category $\{\alpha = d_{z^2}, \beta = d_{z^2}, \boldsymbol{d} = \boldsymbol{0}\}$. A similar procedure is applied to Fig. 2(b) and to other on-site and hopping parameters. The fitting procedure is likewise applied to the WTB parameters in Eq. (9).

After constructing the training dataset, we fit the parameters using the functions $t^{\lambda,\mathrm{P\text{-}GI}}_{I\alpha,J\beta}(\boldsymbol{R}; \{g^\lambda_\ell\})$ and $t^{\lambda,\mathrm{P}}_{I\alpha,J\beta}(\boldsymbol{R}; \{g^\lambda_\ell\})$ introduced in Sec. II C 1. As the hopping parameters of the NRs are expected to approach those of the 2D bulk as L and N_a increase toward infinity, we adopt exponential decay functions of L and N_a to capture this asymptotic behavior. For datasets that are symmetric under the interchange of $L \to -L$, we adopt the fitting function $t^{\lambda,\mathrm{P\text{-}GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) = \delta_1 + \delta_2 \exp(-\delta_4|L|) + \delta_3 \exp(-\delta_5 N_a)$ and $t^{\lambda,\mathrm{P}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) = \gamma_1 + \gamma_2 \exp(-\gamma_4|L|) + \gamma_3 \exp(-\gamma_5 N_a)$, where δ_μ and γ_μ are fitting parameters. For datasets that are antisymmetric under $L \to -L$, we adopt the fitting function $t^{\lambda,\mathrm{P\text{-}GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) = \mathrm{sgn}(L)\,[\delta_1 + \delta_2 \exp(-\delta_4|L|) + \delta_3 \exp(-\delta_5 N_a)]$ and $t^{\lambda,\mathrm{P}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) = \mathrm{sgn}(L)\,[\gamma_1 + \gamma_2 \exp(-\gamma_4|L|) + \gamma_3 \exp(-\gamma_5 N_a)]$, where sgn() denotes the sign function, and $\{\gamma_4, \gamma_5, \delta_4, \delta_5\}$ are assumed to be positive.

To determine the fitting parameters δ_μ and γ_μ, we employ the least-squares method to minimize the residual functions

$$\Delta_{\alpha,\beta,\boldsymbol{d}}(\boldsymbol{\delta}) = \sum_{N_a}\sum_{L} \left| t^{N_a\text{-NR,P-GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) - t^{N_a\text{-NR,GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) \right|^2 \tag{18}$$

and

$$\Gamma_{\alpha,\beta,\boldsymbol{d}}(\boldsymbol{\gamma}) = \sum_{N_a}\sum_{L} \left| t^{N_a\text{-NR,P}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) - t^{N_a\text{-NR}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L) \right|^2, \tag{19}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$, and to find out the best-fitted parameters. At first glance, the indices $\{I, J, \boldsymbol{R}\}$ in Eqs. (18) and (19) may appear undetermined. In fact, they are constrained by the condition $\boldsymbol{d} = (\boldsymbol{\tau}_J + \boldsymbol{R}) - \boldsymbol{\tau}_I$, which defines the training dataset. Each $\{I, J, \boldsymbol{R}\}$ satisfying this condition corresponds uniquely to a coordinate (N_a, L) in the geometric variable space. Therefore, when summing over all (N_a, L) points in the training dataset, all valid $\{I, J, \boldsymbol{R}\}$ are automatically included, ensuring that no indices remain ambiguous.

To secure the correct asymptotic behavior, we impose boundary conditions during the optimization. These conditions require the fitting functions to converge to the WTB parameters of 2D-bulk MoS_2 monolayer, $t^{N_a\text{-NR,P-GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a \to \infty, L \to \pm\infty) = t^{N_a\text{-NR,P}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a \to \infty, L \to \pm\infty) = t^{\mathrm{2D}}_{I\alpha,J\beta}(\boldsymbol{R})$, during the minimization of Eqs. (18)

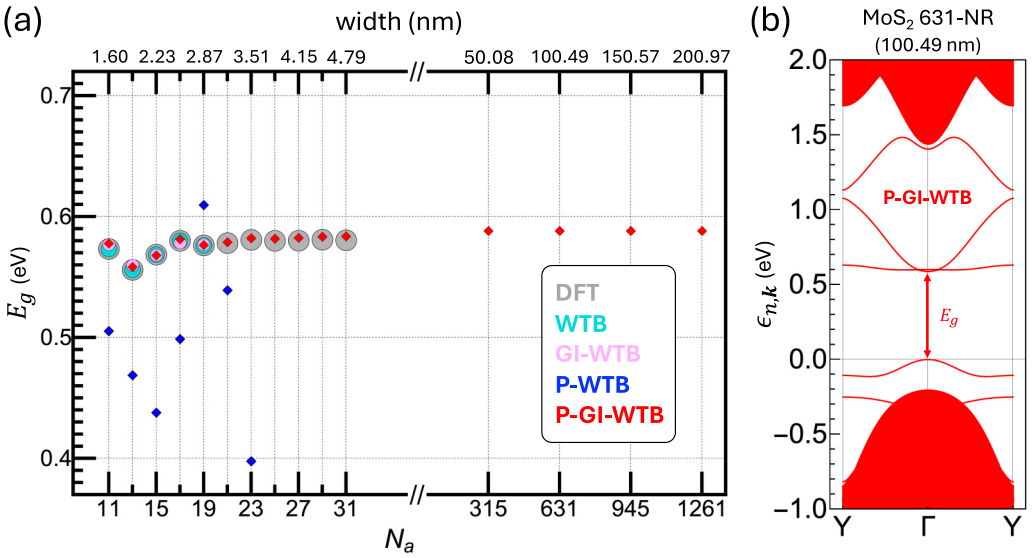


FIG. 3. (a) Dependence of the energy gap E_g on N_a, ranging from $N_a = 11$ to $N_a = 1261$. The upper axis shows the corresponding widths of the monolayer MoS_2 N_a-NR. Data points are color coded according to the model, with gray for DFT, cyan for WTB, pink for GI-WTB, blue for P-WTB, and red for P-GI-WTB. (b) Band structure of the sub-μm-wide monolayer MoS_2 631-NR computed using the P-GI-WTB model.

and (19). This essentially improves the transferability of the parameters for nanostructures of sizes ranging over wide length scales. By applying the replacement in Eq. (13) to the Hamiltonian matrix in Eq. (5), we can obtain the P-GI-WTB model. The same procedure is also used to construct the P-WTB model (see Sec. II C 1).

In Figs. 2(a) and 2(b), the P-WTB parameters, $t^{N_a\text{-NR,P}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L)$, and the P-GI-WTB parameters, $t^{N_a\text{-NR,P-GI}}_{I\alpha,J\beta}(\boldsymbol{R}; N_a, L)$, are shown as blue and red dashed lines, respectively. A clear discrepancy is observed between the WTB and P-WTB results for the on-site energies when $N_a > 15$, as shown in Fig. 2(a), with further deviations evident in the hopping terms shown in Fig. 2(b). In contrast, the GI-WTB data points exhibit excellent agreement with the corresponding fitting curves in the P-GI-WTB model for both on-site energies and hoppings, highlighting the robustness of the proposed GI-WTB model.

Using the P-WTB and P-GI-WTB models, we can predict the parameters for MoS_2 N_a-NRs with $N_a \gg 10^2$, which lie beyond the range of the training dataset [see Figs. 2(a) and 2(b)]. In this regime, the predictions from the P-WTB model are expected to be inaccurate, while those from the P-GI-WTB model exhibit to be reliable. Based on the predicted parameters, the corresponding TB Hamiltonian matrices are constructed using Eq. (5). Figures 2(c) and 2(d) compare the DFT band structure (gray dashed lines) of monolayer MoS_2 31-NR with the results obtained from the P-WTB model (blue lines) and the P-GI-WTB model (red lines), respectively. The P-GI-WTB model reproduces the DFT bands with excellent accuracy, whereas the P-WTB model produces significant deviations.

To further demonstrate the effectiveness of our approach, Fig. 3(a) presents the energy band gap E_g for monolayer MoS_2 N_a-NRs, starting from $N_a = 11$ (1.60 nm) to $N_a = 1261$ (200.97 nm). As a reference, we compute the DFT results for E_g up to $N_a = 31$. Within this range, the P-GI-WTB model exhibits excellent agreement with the DFT results. In contrast, the P-WTB model yields disorganized and significantly deviated E_g values, reflecting the limitations introduced by the gauge freedom in Wannier functions for parameter fitting purpose. As seen in Fig. 3(a), the band gap saturates to a constant value as ribbon width increases. For illustration, Fig. 3(b) presents the band structure of a MoS_2 631-NR (100.49 nm wide), where the spectrum exhibits nearly continuous valence and conduction bands at higher energies.

D. Electron density distributions of sub-μm wide NRs

In NRs, identifying the spatial probability distribution of eigenstates, including bulk and edge states, is essential for practical applications [60–62]. The proposed P-GI-WTB model is a powerful tool for this purpose, as it can directly provide microscopic charge density distributions of electrons in realistically wide NRs, far beyond the reach of conventional DFT calculations.

To facilitate the following analysis, we set the origin of the x axis at the ribbon center. Accordingly, we define the probability of finding an electron around the atomic chain located at $x = x_L$ and indexed by L in a NR [see Fig. 1(a)] as

$$P^L_{n,\boldsymbol{k}} = \sum_{\substack{I \\ \tau_{I,x}=x_L}} \sum_{\alpha\in\mathcal{A}_{\Theta(I)}} \left|C^{(n)}_{I\alpha}(\boldsymbol{k})\right|^2, \tag{20}$$

where $x_L = \pm(\frac{N_a-1}{2})x_0 \mp |L|\,x_0$ is the x position of the atomic chain for $L = \pm|L|$, x_0 is the spacing between atomic chains, and $\Theta(I) = \delta_{\mathrm{mod}(I,3),0} + 1$ is the atomic-species function. For TMDs, the orbital set $\mathcal{A}_1 = \{p_z, p_x, p_y\}$ corresponds to the p orbitals of the chalcogen atoms, and the orbital set $\mathcal{A}_2 = \{d_{z^2}, d_{xz}, d_{yz}, d_{x^2-y^2}, d_{xy}\}$ corresponds to the d orbitals of the transition-metal atoms (see Sec. S.III of the Supplemental Material [57] for details).

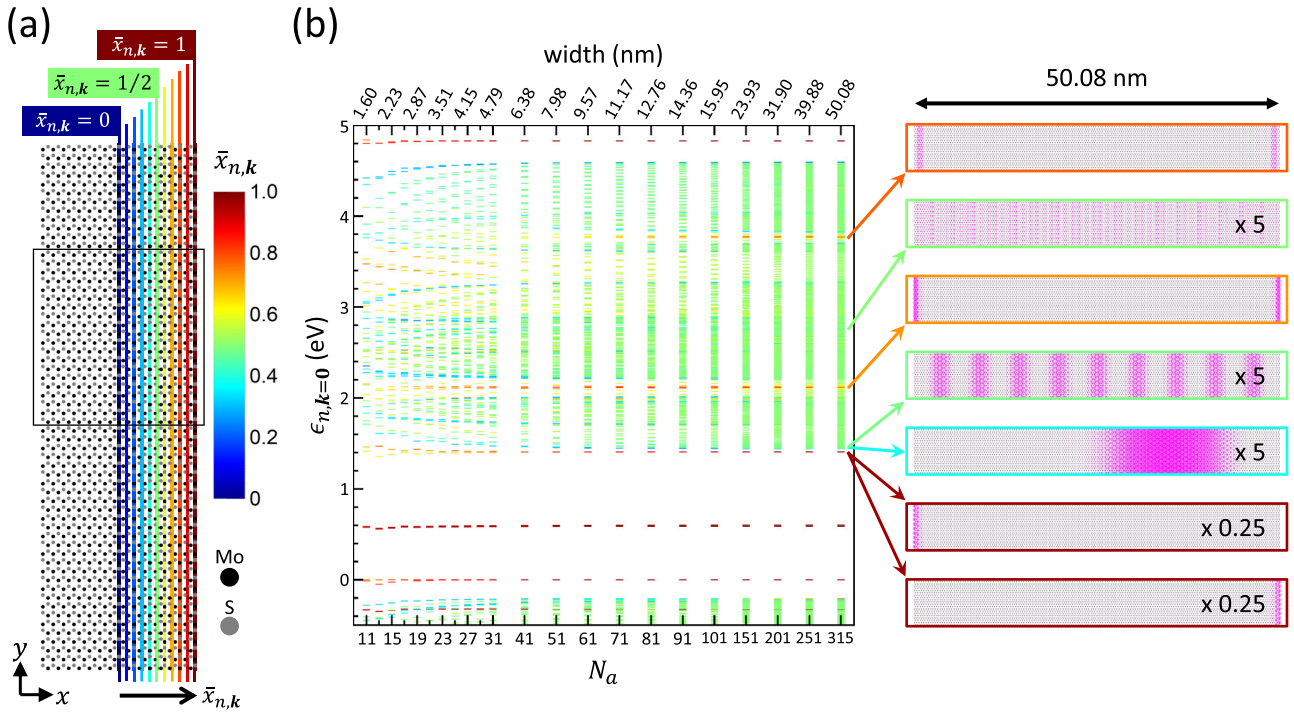


FIG. 4. (a) Schematic illustration of the relative average position $\bar{x}_{n,k}$ for the Bloch state $|\psi_{n,k}\rangle$ [see definition in Eq. (21)] in monolayer MoS_2 N_a-NR. The deep blue line marks $\bar{x}_{n,k} = 0$ at the ribbon center. The deep red line marks $\bar{x}_{n,k} = 1$ at the ribbon edge. The light green line marks $\bar{x}_{n,k} = 1/2$ at the midpoint between the ribbon center and the edge. Intermediate values of $\bar{x}_{n,k}$ between 0 and 1 are color coded according to the scale and shown as colored lines. Mo atoms are shown in black and S atoms are shown in gray to avoid confusion from overuse of colors. (b) Energy levels at the Γ point for MoS_2 N_a-NRs. Each Bloch state is color-coded by its corresponding $\bar{x}_{n,k}$ value using the scale in panel (a). For the 50.08 nm wide NR ($N_a = 315$), several states are selected as representative examples of the probability distributions (indicated by color coded arrows). The radius of the magenta circles represents the probability at each atomic site. Numbers at the upper right of each probability plot indicate the applied scaling factors.

We characterize the spatial charge distribution of a Bloch state in an NR, $|\psi_{n,k}\rangle$, by defining

$$\bar{x}_{n,k} = \frac{\sum_L P^L_{n,k} |x_L|}{w/2}, \tag{21}$$

which measures the averaged location of electron along the positive x axis. Here, $w = (N_a - 1)x_0$ is the ribbon width. In Eq. (21), the numerator gives the average position of the Bloch state, and dividing by $w/2$ yields its relative average position within the NR. Figure 4(a) illustrates the interpretation of $\bar{x}_{n,k}$. Bulk states may fall in the range $0 \leqslant \bar{x}_{n,k} \leqslant 0.5$, edge states may fall in the range $0.5 \leqslant \bar{x}_{n,k} \leqslant 1$, and bulk-edge mixed states may appear around $\bar{x}_{n,k} \approx 0.5$.

Figure 4(b) presents the energy spectra of band states at the Γ point for MoS_2 N_a-NRs of different widths, where the relative average position $\bar{x}_{n,k}$ is color coded according to the scale in Fig. 4(a). The states near 0 and 0.6 eV correspond to the valence band maximum and conduction band minimum, respectively, which remain stable with varying ribbon width, consistent with the band gap E_g behavior shown in Fig. 3(a). Conduction edge states around 1.5 eV remain localized in energy and shift only slightly as the width increases. Their color gradually changes from yellow to red, showing that their localization centers move progressively closer to the ribbon edges. Conduction states below 2 eV shown in cyan shift toward the band gap and become concentrated within a narrower energy range for larger widths. High-lying conduction edge states between 3 and 4 eV, which overlap in energy with extended states (green) in narrow ribbons, contract into a narrow energy window in wide ribbons and separate clearly from the extended states. From the spectra of conduction states, it is evident that the number of extended states (green) increases with width, while the number of edge states (yellow to red) remains nearly unchanged. In the valence band, states near −0.2 eV change from cyan to orange, indicating that their localization centers move from the interior region of the ribbon toward the edges. These observations reveal a consistent trend. As the ribbon width increases, states with stronger localization (cyan, yellow, orange) tend to move toward the edges, while extended states (green) become more numerous and occupy the spectral regions between the localized states. This increase in extended states is advantageous for charge transport in wide NRs.

To confirm that $\bar{x}_{n,k}$ provides a reliable measure of the spatial distribution of band states, we also plot the probability distribution from the first sum (the sum over α) in Eq. (20) for the 50.08 nm MoS_2 315-NR, as indicated by the color-coded arrows in Fig. 4(b). The results reproduce the same trends captured by $\bar{x}_{n,k}$. With the use of the developed P-GI-WTB model and the analysis of the delivered real-space probability distribution, one can further resolve the distinct spatial distribution of bulk-edge mixed states, highlighted in cyan and green.

IV. CONCLUSIONS

In this work, we developed a Wannier TB fitting framework that provides a practical and physically transparent route for simulating mesoscopic nanostructures. While exascale HPC platforms and modern deep-learning architectures nowadays are available to overcome the computational limitations of DFT, accessing such computational resources is prohibitively expensive for most users, and neural-network-based models often behave as black boxes whose internal decisions are difficult to interpret or validate. Our approach offers an alternative solution by using WTB parameters obtained from DFT calculations of small nanostructures as the training dataset in the geometric variable space for fitting, thereby overcoming the computational bottleneck without sacrificing interpretability.

A key challenge in this approach is the gauge freedom of Wannier functions, which introduces arbitrariness in WTB parameters and obscures their dependence on nanostructure size and geometry, therefore hindering systematic parameter fitting. To resolve this, we constructed atomic-orbital-like GI bases and transformed the WTB model into a GI-WTB model. This GI formulation restores clear geometric trends in the parameters, enabling robust fitting and interpolation across the geometric variable space and yielding the P-GI-WTB model capable of simulating nanostructures at realistic scales.

As a demonstration, we applied our fitting scheme to MoS_2 armchair-edge NRs. The framework reproduced DFT band structures with high accuracy. Building on this agreement, we further used P-GI-WTB model to predict parameter variations with respect to geometric variables and to simulate both energy band structures and wavefunctions for realistically wide ribbons with the sub-μm-scaled width.

The results reveal a clear evolution of electronic states as the ribbon widens. States with strong localization, regardless of whether they localize in the interior of the ribbon or near the edges, gradually shift toward the edge regions with increasing width. Meanwhile, the number of extended states increases and these states fill the spectral regions between the localized bands. This growing population of extended states is beneficial for charge transport in wide NRs. In addition to energy spectra, the P-GI-WTB model provides complete real-space wavefunction information for all band states. By examining the relative average position at the Γ point, we can clearly distinguish bulk states, edge states, and states with mixed bulk-edge character with high spatial and spectral resolution.

In conclusion, the P-GI-WTB model establishes a powerful and scalable methodology that combines first-principles-level reliability with computational efficiency. This framework enables predictive modeling of nanostructures at mesoscopic scales, provides a foundation for systematic studies of size- and geometry-dependent electronic properties, and offers significant potential for guiding the design of next-generation quantum devices.

ACKNOWLEDGMENTS

The successful completion of this research was made possible by the academic resources and advanced research infrastructure provided by the National Center for High-Performance Computing (NCHC), Taiwan. We gratefully acknowledge their invaluable support. This work was supported by the National Science and Technology Council (NSTC) of Taiwan under Grants No. NSTC 113-2112-M-A49-035-MY3, No. NSTC 114-2124-M-A49-004-, No. NSTC 114-2923-M-A49-005-MY3, No. NSTC 114-2112-M-A49-016-MY2, and No. NSTC-DAAD 114-2927-I-A49-501.

S.J.C. conceived and supervised the project. G.H.P. developed the theoretical formulation, framework, and analysis. C.J.H. and W.T.Y. performed the density functional theory calculations and Wannierization for MoS_2 nanoribbons. C.J.H. carried out the numerical evaluation of the basis transformation matrix. W.T.Y. performed the parameter fitting to develop the P-GI-WTB model. J.D.L. carried out the numerical evaluation of the Wannier function inner product matrix used to validate orthonormality. G.H.P. and S.J.C. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

The authors declare no competing interests.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data and code are available from the authors upon reasonable request.

[1] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations, Comput. Mater. Sci. **58**, 227 (2012).

[2] 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 Mater. **1**, 011002 (2013).

[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] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, Nature (London) **624**, 80 (2023).

[5] F. A. Rasmussen and K. S. Thygesen, Computational 2D materials database: Electronic structure of transition-metal dichalcogenides and oxides, J. Phys. Chem. C **119**, 13169 (2015).

[6] M. Del Ben, F. H. da Jornada, A. Canning, N. Wichmann, K. Raman, R. Sasanka, C. Yang, S. G. Louie, and J. Deslippe, Large-scale GW calculations on pre-exascale HPC systems, Comput. Phys. Commun. **235**, 187 (2019).

[7] M. D. Ben, C. Yang, Z. Li, F. H. d. Jornada, S. G. Louie, and J. Deslippe, Accelerating large-scale excited-

state GW calculations on leadership HPC systems, in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis, Atlanta, GA, USA* (IEEE, Piscataway, NJ, 2020), pp. 1–11.

[8] B. Zhang, D. Weinberg, C.-E. Hsu, A. R. Altman, Y. Shi, J. B. White III, D. Vigil-Fowler, S. G. Louie, J. R. Deslippe, F. H. da Jornada, Z. Li, and M. Del Ben, Advancing quantum many-body GW calculations on exascale supercomputing platforms, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '25 (Association for Computing Machinery, New York, 2025), pp. 48–59.

[9] 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, Nat. Comput. Sci. **2**, 367 (2022).

[10] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian, Nat. Commun. **14**, 2848 (2023).

[11] Y. Wang, Y. Li, Z. Tang, H. Li, Z. Yuan, H. Tao, N. Zou, T. Bao, X. Liang, Z. Chen, S. Xu, C. Bian, Z. Xu, C. Wang, C. Si, W. Duan, and Y. Xu, Universal materials model of deep-learning density functional theory Hamiltonian, Sci. Bull. **69**, 2514 (2024).

[12] X. Gong, S. G. Louie, W. Duan, and Y. Xu, Generalizing deep learning electronic structure calculation to the plane-wave basis, Nat. Comput. Sci. **4**, 752 (2024).

[13] M. Su, J.-H. Yang, H.-J. Xiang, and X.-G. Gong, Efficient determination of the Hamiltonian and electronic properties using graph neural network with complete local coordinates, Mach. Learn.: Sci. Technol. **4**, 035010 (2023).

[14] Y. Zhong, H. Yu, M. Su, X. Gong, and H. Xiang, Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids, npj Comput. Mater. **9**, 182 (2023).

[15] Z. Wang, S. Ye, H. Wang, J. He, Q. Huang, and S. Chang, Machine learning method for tight-binding Hamiltonian parameterization from *ab initio* band structure, npj Comput. Mater. **7**, 11 (2021).

[16] C. Schattauer, M. Todorović, K. Ghosh, P. Rinke, and F. Libisch, Machine learning sparse tight-binding parameters for defects, npj Comput. Mater. **8**, 116 (2022).

[17] Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang, and W. E, Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with *ab initio* accuracy, Nat. Commun. **15**, 6772 (2024).

[18] H. P. Fried, D. Barragan-Yani, F. Libisch, and L. Wirtz, A machine learning approach to predict tight-binding parameters for point defects via the projected density of states, npj Comput. Mater. **11**, 176 (2025).

[19] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[20] 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).

[21] 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).

[22] L. Brey and H. A. Fertig, Edge states and the quantized Hall effect in graphene, Phys. Rev. B **73**, 195408 (2006).

[23] L. Brey and H. A. Fertig, Electronic states of graphene nanoribbons studied with the Dirac equation, Phys. Rev. B **73**, 235411 (2006).

[24] J. C. Slater and G. F. Koster, Simplified LCAO method for the periodic potential problem, Phys. Rev. **94**, 1498 (1954).

[25] M. Bieniek, L. Szulakowska, and P. Hawrylak, Effect of valley, spin, and band nesting on the electronic properties of gated quantum dots in a single layer of transition metal dichalcogenides, Phys. Rev. B **101**, 035401 (2020).

[26] M. Korkusinski, O. Voznyy, and P. Hawrylak, Fine structure and size dependence of exciton and biexciton optical spectra in CdSe nanocrystals, Phys. Rev. B **82**, 245304 (2010).

[27] Z. Z. Zhang, K. Chang, and F. M. Peeters, Tuning of energy levels and optical properties of graphene quantum dots, Phys. Rev. B **77**, 235411 (2008).

[28] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, Edge state in graphene ribbons: Nanometer size effect and edge shape dependence, Phys. Rev. B **54**, 17954 (1996).

[29] G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, and D. Xiao, Three-band tight-binding model for monolayers of group-VIB transition metal dichalcogenides, Phys. Rev. B **88**, 085433 (2013).

[30] E. Ridolfi, L. R. F. Lima, E. R. Mucciolo, and C. H. Lewenkopf, Electronic transport in disordered MoS_2 nanoribbons, Phys. Rev. B **95**, 035430 (2017).

[31] S.-Y. Chen, T. Goldstein, T. Taniguchi, K. Watanabe, and J. Yan, Coulomb-bound four-and five-particle intervalley states in an atomically-thin semiconductor, Nat. Commun. **9**, 3717 (2018).

[32] E. Liu, J. van Baren, C.-T. Liang, T. Taniguchi, K. Watanabe, N. M. Gabor, Y.-C. Chang, and C. H. Lui, Multipath optical recombination of intervalley dark excitons and trions in monolayer WSe_2, Phys. Rev. Lett. **124**, 196802 (2020).

[33] J. Madéo, M. K. L. Man, C. Sahoo, M. Campbell, V. Pareek, E. L. Wong, A. Al-Mahboob, N. S. Chan, A. Karmakar, B. M. K. Mariserla, X. Li, T. F. Heinz, T. Cao, and K. M. Dani, Directly visualizing the momentum-forbidden dark excitons and their dynamics in atomically thin semiconductors, Science **370**, 1199 (2020).

[34] G.-H. Peng, P.-Y. Lo, W.-H. Li, Y.-C. Huang, Y.-H. Chen, C.-H. Lee, C.-K. Yang, and S.-J. Cheng, Distinctive signatures of the spin-and momentum-forbidden dark exciton states in the photoluminescence of strained WSe_2 monolayers under thermalization, Nano Lett. **19**, 2299 (2019).

[35] P.-Y. Lo, G.-H. Peng, W.-H. Li, Y. Yang, and S.-J. Cheng, Full-zone valley polarization landscape of finite-momentum exciton in transition metal dichalcogenide monolayers, Phys. Rev. Res. **3**, 043198 (2021).

[36] C.-H. Shih, G.-H. Peng, P.-Y. Lo, W.-H. Li, M.-L. Xu, C.-H. Chien, and S.-J. Cheng, Signatures of valley drift in the diversified band dispersions of bright, gray, and dark excitons in MoS_2 monolayers under uniaxial strains, Phys. Rev. B **111**, 245422 (2025).

[37] D. Y. Qiu, T. Cao, and S. G. Louie, Nonanalyticity, valley quantum phases, and lightlike exciton dispersion in monolayer

transition metal dichalcogenides: Theory and first-principles calculations, Phys. Rev. Lett. **115**, 176801 (2015).

[38] T. Deilmann and K. S. Thygesen, Finite-momentum exciton landscape in mono- and bilayer transition metal dichalcogenides, 2D Mater. **6**, 035003 (2019).

[39] G. H. Wannier, The structure of electronic excitation levels in insulating crystals, Phys. Rev. **52**, 191 (1937).

[40] N. Marzari and D. Vanderbilt, Maximally localized generalized Wannier functions for composite energy bands, Phys. Rev. B **56**, 12847 (1997).

[41] 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).

[42] 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).

[43] 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).

[44] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, WANNIER90: A tool for obtaining maximally-localised Wannier functions, Comput. Phys. Commun. **178**, 685 (2008).

[45] S. Chen, S. Kim, W. Chen, J. Yuan, R. Bashir, J. Lou, A. M. van der Zande, and W. P. King, Monolayer MoS_2 nanoribbon transistors fabricated by scanning probe lithography, Nano Lett. **19**, 2092 (2019).

[46] J. Have, N. M. R. Peres, and T. G. Pedersen, Excitonic magneto-optics in monolayer transition metal dichalcogenides: From nanoribbons to two-dimensional response, Phys. Rev. B **100**, 045411 (2019).

[47] M. Gibertini and N. Marzari, Emergence of one-dimensional wires of free carriers in transition-metal-dichalcogenide nanostructures, Nano Lett. **15**, 6229 (2015).

[48] P. Bhattacharyya, R. Chaudhari, N. Alaal, T. Rana, and A. Shukla, Influence of edge functionalization on electronic and optical properties of armchair phosphorene nanoribbons: A first-principles study, Electron. Struct. **2**, 025001 (2020).

[49] N. T. Han, K. D. Vo, T. Le Manh, O. K. Le, and D. T. Van, Electronic, magnetic, and optical properties of Np and Pu decorated armchair graphene nanoribbons: A DFT study, Nanoscale Adv. **6**, 3878 (2024).

[50] D. Hoat, V. K. Dien, Q. D. Ho, D. P. Dam, N. T. Tien, and D. K. Nguyen, Rich essential properties of silicon-substituted graphene nanoribbons: A comprehensive computational study, Phys. Chem. Chem. Phys. **26**, 15939 (2024).

[51] Y.-H. Chen, C.-H. Lee, S.-J. Cheng, and C.-K. Yang, Electronic structures of WS_2 armchair nanoribbons doped with transition metals, Sci. Rep. **10**, 16452 (2020).

[52] H. Pan and Y.-W. Zhang, Edge-dependent structural, electronic and magnetic properties of MoS_2 nanoribbons, J. Mater. Chem. **22**, 7280 (2012).

[53] R. Wang, X. Zhou, X. Xu, J. Hu, and J. Pan, The indirect–direct band gap tuning in armchair MoS_2 nanoribbon by edge passivation, J. Phys. D: Appl. Phys. **50**, 095102 (2017).

[54] L. Zhang, L. Wan, Y. Yu, B. Wang, F. Xu, Y. Wei, and Y. Zhao, Modulation of electronic structure of armchair MoS_2 nanoribbon, J. Phys. Chem. C **119**, 22164 (2015).

[55] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. **140**, A1133 (1965).

[56] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).

[57] See Supplemental Material at http://link.aps.org/supplemental/10.1103/zh4w-q4vm for the verification of the orthonormality of Wannier functions, the demonstration of the nontransferability of TB parameters, the detailed construction of the GI basis sets for MoS_2 nanoribbons, and additional band results for MoS_2 nanoribbons obtained from the GI-WTB model.

[58] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[59] I. Souza, N. Marzari, and D. Vanderbilt, Maximally localized Wannier functions for entangled energy bands, Phys. Rev. B **65**, 035109 (2001).

[60] F. Zhang, C.-H. Lee, J. A. Robinson, and J. Appenzeller, Exploration of channel width scaling and edge states in transition metal dichalcogenides, Nano Res. **11**, 1768 (2018).

[61] F. Cheng, H. Xu, W. Xu, P. Zhou, J. Martin, and K. P. Loh, Controlled growth of 1D $MoSe_2$ nanoribbons with spatially modulated edge states, Nano Lett. **17**, 1116 (2017).

[62] W. Fu, M. John, T. D. Maddumapatabandi, F. Bussolotti, Y. S. Yau, M. Lin, and K. E. Johnson Goh, Toward edge engineering of two-dimensional layered transition-metal dichalcogenides by chemical vapor deposition, ACS Nano **17**, 16348 (2023).