A machine learning route between band mapping and band structure

R. Patrick Xian,[1,3]* Vincent Stimper,[2,3] Marios Zacharias,[1] Shuo Dong,[1] Maciej Dendzik,[1] Samuel Beaulieu,[1] Bernhard Schölkopf,[2] Martin Wolf,[1] Laurenz Rettig,[1] Christian Carbogno,[1] Stefan Bauer,[2]* and Ralph Ernstorfer[1]*

[1]Fritz Haber Institute of the Max Planck Society, 14195 Berlin, Germany.
[2]Department of Empirical Inference, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany.
[3]These authors contributed equally to this work.

*Correspondence authors: xian@fhi-berlin.mpg.de, stefan.bauer@tue.mpg.de, ernstorfer@fhi-berlin.mpg.de.

The electronic band structure (BS) of solid state materials imprints the multidimensional and multi-valued functional relations between energy and momenta of periodically confined electrons [1]. Photoemission spectroscopy is a powerful tool for its comprehensive characterization. A common task in photoemission band mapping is to recover the underlying quasiparticle dispersion [2], which we call band structure reconstruction. Traditional methods often focus on specific regions of interests [3, 4] yet require extensive human oversight. To cope with the growing size and scale of photoemission data [5, 6], we develop a generic machine-learning approach leveraging the information within electronic structure calculations for this task. We demonstrate its capability by reconstructing all fourteen valence bands of tungsten diselenide and validate the accuracy on various synthetic data. The reconstruction uncovers previously inaccessible momentum-space structural information on both global and local scales in conjunction with theory, while realizing a path towards integrating band mapping data into materials science databases.

The modelling and characterization of the electronic band structure of materials play an essential role in materials design [7] and device simulation [8]. The BS lives in the momentum space, $\Omega(k_x, k_y, k_z, E)$, spanned by momentum (k_x, k_y, k_z) and energy (E) coordinates of the electrons. Photoemission band mapping [2] (see Fig. 1a) using momentum- and energy-resolved photoemission spectroscopy (PES), including angle-resolved PES (ARPES) [9] and multidimensional PES [5, 6] measures the BS as an intensity-valued multivariate probability distribution directly in Ω. The proliferation of band mapping datasets and their public availability brought about by recent hardware upgrades [5, 6, 10, 11] have ushered in the possibilities of comprehensive benchmarking between theories and experiments. Interpreting the photoemission spectra often requires least-squares fitting of the 1D energy distribution curves (EDCs) to the single-particle spectral function [9] at selected momentum locations through heuristics derived from physical knowledge of the materials and the experimental settings. Although, in principle, such a physics-informed data model guarantees the highest accuracy and interpretability, upscaling the pointwise fitting (or estimation) to large, densely sampled regions in the momentum space (e.g. including 10^4 or more momentum locations) presents challenges due to limited numerical stability, nor does it form a constructive mathematical framework for progressive refinement. The situation becomes especially challenging for multidimensional and multiband materials with complex band dispersions [12, 13, 14].

To navigate and comprehend band mapping data globally, we propose an analytical framework (see Fig. 1b) for reconstructing the photoemission (or quasiparticle) BS as a set of energy (or electronic) bands, formed by energy values (i.e. band loci) connected along momentum coordinates. Because the local maxima of photoemission intensities are not always good indicators of band loci, we exploit the connection between theory and experiment in our framework, based on a probabilistic machine learning [15, 16] model to approximate the intensity data from band mapping experiments. The gist of the model is rooted in the Bayes rule,

$$p(\boldsymbol{\Theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}), \tag{1}$$

where the model parameters $\boldsymbol{\Theta} = \{\theta_i\}$ and the data $\mathcal{D}$ are mapped directly onto unknowns and experimental observables. We assign the energy values of the photoemission BS as the model parameters to extract from data, and a nearest-neighbor (NN) Gaussian distribution as prior, $p(\boldsymbol{\Theta})$, to describe the proximity of energy values at nearby momenta. The EDC at every momentum grid point relates to the likelihood, $p(\mathcal{D}|\boldsymbol{\Theta})$, when we interpret the photoemission in-

tensity probabilistically. The optimal parameters are obtained via *maximum a posteriori* (MAP) estimation, or maximizing the posterior, $p(\Theta|\mathcal{D})$, in statistical inference [15] (see Methods and Supplementary Figs. 2-3). The posterior in the current setting forms a Markov random field (MRF) [15, 17], which encapsulates the energy band continuity assumption and the measured intensities distribution of photoemission in a probabilistic graphical model. For one benefit, the probabilistic formulation can incorporate imperfect physical knowledge algebraically in the model or numerically as initialization (i.e. warm start, see Methods) of the MAP estimation, without requiring *de facto* ground truth and training as in supervised machine learning [18]. For another, the graphical model representation allows convenient optimization and extension to other dimensions (see Supplementary Fig. 1 and section S1).

To demonstrate the effectiveness of the algorithm, we have reconstructed the entire 3D dispersion surfaces, $E(k_x, k_y)$, of all 14 valence bands within the projected first Brillouin zone, in (k_x, k_y, E) coordinates, of the semiconductor tungsten diselenide (WSe_2), spanning ~ 7 eV in energy and ~ 3 Å^{-1} along each momentum direction. We adapt informatics tools to BS data to sample and compare the reconstructed and theoretical BSs globally. The accuracy of the reconstruction is validated using synthetic data and the extracted local structural parameters in comparison with pointwise fitting. The available data and BS informatics enable detailed comparison of energy dispersion at a resolution of < 0.02 Å^{-1}.

Valence band mapping. The 2D layered semiconductor WSe_2 with 2H interlayer stacking (2H-WSe_2) is a model system for band mapping experiments [20, 21, 12]. Its valence BS contains 14 strongly dispersive energy bands, formed by a mixture of the $5d^4$ and $6s^2$ orbitals of the W atoms and the $4p^4$ orbitals of the Se atoms in its hexagonal unit cell. The strong spin-orbit coupling due to heavy elements produces large momentum- and spin-dependent energy splitting and modifications to the BS [12, 22]. The photoemission band mapping experiment captures the photoelectrons directly in their 3D coordinates, (k_x, k_y, E), by a commercial electron momentum microscope (METIS 1000, SPECS GmbH, see Methods) [5, 6]. Earlier valence band mapping and reconstruction in ARPES experiments on WSe_2 have demonstrated a high degree of similarity between theory and experiments [20, 21, 12], but a quantitative assessment within the entire (projected) Brillouin zone is still lacking. Effects of sample degradation has also been reported [21] during the course of long-duration angular scanning in ARPES measurements.
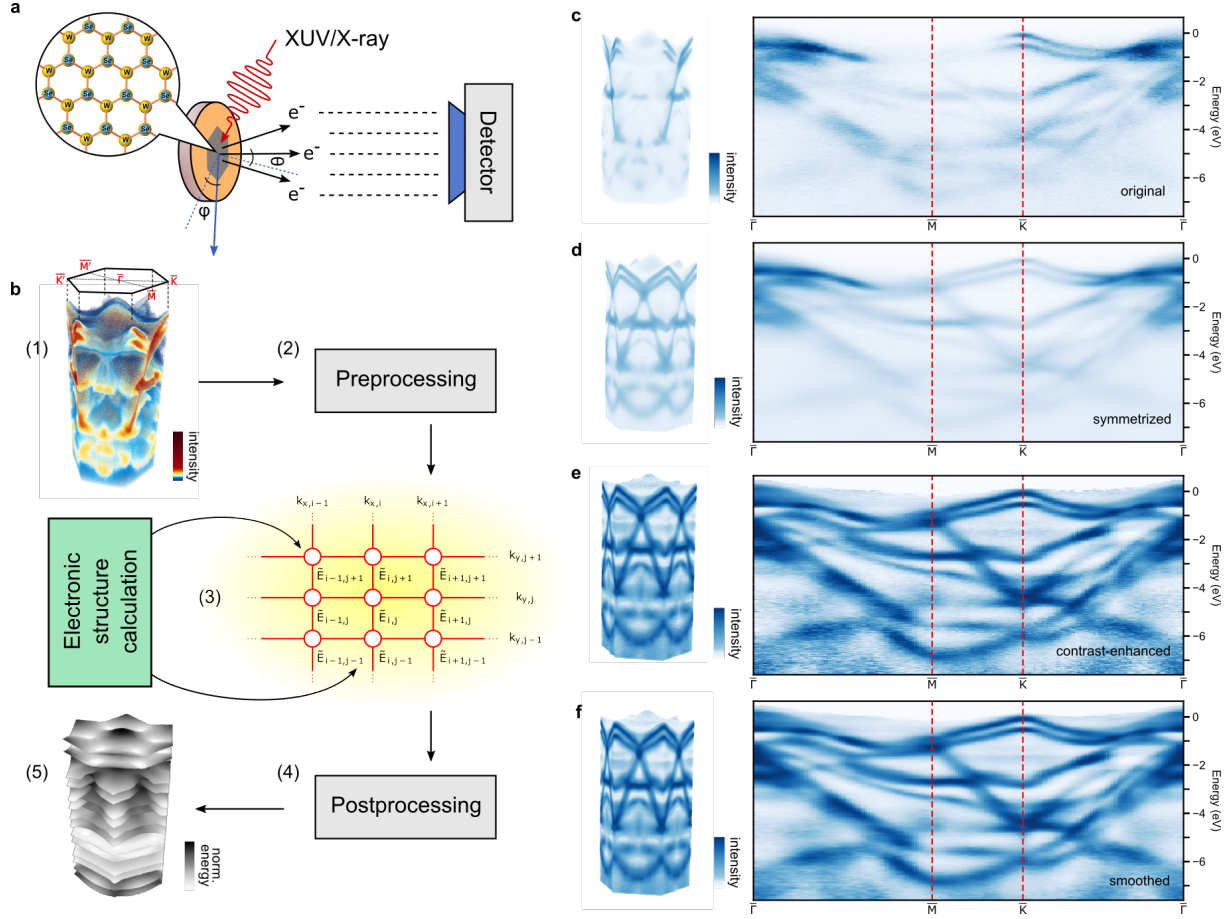
Figure 1: **From band mapping to band structure. a**, Schematic of a photoemission band mapping experiment. The electrons from a crystalline sample's surface are liberated by extreme UV (XUV) or X-ray pulses and collected by a detector through either angular scanning or time-of-flight detection schemes. **b**, Overview of the data-analytical framework for reconstruction of the photoemission (or quasiparticle) band structure: (1) The volumetric data obtained from a band mapping experiment first (2) go through preprocessing steps, then are (3) fed into the probabilistic machine learning algorithm along with electronic structure calculations as initialization of the optimization. The reconstruction algorithm for volumetric band mapping data is represented as a 2D probabilistic graphical model with the band energies as parameters at each node and tens of thousands of nodes in practice. (4) The outcome of the reconstruction is postprocessed (e.g. symmetrization) to (5) yield the dispersion surfaces (i.e. energy bands) of the photoemission band structure ordered by band indices. **c**-**f**, Effects of the intensity transforms in preprocessing viewed in both 3D and along high-symmetry lines of the projected Brillouin zone (see **b**(1)), from the original data (**c**) through intensity symmetrization (**d**), contrast enhancement [19] (**e**) and Gaussian smoothing of intensities (**f**). The intensity data in **c**-**f** are normalized individually for visual comparison.

With our high-repetition-rate photon source [11] and the fast electronics of the momentum microscope, band mapping of WSe$_2$ achieves sufficient signal-to-noise ratio for valence band reconstruction within only tens of minutes of data acquisition, without the need for angular scanning and subsequent reconstruction from momentum-space slices.

Band structure reconstruction and digitization. We use a 2D MRF to model the loci of an energy band within the intensity-valued 3D band mapping data, regarded as a collection of momentum-ordered EDCs. It is graphically represented by a rectangular grid overlaid on the momentum axes with the indices (i, j) (i, j are nonnegative integers), as shown in step (3) of Fig. 1b. The undetermined band energy of the EDC at (i, j), with the associated momentum coordinates $(k_{x,i}, k_{y,j})$, is considered a random variable (or model parameter), $\tilde{E}_{i,j}$, of the MRF. Together, the probabilistic model is characterized by a joint distribution, expressed as the product of the likelihood and the Gaussian prior, in Eq. (1). To maintain its simplicity, we don't explicitly account for the intensity modulations of various origins (such as imbalanced transition matrix elements [23]) in the original band mapping data, which cannot be remediated by upgrading the photon source or detector. Instead, we preprocess the data to minimize their effects on the reconstruction (see Fig. 1c-f). The preprocessing steps include (1) intensity symmetrization, (2) contrast enhancement [19], followed by (3) Gaussian smoothing (see Methods), whereafter the continuity of band-like features is restored. The EDCs from the preprocessed data, $\tilde{I}$, are used effectively as the likelihood to calculate the MRF joint distribution,

$$p(\{\tilde{E}_{i,j}\}) = \frac{1}{Z} \prod_{ij} \tilde{I}(k_{x,i}, k_{y,j}, \tilde{E}_{i,j}) \cdot \prod_{(i,j)(l,m)|\text{NN}} \exp\left[-\frac{(\tilde{E}_{i,j} - \tilde{E}_{l,m})^2}{2\eta^2}\right]. \tag{2}$$

Here, Z is a normalization constant, η is a hyperparameter defining the width of the Gaussian prior, $\prod_{ij}$ denotes the product over all discrete momentum values sampled in the experiment and $\prod_{(i,j)(l,m)|\text{NN}}$ the product over all the NN terms. Detailed derivation of Eq. (2) is given in Supplementary Information section S1. Reconstruction of the bands in the photoemission BS is carried out sequentially and relies on local optimization of the MRF model parameters, $\{\tilde{E}_{i,j}\}$, and operates efficiently on scalable hardware (see Methods). We further inject relevant physical knowledge to correctly resolve band crossings and nearly degenerate energies by using density functional theory (DFT) band structure calculation with semi-local approximation [24] as a starting point for the reconstruction. The calculation qualitatively entails such physical

symmetry information for WSe_2, albeit not quantitatively reproducing the experimental quasi-particle BSs [24] at all momentum coordinates. As shown with four DFT calculations with different exchange-correlation functionals [24] to initiate the reconstruction for WSe_2 and in various cases using synthetic data with known ground truth (see Methods, Supplementary Table 1 and Supplementary Figs. 4-8), the reconstruction algorithm is not particularly sensitive to the initialization as long as the information about band crossings is present.

The reconstructed 14 valence bands of WSe_2 initialized by LDA-level DFT are shown in Fig. 2b-d and Supplementary Videos. To globally compare the computed and reconstructed bands at a consistent resolution, we expand the BS in shape descriptors, such as orthonormal polynomial bases [25], which forms a unified representation of BS data unbiased by the underlying electronic detail. The geometric featurization of band dispersion allows multiscale sampling and comparison using the coefficient (or feature) vectors [26]. Although the choice of basis is not unique, Zernike polynomials (ZPs) are used to decompose the 3D dispersion surfaces (see Fig. 3 and Methods) because of their existing adaptations to various boundary conditions [27]. In Fig. 3a-b, the band dispersions show generally decreasing dependence (seen from the magnitude of coefficients) on basis terms with increasing complexities (see Fig. 3a), and the majority of dispersion is encoded into a subset of the terms (see Fig. 3b). The example in Fig. 3b and more numerical evidence in Supplementary Fig. 10 illustrate the approximation capability of the hexagonal ZPs. Concisely, the coefficient amplitudes and their distribution function as geometric fingerprints of the energy bands to describe their dispersions on a global scale. We can then use similarity or distance metrics (see Methods) for their classification and comparison [26], an approach inspired by content-based image retrieval from databases [28]. For example, in Fig. 3c, the positive cosine similarity confirms the strong shape (or dispersion) resemblance of the 7 pairs of spin-split energy bands in the reconstructed BS of WSe_2, while the low negative values, such as those between bands 1-2 and 13-14, reflect the opposite directions of their respective dispersion (see Fig. 2d). These observations are consistent with the outcome obtained from DFT calculations (see Supplementary Fig. 9).

Moreover, we introduce a BS distance metric (see Methods), invariant to the global energy shift frequently used to adjust the energy zero, to quantify the differences in band dispersion and the relative spacing between bands. The distance is calculated using the geometric feature vectors to bypass interpolation errors while reconciling the coordinate spacing difference

6
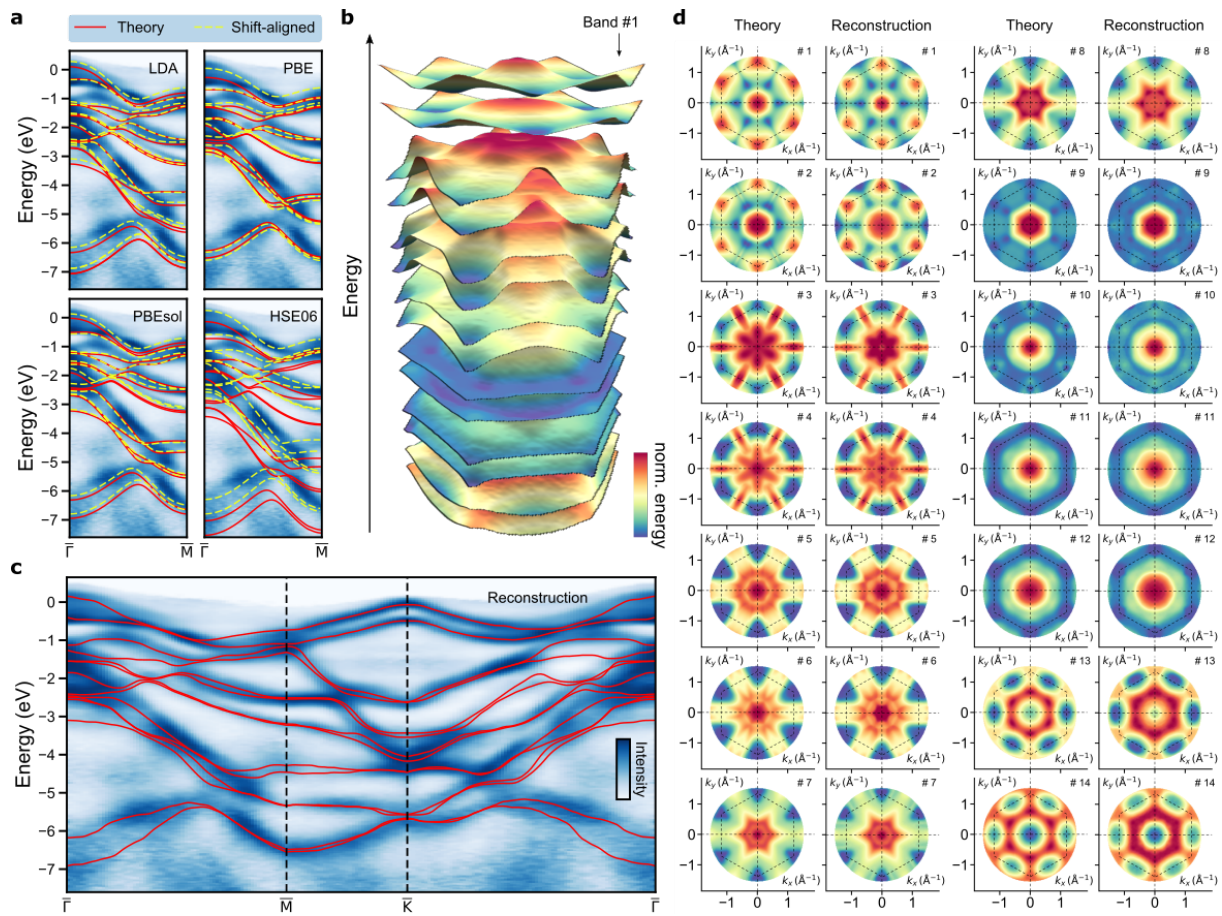
Figure 2: **Reconstructed band structure from WSe$_2$ photoemission data. a,** Comparison between the preprocessed WSe$_2$ valence band photoemission data along $\overline{\Gamma}$-$\overline{M}$ direction, DFT band structure calculated with different exchange-correlation functionals (solid red lines), and their final positions after band-wise rigid-shift alignment (dashed yellow lines) as part of hyperparameter tuning. The energy zero of each DFT calculation is set at the $\overline{K}$ point (see also Supplementary Fig. 4). **b,** Exploded view (with enlarged spacing between bands for better visibility) of reconstructed energy bands of WSe$_2$. **c,** Overlay of reconstructed band dispersion (red lines) on preprocessed photoemission band mapping data cut along the high-symmetry lines in the hexagonal Brillouin zone of WSe$_2$. The residual intensities on the low-energy end are from contrast-amplified background signals unrelated to the band structure. **d,** Band-wise comparison between LDA-level DFT (LDA-DFT) calculation used to initialize the optimization and the reconstructed 14 valence bands of WSe$_2$ (symmetrized in postprocessing). The boundaries of the first Brillouin zone are traced out by the dashed hexagons. The band indices on the upper right corners in **d** follow the ordering of the electronic orbitals in this material obtained from LDA-DFT. The color scale of band energies in **b** and **d** are normalized within each band to improve visibility.

7
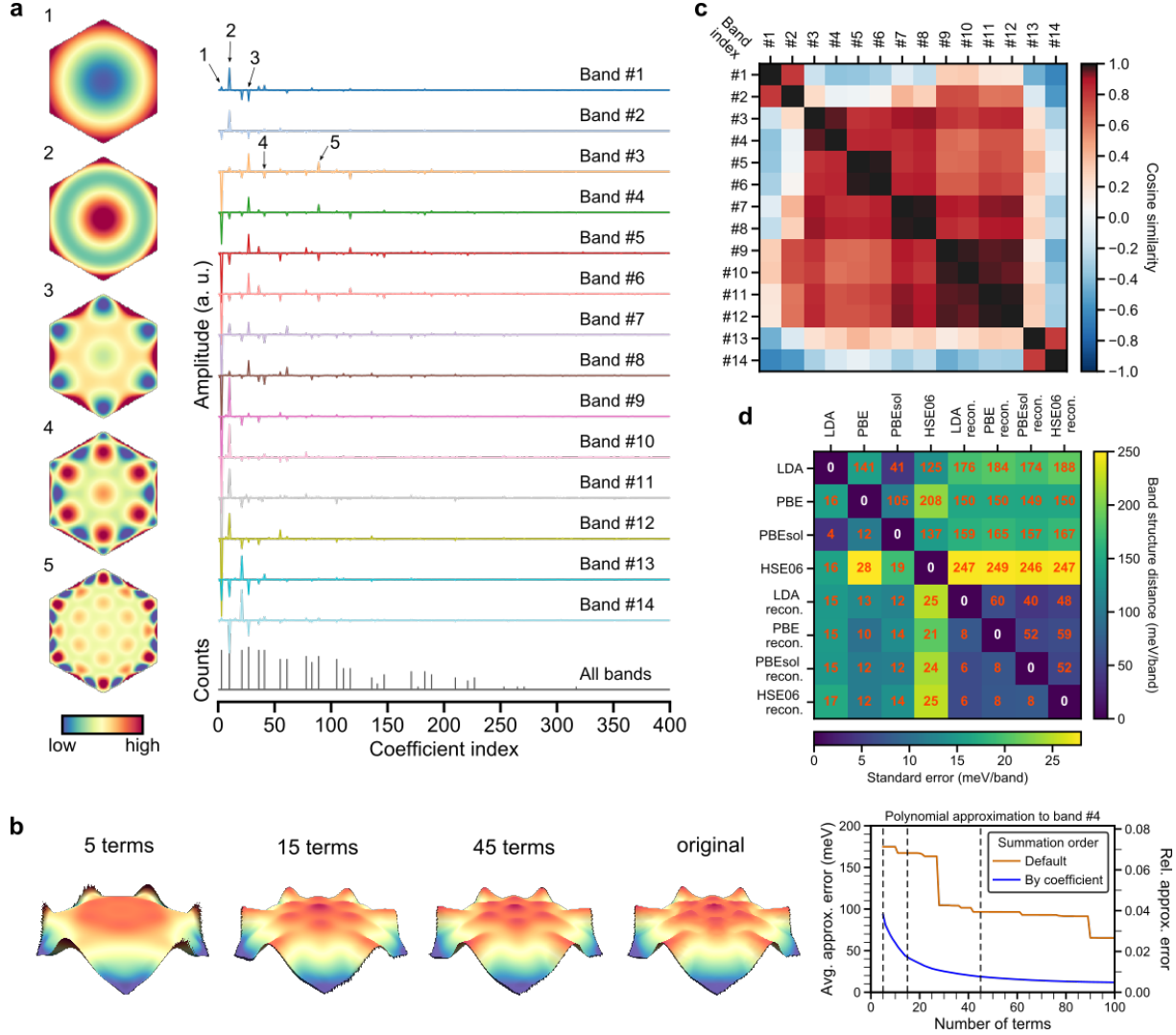
Figure 3: **Digitization and comparison of WSe$_2$ band structures. a**, Decomposition of the 14 energy bands of WSe$_2$ into hexagonal Zernike polynomials (ZPs) with selected major terms displayed on the left. The zero spatial frequency term in the decomposition is subtracted for each band. The counts of large ($> 10^{-2}$ by absolute value) coefficients of all 14 bands are accumulated at the bottom row of the decomposition to illustrate their distribution, which decrease in value towards higher-order terms. **b**, Approximation of the shape (or dispersion) of the fourth energy band using different numbers of hexagonal ZPs. Both of the average and relative approximation errors (see Methods) shown on the far right drop as more terms are included, while summation in the coefficient-ranked order (used to generate the intermediate results in **b**) achieves faster convergence than in the default polynomial order. **c**, Cosine similarity matrix for pairwise comparison of the reconstructed band dispersion in Fig. 2. The band indices follow those in Fig. 2d. **d**, Two-part similarity matrix showing band structure distances (in the upper triangle) and their corresponding standard errors (in the lower triangle) between the computed and reconstructed band structures of WSe$_2$. The abbreviation "LDA recon." denotes reconstruction with LDA-level DFT band structure as the initialization.

8

between reconstructed and theoretical BSs, essential for differentiating BS data from heterogeneous sources in databases. The results in Fig. 3d refer to the valence BS of WSe_2 discussed in this work, where the distances and their spread (i.e. standard errors) are displayed in the upper and lower triangles, respectively. A high degree of consistency exists among the reconstructions (pairwise distance no larger than 60±8 meV/band) regardless of the level of DFT calculation used for initialization, indicating the robustness of the probabilistic reconstruction algorithm, whereas the distances between the DFT calculations are much larger, both in energy shifts and their spread. The current framework can also support the initialization from more advanced electronic structure methods, such as GW [29] or that including electronic self-energies renormalized by electron-phonon coupling [30], when semi-local approximation yields not only quantitatively, but also qualitatively wrong quasiparticle BSs compared with experiment. However, a systematic benchmark of theory and experiment goes beyond the scope of this work. As shown in Fig. 3d and Supplementary Fig. 4, the learning algorithm can bridge the epistemic gap between theories to obtain a consistent reconstruction.

Besides providing global structural information, the reconstruction improves the robustness of traditional region-of-interest lineshape fitting in extended regions of the momentum space (see Supplementary Fig. 11), when used as initial guess. Pointwise fitting in turn acts as refinement of local details not explicitly included in the probabilistic reconstruction model, which prioritizes efficiency. A compendium of local parameters are retrieved using this approach (see Supplementary Table 2). We obtain the trigonal warping parameter of the first two valence bands around $\overline{\mathrm{K}}$-point, 5.8 eV·Å^3 and 3.9 eV·Å^3, respectively, confirming the magnitude difference between these spin-split bands predicted by theory [22]. Fitting around $\overline{\mathrm{M}}'$ (and $\overline{\mathrm{M}}$) reveals that the gap opened by spin-orbit interaction extends beyond the saddle point in the dispersion surface with the minimum gap at 338 meV, markedly larger than DFT results. Overall, the reconstruction yields local structural information consistent with the more laborious pointwise fitting. Finally, our approach reduces the data size by over 5000 times from 3D band mapping data to geometric features vectors (see Methods), facilitating database integration.

Conclusion and outlook. We have formulated band structure reconstruction ubiquitous in photoemission band mapping in an inference perspective and described an efficient reconstruction procedure by combining probabilistic machine learning with the physical knowledge embedded

in electronic structure calculations, as demonstrated for the energy-dispersive, multiband material of WSe_2. The reconstruction reveals global and local structural information challenging to access by pointwise fitting. Our approaches lend valuable insights to automating data analysis in materials characterization and future upgrade into an end-to-end framework balancing physical constraints and computational efficiency to achieve desired accuracy. The reconstruction outcome should assist interpretation of deep-lying bands, parametrizing multiband Hamiltonian models [31], simulation of realistic devices [8], and complement theoretical data in materials science databases. Implementation of the reconstruction across multiple materials and to higher-dimensional data [5], including temperature, photon energy, dynamical time delay and spin as resolved quantities, will generate comprehensive knowledge about the (non)equilibrium electronic structure of materials to benchmark theories. Likewise, the method is transferable to extract the band structure of other quasiparticles (e.g. phonons [32] or polaritons) in periodic systems, given the availability of these alternative band mapping data. The multidisciplinary methodology provides an example for building next-generation high-throughput materials characterization toolkits combining learning algorithms with physical knowledge [33] to arrive at a comprehensive understanding of materials properties unattainable before.

Methods

Band mapping measurements. Photoemission band mapping of WSe_2 using multidimensional photoemission spectroscopy were conducted using a laser-driven, high harmonic generation-based extreme UV light source [11] operating at 21.7 eV and 500 kHz and a METIS 1000 (SPECS GmbH) momentum microscope featuring a delay-line detector coupled to a time-of-flight drift tube [6, 34]. Single crystal samples of WSe_2 ($> 99.995\%$ pure) were purchased from HQ graphene and were used directly for measurements without further purification. Before measurements, the WSe_2 samples were attached to the Cu substrate by conductive epoxy resin (EPO-TEK H20E). The samples were cleaved by cleaving pins attached to the sample surface upon transfer into the measurement chamber, which operates at an ambient pressure of 10^{-11} mbar during photoemission experiments. No effect of surface termination has been observed in the measured WSe_2 photoemission spectra, similar to previous experimental observations [20, 12]. For the valence band mapping experiments, the energy focal plane of the photoelectrons within the time-of-flight drift tube was set close to the top valence band.

Data processing and reconstruction. The raw data, in the form of single-electron events recorded by the delay-line detector, were preprocessed using home-developed software packages [35]. The events were first binned to the (k_x, k_y, E) grid with a size of $256 \times 256 \times 470$ to cover the full valence band range in WSe_2 within the projected Brillouin zone, which amounts to a pixel size of ~ 0.015 Å^{-1} along the momentum axes and ~ 18 meV along the energy axis. The bin sizes are within the limits of the momentum resolution (< 0.01 Å^{-1}) and energy resolution (< 15 meV) of the photoelectron spectrometer [36].

Data binning is carried out in conjunction with the necessary lens distortion correction [37] and calibrations as described in [35]. The outcome provides a sufficient level of granularity in the momentum space to resolve the fine features in band dispersion while achieving higher signal-to-noise ratio than using single-event data directly. Afterwards, we applied intensity symmetrization to the data along the sixfold rotation symmetry and mirror symmetry axes [12] of the photoemission intensity pattern in the (k_x, k_y) coordinates, followed by contrast enhancement using the multidimensional extension of the contrast limited adaptive histogram equalization (MCLAHE) algorithm, where the intensities in the image are transformed by a look-up table built from the normalized cumulative distribution function of local image patches [19]. Finally, we applied Gaussian smoothing to the data along the k_x, k_y and E axes with a standard deviation of 0.8, 0.8 and 1 pixels (or about 0.012 Å^{-1}, 0.012 Å^{-1}, and 18 meV), respectively.

After data preprocessing, we sequentially reconstructed every energy band of WSe_2 from the photoemission data using the *maximum a posteriori* (MAP) approach described in the main text. The reconstruction requires tuning of three hyperparameters: (1) the momentum scaling and (2) the rigid energy shift to coarse-align the computed energy band, e.g. from density functional theory (DFT), to the photoemission data, and (3) the width of the nearest-neighbor Gaussian prior (η in Eq. (2)). The hyperparameter tuning is also carried out individually for each band to adapt to their specific environment. An example of hyperparameter tuning is given in Supplementary Fig. 3. The MAP reconstruction method involves optimization of the model parameters (i.e. band energy random variables), $\{\tilde{E}_{i,j}\}$ to maximize the posterior probability $p = p(\{\tilde{E}_{i,j}\})$ or to minimize the negative log-probability loss function, $\mathcal{L} := -\log p$, obtained from Eq. (2) as is used in our actual implementation.

$$\mathcal{L}(\{\tilde{E}_{i,j}\}) = -\sum_{i,j} \log I(k_{x,i}, k_{y,j}, \tilde{E}_{i,j}) + \sum_{(i,j),(l,m)|\text{NN}} \frac{(\tilde{E}_{i,j} - \tilde{E}_{l,m})^2}{2\eta^2} + \text{const.} \qquad (3)$$

We implemented the optimization using a parallelized version of the iterated conditional mode (ICM) [38] method in Tensorflow [39] in order to run on multicore computing clusters and GPUs. The parallelization involves a checkerboard coloring scheme (or coding method) of the graph nodes [40] and subsequent hierarchical grouping of colored nodes, which allows alternating updates on different subgraphs (i.e. subsets of the nodes) of the Markov random field during optimization. Typically, the optimization process in the reconstruction of one band converges within and therefore is terminated after 100 epochs, which takes ~ 7 seconds on a single NVIDIA GTX980 GPU for the above-mentioned data size. Details on the parallelized implementation are provided in section S1 of the Supplementary Information. In addition, because symmetry information is not explicitly included in the MRF model, the reconstructed bands generally requires further symmetrization as refinement or post-processing to be ready for database integration.

We described our approach of using band structure calculations to initialize the MAP optimization as a warm start. The term "warm start" in the context of numerical optimization generally refers to the initialization of an optimization using the outcome of an associated and yet more solvable problem (e.g. surrogate model) obtained beforehand that yields an approximate answer, instead of starting from scratch (i.e. cold start). Warm-starting an optimization improves the effective use of prior knowledge and its convergence rate [41]. In the current context, we regard the band structure reconstruction from photoemission band mapping data as the optimization problem to warm start, and the outcome from an electronic structure calculation can produce a sufficiently good approximate to the solution of the optimization problem. For WSe$_2$, straightforward DFT calculations with semi-local approximation (which in itself involves explicit optimizations such as geometric optimization of the crystal structures) are sufficient, but our approach is not limited to DFT. Therefore, the use of "warm start" in our application is conceptually well-aligned with the origin of the term.

To validate the MAP reconstruction algorithm in a variety of scenarios, we used synthetic photoemission data where the nominal ground-truth band structures are available. The band structures are constructed using analytic functions, model Hamiltonians or DFT calculations. The initializations are generated by tuning the numerical parameters used to generate the ground-truth band structures. The procedures and results are presented in section S2 of the Supplementary Information. In simple cases, such as single or well-isolated bands, the reconstruction

yields a close solution to the ground truth even with a flat band initialization. In the more general multiband scenario with congested bands and band crossings (or anti-crossings), an approximate dispersion (or shape) of the band and the crossing information is required in the initialization (i.e. warm start) in order to converge to a realistic solution. We further tested the robustness of the initializations by (1) scaling the energies of the ground truth and by (2) using DFT calculations with different exchange-correlation (XC) functionals, in order to capture sufficient variability of available band structure calculations in the real world. We quantify the variations in the initializations and the performance of the reconstruction using the average error (Eq. (8), or Fig. 3b), calculated with respect to the ground truth. Among the different numerical experiments, we find that the optimization converges consistently to a set of bands that better matches the experimental data than the initialization. This is manifest in that the average errors of the initializations are reduced to a similar level in the corresponding reconstruction outcomes, a trend seen over all bands regardless of their dispersion. In the synthetic data with an energy spacing of $\sim$ 18 meV, the average error in the reconstruction is on the order of 40-50 meV for each band, which amounts to an average inaccuracy of $<$ 3 bins along the energy dimension at a momentum location. The inaccuracy is, however, dependent on the bin sizes used in the preprocessing and the fundamental resolution in the experiment. We have made the code for the MAP reconstruction algorithm and the synthetic data generation publicly accessible from the online repository fuller [42] for broader applications.

Band structure calculations. Electronic band structures were calculated within (generalized) DFT using the local density approximation (LDA) [43, 44], the generalized-gradient approximation (GGA-PBE) [45] and GGA-PBEsol [46]), and the hybrid XC functional HSE06 [47], which incorporates a fraction of the exact exchange. All calculations were performed with the all-electron, full-potential numeric-atomic orbital code, FHI-aims [48]. They were conducted for the geometries obtained by fully relaxing the atomic structure with the respective XC-functional to keep the electronic and atomic structures consistent. Spin-orbit coupling was included in a perturbational fashion [49]. The momentum grid used for the calculation was equally sampled with a spacing of 0.012 $\mathring{\text{A}}^{-1}$ in both k_x and k_y directions that covers the irreducible part of the first Brillouin zone at k_z = 0.35 $\mathring{\text{A}}^{-1}$, estimated using the inner potential of WSE$_2$ from a previous measurement [12]. The calculated band structure is symmetrized to fill the entire hexagonal Brillouin zone to be used to initialize the band structure reconstruction and

synthetic data generation. We note here that for the MAP reconstruction, the momentum grid size used in theoretical calculations (such as DFT at various levels used here) need not be identical to that of the data (or instrument resolution) and in those cases an appropriate upsampling (or downsampling) should be applied to the calculation to match their momentum resolution. Further details are presented in section S3 of the Supplementary Information.

Band structure informatics. The shape feature space representation of each electronic band is derived from the decomposition,

$$E_b(\mathbf{k}) = \sum_l a_l \phi_l(\mathbf{k}) = \mathbf{a} \cdot \Phi \tag{4}$$

Here, $\mathbf{k} = (k_x, k_y)$ represents the momentum coordinate, $E_b(\mathbf{k})$ is the single-band dispersion relation (e.g. dispersion surface in 3D), a_l and $\phi_l(\mathbf{k})$ are the coefficient and its associated basis term, respectively. They are grouped separately into the feature vector, $\mathbf{a} = (a_1, a_2, ...)$, and the basis vector, $\Phi = (\phi_1, \phi_2, ...)$. The orthonormality of the basis is guaranteed within the projected Brillouin zone (PBZ) of the material.

$$\int_{\mathbf{k} \in \Omega_{\text{PBZ}}} \phi_m(\mathbf{k}) \phi_n(\mathbf{k}) \, d\mathbf{k} = \delta_{mn} \tag{5}$$

For the hexagonal PBZ of WSe_2, the basis terms are hexagonal Zernike polynomials (ZPs) constructed using a linear combination of the circular ZPs via Gram-Schmidt orthonormalization within a regular (i.e. equilateral and equiangular) hexagon [27]. A similar method can be used to generate ZP-derived orthonormal basis adapted to other boundary conditions [27]. The representation in feature space [26] provides a way to quantify the difference (or distance) d between energy bands or band structures at different resolutions or scales without additional interpolation. To quantify the shape similarity between energy bands E_b and $E_{b'}$, we calculate the cosine similarity using the feature vectors,

$$d_{\cos}(E_b, E_{b'}) = \frac{\mathbf{a} \cdot \mathbf{a}'}{|\mathbf{a}| \cdot |\mathbf{a}'|}, \tag{6}$$

The cosine similarity is bounded within $[-1, 1]$, with a value of 0 describing orthogonality of the feature vectors and a value of 1 and -1 describing parallel and anti-parallel relations between them, respectively, both indicating high similarity. The use of cosine similarity in feature space allows comparison of dispersion while being unaffected by their magnitudes. In comparing the

dispersion between single energy bands using Eq. (6), the first term in the polynomial expansion, or the hexagonal equivalent of the Zernike piston [50], is discarded as it only represents a constant energy offset (with zero spatial frequency) instead of dispersion, which is characterized by a combination of finite and nonzero spatial frequencies.

The electronic band structure is a collection of energy bands $E_B = \{E_{b_i}\}$ ($i = 1, 2, ...$). To quantify the distance between two band structures, $E_{B_1} = \{E_{b_{1,i}}\}$ and $E_{B_2} = \{E_{b_{2,i}}\}$, containing the same number of energy bands while ignoring their global energy difference, we first subtract the energy grand mean (i.e. mean of the energy means of all bands within the region of the band structure for comparison). Then, we compute the Euclidean distance, or the ℓ^2-norm, for the ith pair of bands, $d_{b,i}$.

$$d_{b,i}(E_{b_{1,i}}, E_{b_{2,i}}) = \|\tilde{\mathbf{a}}_{1,i} - \tilde{\mathbf{a}}_{2,i}\|_2 = \sqrt{\sum_l (\tilde{a}_{1,il} - \tilde{a}_{2,il})^2}. \tag{7}$$

Here, $\tilde{\mathbf{a}}$ denotes the feature vector after subtracting the energy grand mean so that any global energy shift is removed. We define the band structure distance as the average distance over all N_b pairs of bands, or $d_B(E_{B_1}, E_{B_2}) = \sum_i^{N_b} d_{b,i}(E_{b_{1,i}}, E_{b_{2,i}})/N_b$. The values of $d_B(E_{B_1}, E_{B_2})$ are shown in the upper triangle of Fig. 3d and their corresponding standard errors (over the 14 valence bands of WSe$_2$) in the lower triangle. The distance in Eq. (7) is independent of basis and allows energy bands calculated on different resolutions or from different materials with the same symmetry (e.g. differing only by Brillouin zone size) to be compared.

We use same-resolution error metrics to evaluate the approximation quality of the expansion basis and to quantify the reconstruction outcome with a known ground-truth band structure. Specifically, we define the average approximation error (with energy unit), η_{avg}, for each energy band using the energy difference at every momentum location,

$$\eta_{\mathrm{avg}}(E_{\mathrm{approx}}, E_{\mathrm{recon}}) = \sqrt{\frac{1}{N_k} \sum_{\mathbf{k} \in \Omega_{\mathrm{PBZ}}} (E_{\mathrm{approx},\mathbf{k}} - E_{\mathrm{recon},\mathbf{k}})^2}, \tag{8}$$

where N_k is the number of momentum grid points and the summation runs over the projected Brillouin zone. In addition, we construct the relative approximation error, η_{rel}, following the definition of the normwise error [51] in matrix computation,

$$\eta_{\mathrm{rel}}(E_{\mathrm{approx}}, E_{\mathrm{recon}}) = \frac{\|E_{\mathrm{approx}} - E_{\mathrm{recon}}\|_2}{\|E_{\mathrm{recon}}\|_2}. \tag{9}$$

Eq. (8)-(9) are used to compute the curves in Fig. 3b as a function of the number of basis terms included in the approximation. The relevant code for the representation using hexagonal ZPs and the computation of the metrics is also accessible in the public repository fuller [42].

Data reduction. The raw data and intermediate results are stored in the HDF5 format [35]. The file sizes quoted here for reference are calculated from storage as double-precision floats or integers (for indices). The photoemission band mapping data of WSe_2 ($256 \times 256 \times 470$ bins) have a size of about 235 MB (240646 kB) after binning from single-event data (7.8 GB or 8176788 kB). The reconstructed valence bands at the same resolution occupy about 3 MB (3352 kB) in storage, and the size further decreases to 46 kB when we store the shape feature vector associated with each band. If only the top-100 coefficient (ranked by the absolute values of their amplitudes) and their indices in the feature vectors are stored, the data amounts to 24 kB. For the case of WSe_2, the top-100 coefficients can approximate the band dispersion with a relative error (see Eq. (9)) of $< 0.8\%$ for every energy band, as shown in Supplementary Fig. 10.

Acknowledgments

Authors contributions

R.P.X. and R.E. conceived the project. S.D. and Sa.B. performed the photoemission band mapping experiment. M.Z., M.D. and C.C. performed the DFT band structure calculations.

R.P.X. processed the raw data, devised the band structure digitization and algorithm validation schemes. V.S. designed and implemented the machine learning algorithm under the supervision of St.B. and B.S. with inputs from R.P.X.. R.P.X., V.S. and M.Z. co-wrote the first draft of the manuscript. All authors contributed to discussion and revision of the manuscript to its final version.

Data and code availability

The electronic structure calculations are available from the NOMAD repository (DOI: 10.17172/NOMAD/2020.03.28-1). The photoemission dataset used in this work will be made available on Zenodo. The code developed for this work will be made available on GitHub at https://github.com/mpes-kit/fuller.

Competing interests

The authors declare no competing interests in the content of the article.

Supplementary Information

A machine learning route between band mapping and band structure

S1 Band structure reconstruction

Physical foundations. The three quantities of common interest for the interpretation of photoemission spectra are (1) the bare band energy, $\epsilon_{\mathbf{k}}$, (2) the complex-valued electron self-energy, $\Sigma(\mathbf{k}, E) = \mathrm{Re}\Sigma(\mathbf{k}, E) + i\mathrm{Im}\Sigma(\mathbf{k}, E)$, and (3) the transition matrix elements connecting the final (f) and initial (i) electronic states, $M_{f,i}(\mathbf{k}, E)$. An established interface between theory and experiment for quantitating and interpreting the photoemission signal is the formalism of an experimental observable—the single-particle spectral function [52, 9], $A(\mathbf{k}, E)$. For a single energy band of a many-body electronic system,

$$A(\mathbf{k}, E) = \frac{1}{\pi} \frac{\mathrm{Im}\Sigma(\mathbf{k}, E)}{[E - \epsilon_{\mathbf{k}} - \mathrm{Re}\Sigma(\mathbf{k}, E)]^2 + [\mathrm{Im}\Sigma(\mathbf{k}, E)]^2}. \tag{10}$$

Within this framework, the band loci of the photoemission (or quasiparticle) band structure (BS), $b(\mathbf{k}, E) = \epsilon_{\mathbf{k}} + \mathrm{Re}\Sigma(\mathbf{k}, E)$, correspond to the bare band dispersion modulated by the real part of the electron self-energy, and they occupy the local maxima of the spectral function evaluated at different momenta. However, in the photoemission process, the intensity counts registered by the detector are modulated by the transition matrix elements [23], the Fermi-Dirac occupation function, $f_{\mathrm{FD}}(E)$, and the resolution of the measuring instrument, $G(E, \sigma_E, \sigma_{\mathbf{k}})$, typically a multidimensional Gaussian function. This leads to the expression of the photoemission intensity, $I(\mathbf{k}, E)$, registered on an energy- and momentum-resolved detector,

$$I(\mathbf{k}, E) \propto |M_{f,i}(\mathbf{k}, E)|^2 f_{\mathrm{FD}}(E) A(\mathbf{k}, E) \otimes G(E, \sigma_E, \sigma_{\mathbf{k}}). \tag{11}$$

For a multiband electronic structure, band mapping measurements, in principle, have access to the spectral functions of at least all valence bands. The photoemission intensities are combined

in summation to form the multiband (MB) counterpart of the single-band formula.

$$I_{\text{MB}}(\mathbf{k}, E) = \sum_j I_j(\mathbf{k}, E) \propto \sum_j |M_{f_j, i_j}(\mathbf{k}, E)|^2 f_{\text{FD}}(E) A_j(\mathbf{k}, E) \otimes G(E, \sigma_E, \sigma_{\mathbf{k}}) \tag{12}$$

$$\sim \sum_j A_j(\mathbf{k}, E) \otimes G(E, \sigma_E, \sigma_{\mathbf{k}}), \quad (\text{when} |M_{f_j, i_j}(\mathbf{k}, E)| \to 1, f_{\text{FD}}(E) \to 1). \tag{13}$$

The condition $f_{\text{FD}}(E) \to 1$ applies to valence bands, while $|M_{f_j, i_j}(\mathbf{k}, E)| \to 1$ may be achieved through nonlinear intensity normalization or contrast enhancement in data processing. The expression of the multiband photoemission intensity in Eqs. (12)-(13) provides the physical foundation and inspiration for the approximate generation of band mapping data (see Section S2) that we employ to validate the reconstruction algorithm introduced in this work.

Markov random field modeling. The Markov random field (MRF) model for the photoemission band structure in photoemission band mapping data can be constructed similarly for data in multiple dimensions. In traditional angle-resolved photoemission spectroscopy (ARPES), photoemission intensities are measured in the (k, E) coordinates, the proximity of the momentum positions in the band structure can be modeled using an MRF composed of a 1D chain of random variables as shown in Supplementary Fig. 1a. Band mapping data in (k_x, k_y, E) coordinates, as described in the main text, can be modelled using a 2D MRF. In addition, the algorithm can be extended to higher dimensions involving coordinates beyond energy and momenta. For example, time-resolved photoemission data recorded in (k_x, k_y, E, t) coordinates can be modelled using a 3D MRF as shown in Supplementary Fig. 1c. In the following, we provide a brief introduction to the theory underlying MRF and provide a simplified derivation of the 2D MRF model introduced in the main text.

Deriving the MRF amounts to determining the joint distribution of the random variables associated with its graphical representation. In graphical model theory [53], a graph is constructed from the fundamental components called cliques. Each clique C of a graph is a subset of nodes that shares an edge with another node in C, with the total number of nodes in C defined as its size. The MRFs in Supplementary Fig. 1a-c that model the photoemission data are built out of cliques of sizes 1–2 shown in Supplementary Fig. 1d. Although larger cliques can be constructed similarly [53], their parent graphical models are described by more complex joint distributions with drastically higher computational costs in optimization, therefore are not used in our MRFs. Mathematically, each clique is represented by a so-called potential function, ψ_C,

which is used to derive the joint distribution that characterizes the MRF. The potential function only depends on the node configuration in the cliques, $\mathbf{X}_C$, and satisfies $\psi_C(\mathbf{X}_C) \geqslant 0$. According to the Hammersley-Clifford theorem [54, 55, 53], the joint distribution of a vector of random variables, $\mathbf{X}$, can be written in the factorized form,

$$p(\mathbf{X}) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(\mathbf{X}_C). \tag{14}$$

Here, $\mathcal{C}$ is the set of all cliques in the graph, and the partition function Z is a normalization constant given by

$$Z = \sum_{\mathbf{X}} \prod_{C \in \mathcal{C}} \psi_C(\mathbf{X}_C).$$

The graphical representation of the MRFs relevant to this work are rectangular grids shown in
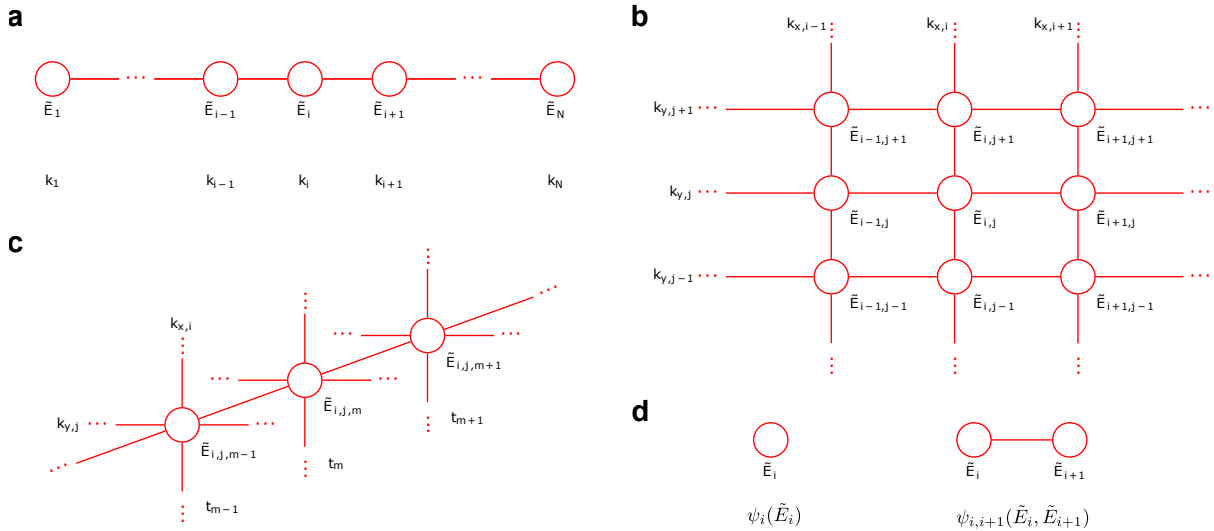
Supplementary Figure 1: **Examples of the MRF models for photoemission spectroscopy data**. **a**, 1D MRF model for data in (k, E) coordinates, represented as a chain of random variables $\tilde{E}_i$. N is the number of measured momentum values. **b**, 2D MRF model of photoemission data in (k_x, k_y, E) coordinates as introduced and demonstrated for use in the main text, with the random variables $\tilde{E}_{i,j}$ connected on two dimensions k_x and k_y. **c**, 3D MRF model for time- and momentum-resolved photoemission spectroscopy data in (k_x, k_y, E, t) coordinates. The random variables $\tilde{E}_{i,j,m}$ are first connected in the graph to the neighboring momentum positions as in **b**, then subsequently along the neighboring time points. The time variable in **c** may also be replaced with other variables without changes in the structure of the graphical model. In **a-c**, the MRFs are constructed using components (cliques) with sizes 1 (left) and 2 (right) in **d**, with their respective potential functions written below the illustrations.

Supplementary Fig. 1. The respective potential functions of the size-1 and size-2 cliques are

interpreted as the likelihood and prior of the probabilistic graphical model, respectively. To cast the band structure reconstruction problem into this framework, we assign the band energies as the random variables (or model parameters) in the model, and the potential function of each node (size-1 clique) as the (preprocessed) photoemission intensity at the respective grid position. For simplicity and computational efficiency, this formulation doesn't explicitly account for the intensity modulations described in Eq. (11) and preprocessing steps are required to neutralize their effects. The continuity assumption (i.e. no sharp jump) of the band energies along momentum directions means that the potential function of size-2 cliques can be represented by a Gaussian on adjacent momentum grid positions. Intuitively, this means that the closer the two adjacent energies is, the more probable they are the actual band loci, and *vice versa*.

In the 1D case (see Supplementary Fig. 1a), the potential function of each node (containing one band energy random variable $\tilde{E}_i$) is given by

$$\psi_i(\tilde{E}_i) = \tilde{I}(k_i, \tilde{E}_i), \tag{15}$$

where $\tilde{I}$ is the photoemission intensity after preprocessing. The potential function of two connected nodes (describing the similarity between two neighboring band energy random variables) is given by

$$\psi_{j,j+1}(\tilde{E}_j, \tilde{E}_{j+1}) = \exp\left[-\frac{(\tilde{E}_j - \tilde{E}_{j+1})^2}{2\eta^2}\right]. \tag{16}$$

Plugging Eqs. (15)-(16) into Eq. (14) yields

$$
\begin{aligned}
p(\tilde{E}_1, ..., \tilde{E}_N) &= \frac{1}{Z} \prod_{i=1}^{N} \psi_i(\tilde{E}_i) \cdot \prod_{j=1}^{N-1} \psi_{j,j+1}(\tilde{E}_j, \tilde{E}_{j+1}) \\
&= \frac{1}{Z} \prod_{i=1}^{N} \tilde{I}(k_i, \tilde{E}_i) \cdot \prod_{j=1}^{N-1} \exp\left[-\frac{\left(\tilde{E}_j - \tilde{E}_{j+1}\right)^2}{2\eta^2}\right]
\end{aligned}
\tag{17}
$$

as the joint distribution of the 1D MRF, with N being the total number of momentum grid points. Analogously, we can derive the joint distribution of the 2D MRF as given in the main text, and that for the 3D MRF in the (k_x, k_y, E, t) coordinates is

$$p(\{\tilde{E}_{i,j,m}\}) = \frac{1}{Z} \prod_{i,j,m} \tilde{I}(k_{x,i}, k_{y,j}, t_m, \tilde{E}_{i,j,m}) \cdot \prod_{(i,j,m),(l,o,q)|\mathrm{NN}} \exp\left[-\frac{(\tilde{E}_{i,j,m} - \tilde{E}_{l,o,q})^2}{2\eta^2}\right].$$

The MRF models in different dimensions discussed here follow the same Bayesian interpretation as the 2D MRF (Eq. (1) in the main text).

21

Optimization procedure. Optimization of the MRF model is a local minima-finding process [53]. The following procedures are described using the 2D MRF in the main text as an example, but the approach can be extended to arbitrary dimensions. Due to the large number of random variables ($\sim 10^4$ for the 2D MRF in the main text) and their complex dependence structure in the MRF, we solved it numerically using iterated conditional mode (ICM) [38] procedure and implemented with efficient parallelization schemes, including the coding method and the hierarchical grouping of random variables. Next, we discuss the motivations and clarify the details of these three aspects. We provide the associated pseudocode in Algorithm 1.

(1) ICM: Originally developed for similar optimization problems arising in image denoising [56, 57, 53], ICM is applicable to optimizing MRF at any dimension. The ICM procedure includes (i) initialization of the random variables (e.g. $\{\tilde{E}_{i,j}\}$ in 2D MRF) and (ii) selection of a single random variable to optimize in the loss function $\mathcal{L}$ while keeping all the other random variables fixed. Each round in (ii) requires to compute at most five terms in the loss (Eq. (3) in the main text Methods section) which depend on the selected random variable $\tilde{E}_{i,j}$. We can simply evaluate these terms at the energy axis values measured in the experiment to determine the energy associated with the lowest loss. (iii) iterate over all other random variables using the same procedure in (ii).

(2) Coding method: The ICM procedure described above operates sequentially over every $\tilde{E}_{i,j}$, which is inefficient for the MAP optimization involving a large number of parameters. To improve the optimization performance, we implement the ICM with a checkerboard parallelization scheme (or coding method) [40] that scales favorably on multicore computing clusters. The scheme assigns the nodes of the MRF alternately with white and black colors, as shown in Supplementary Fig. 2a. If the white nodes are blocked, the black nodes are no longer connected through paths (i.e. sequences of connected edges and nodes). This property is called d-separation [58, 53]. Analogously, blocking the black nodes d-separates the white nodes. Since the MRF models satisfy the Hammersley-Clifford theorem [54], d-separation is equivalent to conditional independence, meaning that the random variables represented by the black nodes are independent if we condition on those represented by the white nodes. Therefore, conditioning on the nodes of one color allows us to compute the terms in the log-probability loss (Eq. (3) in main text Methods) that depends on the nodes of another color in parallel, which means that the nodes associated with different colors can be updated alternately. Further details and proofs
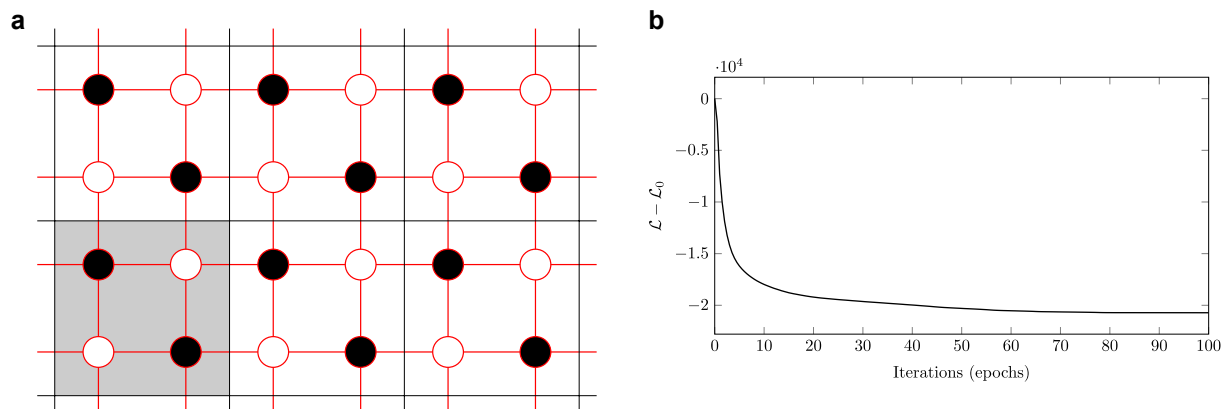
Supplementary Figure 2: **Numerical optimization of the MRF model**. **a**, Schematic of the checkerboard parallelization (or coding method) and hierarchical grouping schemes for speeding up the ICM. The nodes of the MRF are alternately colored white and black (checkerboard parallelization) and each set of four neighboring nodes are group into a unit as colored in grey (hierarchical grouping). The updates in optimization are carried out first at the four-node unit level, then alternately on the white or black nodes within the units. **b**, An example loss curve for reconstructing the second valence band of WSe_2 using the 2D MRF model and parallelized ICM implementation. $\mathcal{L}_0$ is the initial value of the loss at the start of the optimization. Within an epoch in the parallelized scheme, the white nodes and subsequently the black nodes are separately updated, therefore each band energy random variable is effectively updated once. The loss decreases rapidly in the beginning and reaches a minimum after about 90 epochs.

related to the coding method have been elaborated in [59, 55].

(3) Hierarchical grouping: The introduction of the checkerboard parallelization scheme reduces the translation symmetry of the original graph (originally symmetric by translation of an arbitrary number of nodes, now only symmetric by a translation of two nodes in each direction), which complicates the matrix operations needed to update the loss. However, we can restore the translation symmetry and carry out the computation on a higher level by grouping a set of four neighboring nodes into a unit, as illustrated in Supplementary Fig. 2a. In this way, updating the loss requires only standard matrix operations at the unit level followed by consecutive updates of the nodes within the units. During the optimization, the loss is updated by two sets of operations concerning (i) the nearest neighbor nodes within the unit (line 18-19 in Algorithm 1) and (ii) the nearest neighbor nodes of the neighboring unit (line 20-21 in Algorithm 1). The latter operations are carried out by shifting the higher-level rectangular grid formed by the units by one step vertically or horizontally, followed by an operation on nodes of the respective units

---

Algorithm 1 Optimization procedure for reconstructing a single energy band.

Input: I (3D momentum-resolved photoemission data), E_0 (2D initialization from density functional theory calculation), E (1D energy axis)

Parameter: η (hyperparameter of the Markov random field), N (number of epochs)

Output: E_{rec} (Reconstructed 2D energy band)

    # Initialize the momentum index grid for an energy band

 1: size_kx, size_ky, size_E = size(I)

 2: ind_x, ind_y = meshgrid(range(size_kx, step=2), range(size_ky, step=2))

    # Divide data into four-node units. $E_u(i,j,...)$, $I_u(i,j,...)$ are the band energies and

    # photoemission intensities for the node (i,j) in a unit (u) in Supplementary Fig. 2,

    respectively

 3: **for** i **in** [0, 1] **do**

 4:     **for** j **in** [0, 1] **do**

 5:         $E_u[i, j, :, :] = E_0[\text{ind\_x} + i, \text{ind\_y} + j]$

 6:         $\text{log\_}I_u[i, j, :, :, :] = \log(I[\text{ind\_x} + i, \text{ind\_y} + j, :])$

    # Iterative optimization of energy values

 7: **for** n **in** range(N) **do**

    # Update white nodes

 8:     $E_u[0, 0, :, :] = \text{update\_E}(0, 0, \text{log\_}I_u, E_u, E)$

 9:     $E_u[1, 1, :, :] = \text{update\_E}(1, 1, \text{log\_}I_u, E_u, E)$

    # Update black nodes

10:     $E_u[0, 1, :, :] = \text{update\_E}(0, 1, \text{log\_}I_u, E_u, E)$

11:     $E_u[1, 0, :, :] = \text{update\_E}(1, 0, \text{log\_}I_u, E_u, E)$

    # Assemble reconstruction from all nodes in the units

12: **for** i **in** [0, 1] **do**

13:     **for** j **in** [0, 1] **do**

14:         $E_{rec}[\text{ind\_x} + i, \text{ind\_y} + j] = E_u[i, j, :, :]$

    # Function to update the energy of the element (i, j) within a four-node unit

15: **function** UPDATE_E(i, j, log_I_u, E_u, E)

    # Calculate the difference between current and all possible energies

16:     $\text{squ\_diff} = (E_u - E) ** 2 / (2 * \eta ** 2)$

    # Calculate all possible $\log p$ values, start with log-likelihood

17:     $\text{log\_p} = \text{log\_}I_u[i, j, :, :, :]$

    # Substract by energy differences from nearest neighbor nodes within unit

18:     $\text{log\_p} \mathrel{-}= \text{squ\_diff}[(i + 1) \% 2, j, :, :, :]$

19:     $\text{log\_p} \mathrel{-}= \text{squ\_diff}[i, (j + 1) \% 2, :, :, :]$

    # Substract by energy differences from nearest neighbor nodes of the neighboring unit

20:     $\text{log\_p} \mathrel{-}= \text{shift}(\text{squ\_diff}[(i + 1) \% 2, j, :, :, :], 2 * i - 1, \text{axis=2})$

21:     $\text{log\_p} \mathrel{-}= \text{shift}(\text{squ\_diff}[i, (j + 1) \% 2, :, :, :], 2 * j - 1, \text{axis=3})$

    # Return optimal energy values

22:   **return** E[argmax(log_p)]

---

of the original and the shifted grid. The procedure is implemented in the open-source fuller package [42] using Tensorflow [39]. Supplementary Fig. 2b shows an example loss curve (i.e. loss as a function of iteration) in reconstruction of an energy band, where the optimization is essentially complete within $\sim$ 90 iterations.

(4) Robust initialization: Since the current MRF model doesn't include any explicit regularization on the outcome with respect to the initialization, the optimizer is free to explore a large range of values. In other words, the initial band dispersion is able to freely deform to fit to the band loci embedded in the data. This design improves the robustness of the algorithm to initialization. As a result, in scenarios with only non-crossing energy bands, the MAP optimization can simply be initialized with constant energy values to yield consistent results. In general situations involving band crossings, the optimization procedure requires an initialization with approximate energy values that preserves the band-crossing information, such as those provided by electronic structure calculations. In this scenario, the robustness of the algorithm is manifest in the fact that it can tolerate a certain amount of deviation in the initialization and still converges to a satisfactory reconstruction, which, in realistic settings, is closer to the real band structure contained in photoemission data than the initialization (e.g. from electronic structure calculations). Quantitative examples demonstrating the robustness of initialization are provided using synthetic data in Supplementary Figs. 5-6 (see Section S2).

Hyperparameter tuning. The optimization process in the band structure reconstruction involves the tuning of three kinds of hyperparameters, which are the momentum scaling parameter, the rigid energy shift and the width of the nearest-neighbor Gaussian prior.

(1) Momentum scaling: applied to equalize the momentum scale and resolution between the BS calculation (e.g. conducted on relaxed unit cells, see Supplementary Table 1) and the experimental data (measured on real materials). In our reconstruction procedure, the scaling factor is fixed in the reconstruction of all energy bands using a particular level of density functional theory (DFT) calculation as initialization.

(2) Rigid energy shift (ΔE): separately applied to each energy band in the calculated BS to coarse-align to the band mapping data. In our case, the shift is chosen manually by visual inspection of the theoretical band energies overplotted on photoemission data (usually in the energy-momentum slices). In practice, the necessary energy shifts vary between bands and also
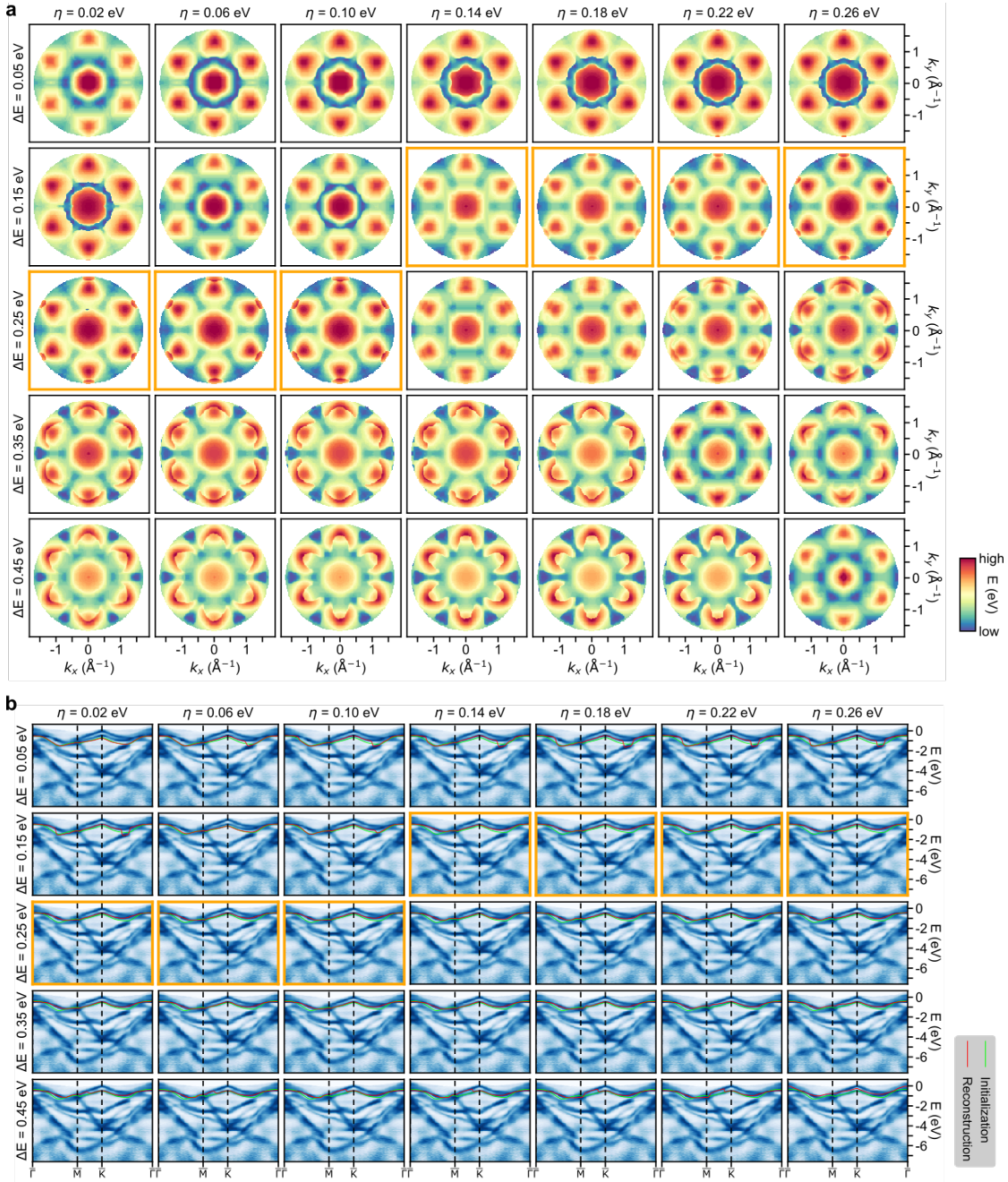
Supplementary Figure 3: **Demonstration of hyperparameter tuning**. An example of tuning the hyperparameters, the rigid shift (ΔE) and the width of the nearest-neighbor Gaussian prior (η), for reconstructing the second valence band of WSe_2. **a**, Evolution of reconstructed energy band during hyperparameter tuning. **b**, Evolution of the initialization and reconstructed band along high-symmetry directions of the hexagonal lattice of WSe_2. The energy bands are overlaid on top of preprocessed data from photoemission band mapping of WSe_2 (Fig. 1f in the main text). In **a**,**b**, the images showing the optimal region for the hyperparameters identified by the scientists are emphasized with orange-colored frames.

26

depend on level of approximation in the BS calculation used as initialization, as illustrated in Fig. 2a of the main text.

(3) Width of the nearest-neighbor Gaussian prior (η): The value of the parameter η is chosen manually from an initial estimate and subsequently optimized by visual inspection of the reconstruction outcome. In the case of WSe$_2$, the momentum grid of the experimental data has a spacing of $\Delta k_x = \Delta k_y \approx 0.015$ Å^{-1}, we used $\eta \in [0.05, 0.2]$ eV. Generally speaking, the initial estimate of η has the order of magnitude proportional to the momentum grid spacing times the dispersion due to the following argument: To obtain a consistent reconstruction, we expect the posterior to stay relatively constant and be independent of the momentum grid spacing, which should be sufficiently fine to ensure band continuity. Since after preprocessing the data, the intensity (i.e. the likelihood) is normalized and stays constant with respect to the momentum grid spacing, the nearest-neighbor Gaussian prior term should stay constant correspondingly. For example, for two nearest-neighbor energy variables along the k_x axis, the reasoning above requires,

$$\text{const} \approx \frac{(\tilde{E}_{i+1,j} - \tilde{E}_{i,j})^2}{\eta^2} \approx \left(\frac{\partial E}{\partial k_x}\right)^2 \frac{\Delta k_x^2}{\eta^2}. \tag{18}$$

Thereby, we obtain $\eta \propto \frac{\partial E}{\partial k_x} \Delta k_x$, which provide an order-of-magnitude estimate of η. The same lines of reasoning apply to the k_y axis, for detector systems with relatively constant momentum resolution. As the grid spacing is the same in both k_x and k_y directions, a single η is used for reconstructing each band in the case of WSe$_2$, but the best η differs somewhat between energy bands due to their various amounts of dispersion and how they are connected to the neighboring bands (i.e. their environment), hence the range of η as specified earlier.

To demonstrate the process of hyperparameter tuning, we provide an example showing the reconstruction of the second valence band of WSe$_2$ (see Supplementary Fig. 3), visualized in the top view of the reconstruction outcome and in the momentum path along high-symmetry lines of the projected Brillouin zone. The orange-framed subfigures represent the range of hyperparameter settings that yield a good reconstruction. In practice, the rule-of-thumb is that the choice of hyperparameters is more flexible for reconstructing more isolated bands or those with less number of crossings, and *vice versa*.

Reconstructions using different theories as initializations. Comparison between reconstructed and theoretical band structures for 2H-WSe$_2$ are presented as a similarity matrix in the main
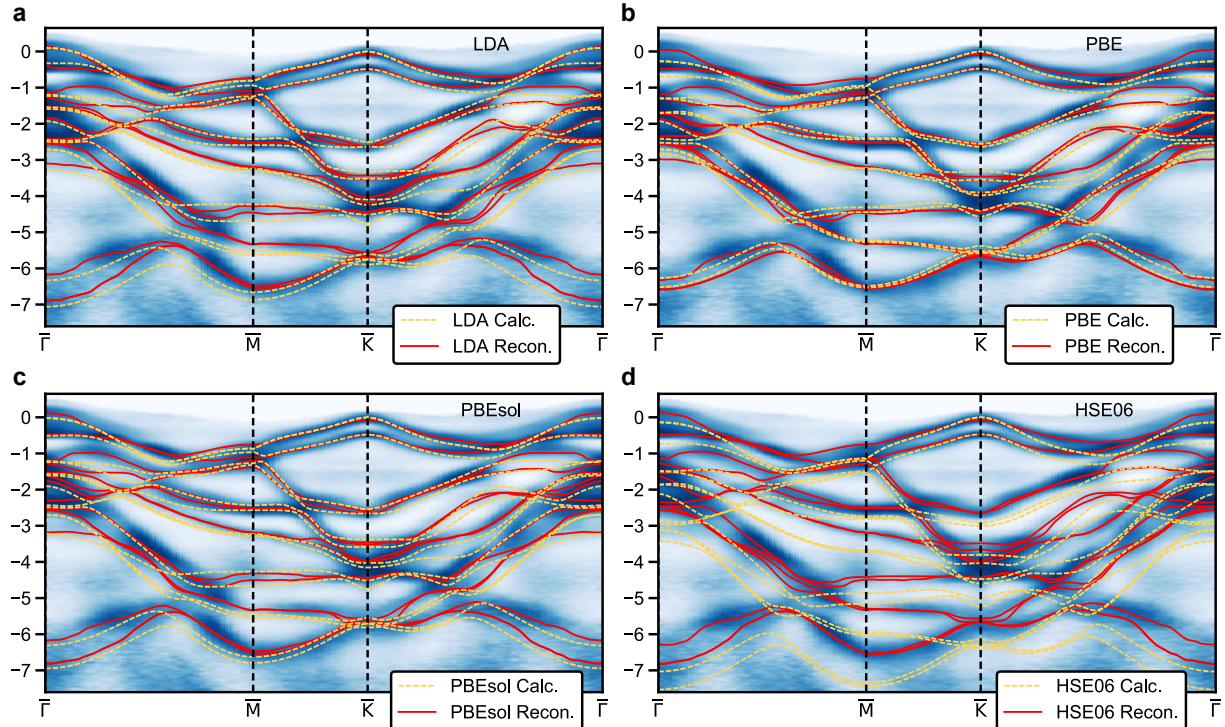
Supplementary Figure 4: **Band structure reconstructions with different theory initializations**. Comparisons between reconstructed photoemission band structures (abbreviated as recon.) and calculated band structures (abbreviated as calc.) from density functional theory (DFT) with different exchange-correlation functionals, including **a,** local density approximation (LDA); **b,** PBE generalized gradient approximation (GGA); **c,** PBEsol GGA; **d,** HSE06 hybrid functional. For each set of DFT band structure, the same energy shift (as in Supplementary Fig. 8) is applied globally to all bands to align the energy zero at the $\overline{\mathrm{K}}$ point with the reconstruction.

text. To provide more intuitive visual guidance in interpreting the BS distance metric used in constructing the similarity matrix, we compare these band structures along the high-symmetry lines of the Brillouin zone in Supplementary Fig. 4.

S2   Generation of and validation on synthetic data

The advantage of using synthetic data is that the underlying band structure (i.e. ground truth) is exactly known such that they can be used for benchmarking the performance of the MAP reconstruction algorithm described in this work. Benchmarking includes numerical experiments on two interrelated aspects: (1) testing the robustness of the reconstruction algorithm using different initializations and comparing the deviations of the outcome from the ground-truth; (2)

testing the accuracy of reconstruction by determining the closest-possible reconstruction outcome from a given initialization. In the following, we first describe the workflow of generating the band structure, the photoemission data and the initializations, which provide all essential components to carry out the tests. Then we present the benchmarking results on various cases.

Generation of band structure data. We have adopted two approaches to generate band structure data to meet the needs for testing the reconstruction algorithm. Firstly, we used analytic functions to describe the band dispersion (see Supplementary Fig. 5). They are computationally efficient, contain tunable parameters, can be produced at any resolution, and are easily extendable to higher dimensions. In 2D momentum space, we constructed a multi-sinusoidal band and two double-crossing parabolic bands. In 3D momentum space, we constructed a scaled version of the strongly oscillating second-order Griewank function [60] and the tight-binding formulation of the two-band graphene band structure [61] as model band dispersion surfaces. The modified Griewank function takes the form,

$$E_{\text{griewank}}(k_x, k_y) = \frac{1}{16000}(k_x^2 + k_y^2) - \cos(2k_x)\cos(\sqrt{2}k_y).$$ (19)

The two-band tight-binding model of graphene has the energy dispersion relations,

$$E_{\pm}(k_x, k_y) = \pm\sqrt{3 + 2\cos\left(\sqrt{3}k_y a\right) + 4\cos\left(\frac{\sqrt{3}}{2}k_y a\right)\cos\left(\frac{3}{2}k_x a\right)}.$$ (20)

Here, E_+ and E_- refer to the conduction band and the valence band, respectively. Secondly, we used numerical band structures from DFT calculations with different exchange-correlation functionals (see section S3). They are more physically realistic, but also require more computation to obtain than generating bands from analytic functions.

Tuning the initialization. For simple bands constructed using analytic functions, tuning can be achieved by modifying the parameters in the functions. In the complex multiband situation such as that of WSe$_2$, we tuned the initialization of the reconstruction algorithm by scaling or perturbing the coefficient amplitudes of the constituent bases of the band structure. In our case, the bases are the terms of the hexagonal Zernike polynomials (ZPs) [62, 27]. Although unconstrained basis tuning is prone to unrealistic results, it achieves a level of ad hoc control for efficient generation of a large amount of distinct initializations. For more physically realistic

29
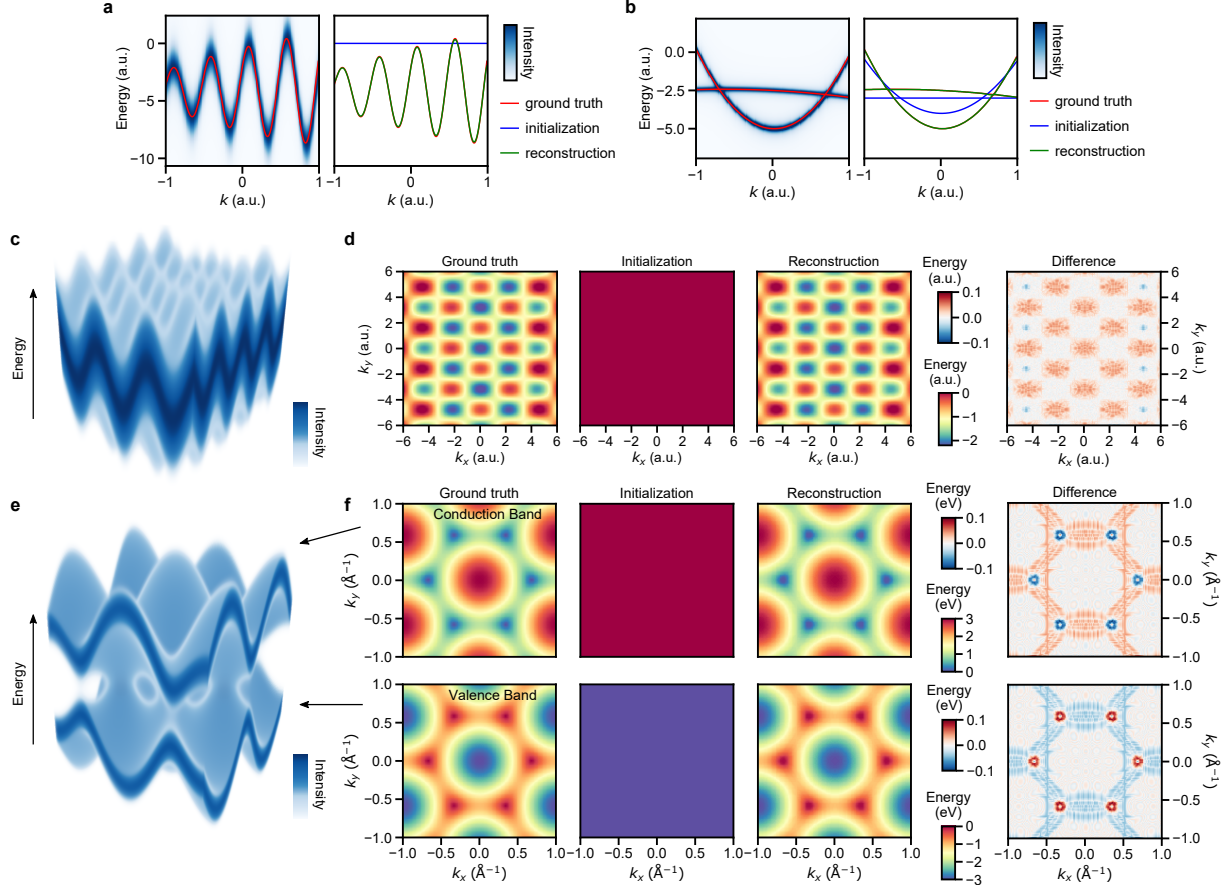
Supplementary Figure 5: **Validations on 2D and 3D synthetic data**. Test results for the reconstruction algorithm on band structures generated with analytic functions. **a**, Reconstruction of a multi-sinusoidal band. **b**, Reconstruction of two double-crossing parabolic bands. **c,d**, Reconstruction of a multi-extrema band with dispersion following the second-order Griewank function (see Eq. (19)) [60]. **e,f**, Reconstruction of the two bands of graphene nearby its Fermi level (**e,f**) formulated in the tight-binding model (see Eq. (20)) [61]. The volumetric renderings in **c,e**, display the synthetic data. The initialization for the reconstruction in **a** is a flat line, while 2D flat bands are used to initialize the cases in **d,f**. In **b**, two double-crossing curves are needed as initialization to preserve the crossing in the reconstruction. The values in the difference plots in **d,f** are calculated by subtracting the ground-truth band energies from the reconstructed ones.

tuning, we used DFT calculations with different exchange-correlation functionals (see section S3).

Approximate generation of photoemission data. We approximately synthesized momentum-resolved photoemission data for each energy band by plugging the band energy and linewidth parameter at each momentum position into the Voigt profile [63] (with Gaussian and Lorentzian
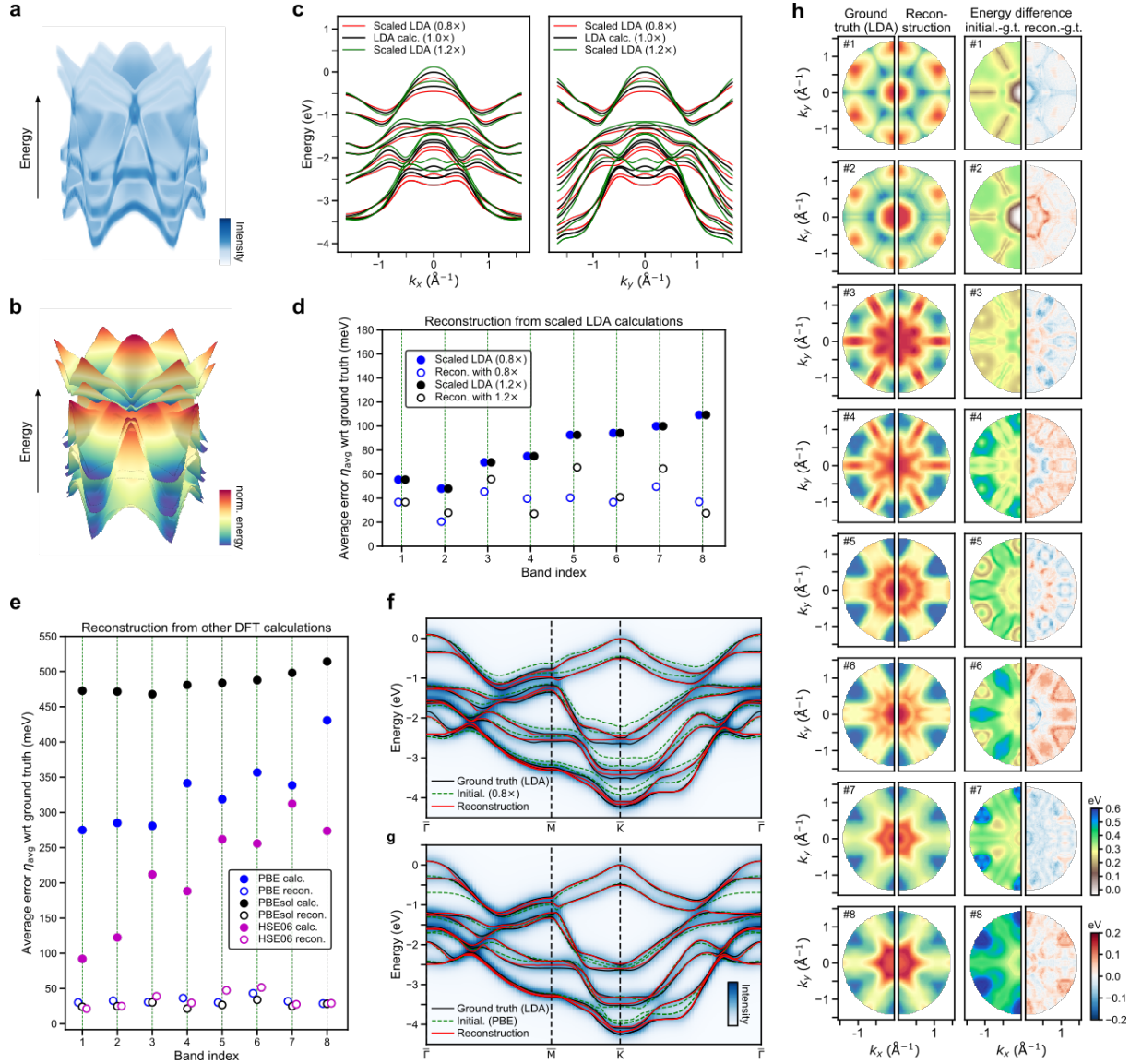
Supplementary Figure 6: **Validation on 3D synthetic multiband photoemission data**. **a**, Synthetic photoemission data with **b**, the underlying band structure obtained from LDA-level DFT calculation of WSe_2 (only the first 8 valence bands are used here). **c**, Comparison of two sets of differently scaled (by 0.8 and 1.2 times, respectively) initial conditions with respect to the ground-truth band structure calculation (LDA calc.), shown for a k_x-E (left) and a k_y-E (right) slice. **d**,**e**, Comparison of the average error η_{avg} for energy bands used as initializations (solid dots) and reconstructions (hollow dots). The initializations are constructed by scaling the ground-truth band energies (**d**) or by using other DFT calculations (**e**). The reconstructions all have reduced η_{avg} compared with the initialization and η_{avg} is consistent across all energy bands. **f**,**g**, Reconstruction, ground truth (LDA), and initialization overlaid on the synthetic data along high-symmetry lines of the hexagonal Brillouin zone, corresponding to two of the cases in **d** and **e**, respectively. The energy zeros of the initialization in **d**-**e** are aligned with the ground truth via a global shift. **h**, Comparisons of ground truth (LDA), reconstructed bands, and the differences between initialization (PBE), reconstruction and ground truth (g.t.) for each energy band.

31

parameters σ and γ, and amplitude B) computed using the Faddeeva function W [64]. The Voigt profile approximates the convolution of a single-particle spectral function (see Section S1), describing the photoemission observable, with a Gaussian energy resolution function. The synthetic photoemission intensity, I_{synth}, for a band structure composed of a set of energy bands, $E_B = \{E_{b_i}\}$, is generated by combining multiple Voigt profiles in summation, similar to Eqs. (12)-(13).

$$I_{\text{synth}}(k_x, k_y, E) = \sum_j \frac{B_j(k_x, k_y)}{\sigma_j \sqrt{2\pi}} \text{Re} \left[W \left(\frac{E - E_{b_j}(k_x, k_y) + \mathrm{i}\gamma_j(k_x, k_y)}{\sigma_j \sqrt{2}} \right) \right] \quad (21)$$

Without loss of generality, we assume the energy resolution in detection for all bands to be the same ($\sigma_j = \sigma$). For the cases shown in Supplementary Figs. 5-6, the linewidth parameter γ are set to a constant throughout the band. In all synthetic data, we omitted the inhomogeneous intensity modifications in realistic photoemission data due to experimental factors such as the experimental geometry, sample condition, matrix element effect, photon energy, etc. This omittance relies on the assumption that the essential preprocessing step, such as symmetrization and contrast enhancement [19] in our workflow (see main text Methods), can sufficiently restore the intensity continuity along the energy bands. The momentum resolution effect is also not accounted for because the instrument (such as METIS 1000 [6, 36]) has a higher momentum resolution than the momentum spacing used in data binning or generation.

Validation of the reconstruction algorithm. Using synthetic data generated from analytic functions of varying complexities as the band structure, we test out the accuracy of reconstruction algorithm (see Supplementary Fig. 5); Using synthetic multiband data generated from the LDA-level DFT (LDA-DFT) band structures of WSe_2 (see section S3), we tested out the sensitivity of reconstruction to the initialization (see Supplementary Fig. 6). In this case, to capture sufficient physical realism similar to the photoemission band mapping of WSe_2 presented in the main text, we set the energy resolution parameter of $\sigma = 100$ meV, the lineshape parameter $\gamma = 50$ meV [65], and the energy spacing of data to ~ 18 meV, identical to the energy bin size for the experimental data. The tests include four sets of numerical experiments summarized below:

(1) Reconstructing non-crossing bands: For isolated bands, we tested synthetic data constructed from a multi-sinusoidal band (Supplementary Fig. 5a), the band generated by the Griewank function (Supplementary Fig. 5c-d), and the two-band tight-binding model of graphene (Sup-

plementary Fig. 5e-f). In these cases, initialization with a flat band without any initial knowledge of the band dispersion (i.e. cold start) is sufficient to recover its shape, regardless of the complexity of the dispersion.

(2) Reconstructing crossing bands: We tested the simplest case of crossing bands with two parabolas of opposite directions of opening (Supplementary Fig. 5b). To recover the dispersion without band index scrambling, the knowledge of crossing needs to be included numerically in the initialization. This means, operationally, that the initialization requires crossing bands at nearby energy values, or that the reconstruction needs a warm-start optimization. For the intercrossing parabolas, the initializations that yield feasible outcomes are generated by slight tuning of the parabola parameters in the range that retains the crossing.

(3) Sensitivity of reconstruction to scaled energies as initialization: We scaled the energies of the LDA-DFT band structure of WSe_2 (using the first 8 valence bands) around the mean energy of each band (see Supplementary Fig. 6c) to for use as the initialization. The accuracy of the reconstruction outcome is evaluated by its average error η_{avg} (Eq. (8) in the main text Methods), calculated with respect to the ground-truth band energies. The results displayed in Supplementary Fig. 6d,f show that the average error and its spread in the reconstruction are reduced from the corresponding values in the initialization. Quantitatively, in the reconstruction, η_{avg} is within the range 20-65 meV, while in the initialization, η_{avg} varies within 45-100 meV for all 8 valence bands.

(4) Sensitivity of reconstruction to differently calculated band structures as initialization: We used DFT band structure calculations of WSe_2 with PBE, PBEsol and HSE06 exchange-correlation functionals (see section S3) to initialize the reconstruction. The accuracy of the reconstruction is quantified similarly as in the previous numerical experiment using η_{avg}. The results displayed in Supplementary Fig. 6e,g,h show that, despite the huge spread in the average error for the different levels of DFT calculations (used as initialization without global shift alignment of energy zero), the corresponding reconstructions all have average errors at around or below 40 meV for every band. The value of η_{avg} varies by up to ~ 30 meV (i.e. between band #1 and #6) in each set of reconstructed bands, much lower than those in the initialization.

The results of the numerical experiments demonstrate that the reconstruction by MAP optimization converges to a consistent range in the tested scenarios and initializations. It should be

noted here that the fundamental accuracy in reconstruction reported here is still limited by the coordinate spacings of the data along all dimensions and the discrete nature of the MRF model (the output is centered only at the bin locations). The former can be improved by casting the experimental data into finer bins in the preprocessing stage (single-electron events can be binned into various sizes) or interpolating between existing bins, while the latter can be improved by using a continuous probabilistic model [53] to formulate the reconstruction problem, albeit at the cost of much increased computational demand.

S3 Band structure calculations

DFT calculations. The crystal structure of bulk WSe_2 with 2H stacking ($2H$-WSe_2) belongs to the $P6_3/mmc$ space group and consists of two Se-W-Se triatomic layers as shown in Supplementary Fig. 7. The stacking order of the two hexagonal layers is $-BAB-ABA-$ and the long c-axis is oriented perpendicular to the layers. Electronic structure calculations were performed within DFT using the local density approximation (LDA), the generalized-gradient approximation (GGA-PBE and GGA-PBEsol), and hybrid (HSE06) exchange-correlation functionals as implemented in FHI-aims [48]. The atomic orbitals basis sets, the integration grids and Hartree potential employed for all calculations are according to the default "tight" numerical settings of FHI-aims. A $16\times16\times4$ uniform **k**-gird was used to sample the Brillouin zone. The Broyden-Fletcher-Goldfarb-Shanno optimization algorithm was used to relax the atomic positions until the residual force component per atom was less than 10^{-2} eV/Å. Supplementary Table 1 shows the optimized lattice constants, a and c, as obtained by the evaluation of the analytical stress tensor [66] using different exchange-correlation functionals. In all BS calculations, we included the effect of spin-orbit coupling, which is known to introduce a large splitting of the outermost valence states of bulk $2H$-WSe_2 [67].

The calculated BSs of bulk $2H$-WSe_2 using different levels of approximation for the exchange-correlation (XC) functional are shown in Supplementary Fig. 8. For each XC functional, the calculations were performed on (1) fully optimized structures (black lines), and on (2) optimized structures by fixing the lattice parameters of the unit cell to the experimental values (colored lines). All calculations using different XC functionals reveal an indirect band gap with the conduction band minimum located along the Γ-K path (Γ and K being the bulk equivalents
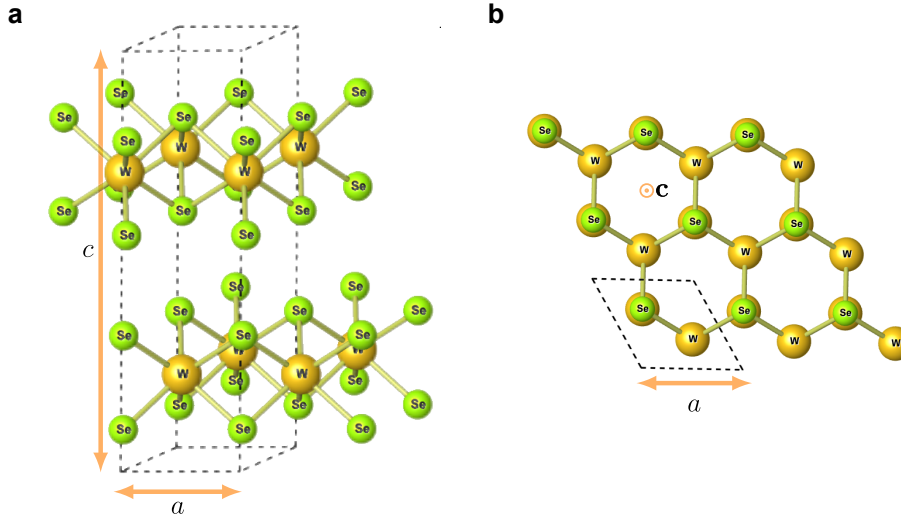
Supplementary Figure 7: **Crystal structure of bulk 2H-WSe$_2$. a**, Side view and **b**, top view of the crystal structure of 2H-WSe$_2$. The space group of the hexagonal structure is P6$_3$/mmc with the c-axis oriented perpendicular to the stacking layers. In each case, the real-space unit cell is labelled by dashed black lines.

Supplementary Table 1: **Parameters from density functional theory calculations.** Optimized lattice constants, spin-orbit splitting of the topmost valence states at the K high-symmetry point, and the band gap of bulk 2H-WSe$_2$ calculated within density functional theory using the LDA, PBE, PBEsol and HSE06 exchange-correlation functionals. For comparison, we also report the corresponding experimental values at room temperature.

| xc-functional	LDA	PBE	PBEsol	HSE06	Experiment
a (Å)	3.250	3.317	3.269	3.295	3.28
c (Å)	12.827	14.921	13.211	13.863	12.98
Spin-orbit splitting at K (eV)	0.485 [1]	0.473 [1]	0.476 [1]	0.467 [1]	0.5 [3]
	0.490 [2]	0.481 [2]	0.484 [2]	0.480 [2]	
Band gap (eV)	1.022 [1]	1.186 [1]	1.105 [1]	1.679 [1]	1.219 [4]
	1.052 [2]	1.074 [2]	1.060 [2]	1.582 [2]	

[1] Fully optimized structure.
[2] Optimized structure by fixing the lattice parameters to experimental values.
[3] Ref. [12].
[4] Ref. [68].

of the $\overline{\Gamma}$ and $\overline{K}$ high-symmetry points). For both sets of optimized structures, the LDA results reveal a valence band maximum at the Γ point, compatible with experimental measurements, while the PBE, PBEsol, and HSE06 band structures obtained for fully optimized structures exhibit a valence band maximum at the K point. Nevertheless, fixing the unit cell dimensions at the experimental lattice constants reproduces the experimental behavior that the valence band
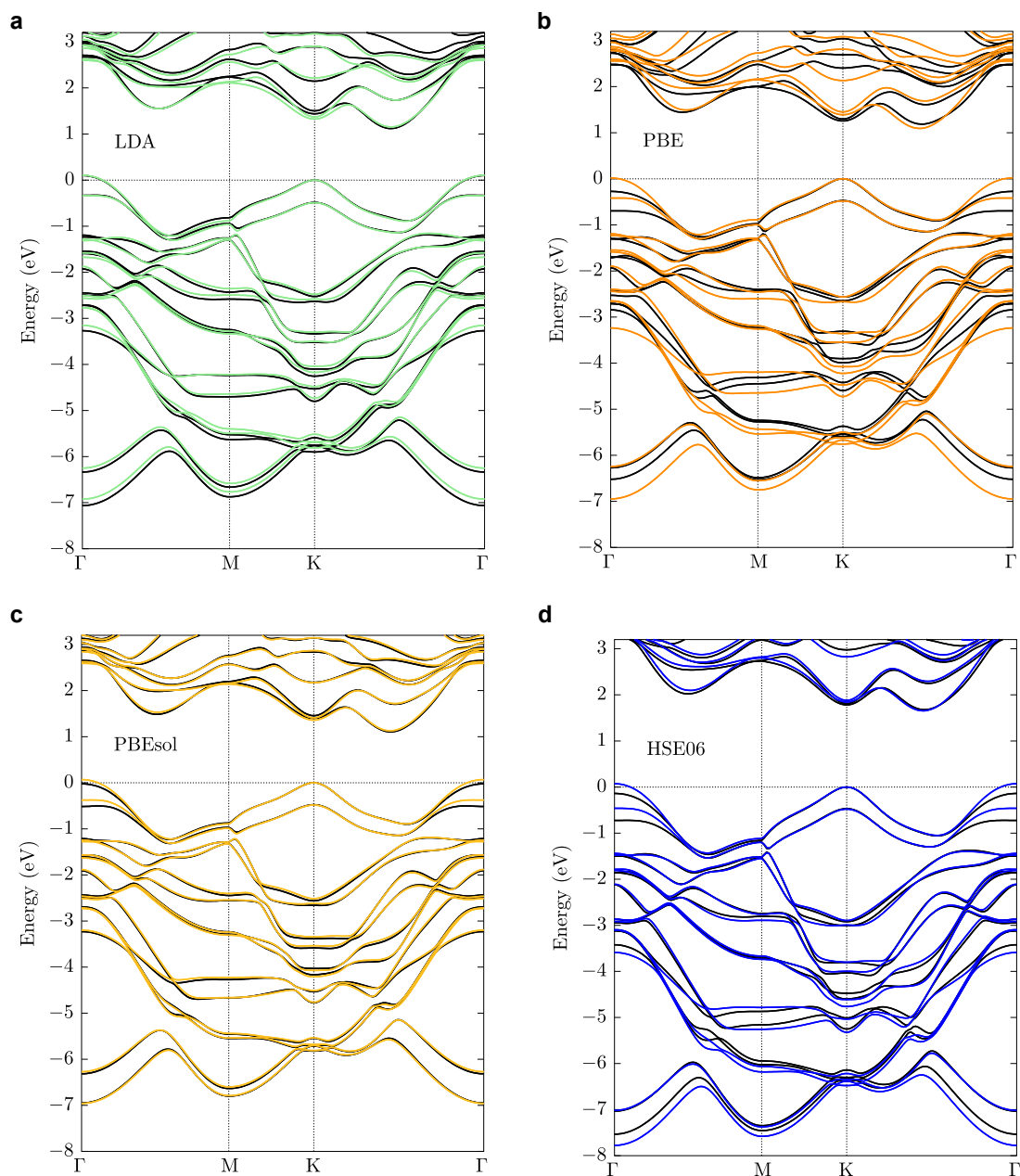
Supplementary Figure 8: **Bulk electronic band structure of 2H-WSe$_2$**. **a-d**, Band structure of bulk 2H-WSe$_2$ along the Γ-K-M-Γ momentum path of its Brillouin zone including the effect of spin-orbit coupling. Calculations were performed using the LDA (green, **a**), PBE (orange, **b**), PBEsol (yellow, **c**), and HSE06 (blue, **d**) exchange-correlation functionals and optimized structures (see Supplementary Table 1) with the unit cell dimensions kept fixed at the experimental lattice constants. Black lines in **a-d** represent the corresponding calculations using fully optimized geometries. For comparison, the two band structures in each plot are rigidly shifted to align their uppermost valence state at the K high-symmetry point, where we also define as the energy zero. All band structure calculations used $k_z = 0.35$ Å^{-1}.

maximum resides at the Γ point. The difference between the two sets of calculations obtained using PBE, PBEsol, and HSE06 functionals is attributed to the overestimation of the lattice parameter c and the residual strain along the c-axis [69]. The calculated indirect band gaps and the spin-orbit splitting of the two topmost valence states at the K point using both sets of optimized structures are shown in Supplementary Table 1.

Brillouin zone tiling. Generation of a large and densely sampled patch of energy bands covering the first Brillouin zone and beyond is crucial for initialization of the MRF model. To balance the computational cost using different XC functionals with the dense sampling similar to the experimental data grid, we used the symmetry properties of the Brillouin zone to tile the calculated momentum-space rectangular patch that covers the Γ, K and M points of the Brillouin zone. The hexagonal Brillouin zone of WSe_2 has a sixfold rotation symmetry axis and two independent mirror planes in the (k_x, k_y) coordinates. The initial rectangular patch is first symmetrized about the two mirror planes in the Γ-K and Γ-M directions to form a larger patch, which is then rotated by $60°$ and $120°$, respectively, and combined with the original mirror-symmetrized patch. The composite patch is then shifted along all six Γ-M directions by one unit cell distance and the result is cut to the required shape compatible with photoemission data.

S4   Band structure informatics

Global scale structure. Unbiased approximators allow us to use informatics tools for data retrieval, representation and comparison for entire bands. We extend the examples given in Fig. 3 of the main text to other bands and band structures used in the present work. Supplementary Fig. 9 displays the band-wise comparison of dispersion surfaces within other DFT calculations. These results contain similar features as Fig. 3a and 3c in the main text, reaffirming that the geometric featurization provides a sparse representation of the band dispersions and that the dispersion similarities are largely preserved despite the use of different exchange-correlation functionals in the DFT calculations. They may, therefore, be regarded as general features of the band structure of WSe_2.

In Supplementary Fig. 10, we demonstrate numerically the approximation capability of the hexagonal ZP basis set to all 14 valence bands of WSe_2. Despite the stark differences in energy dispersion, the approximation to reconstructed bands (Supplementary Fig. 10a-d) and
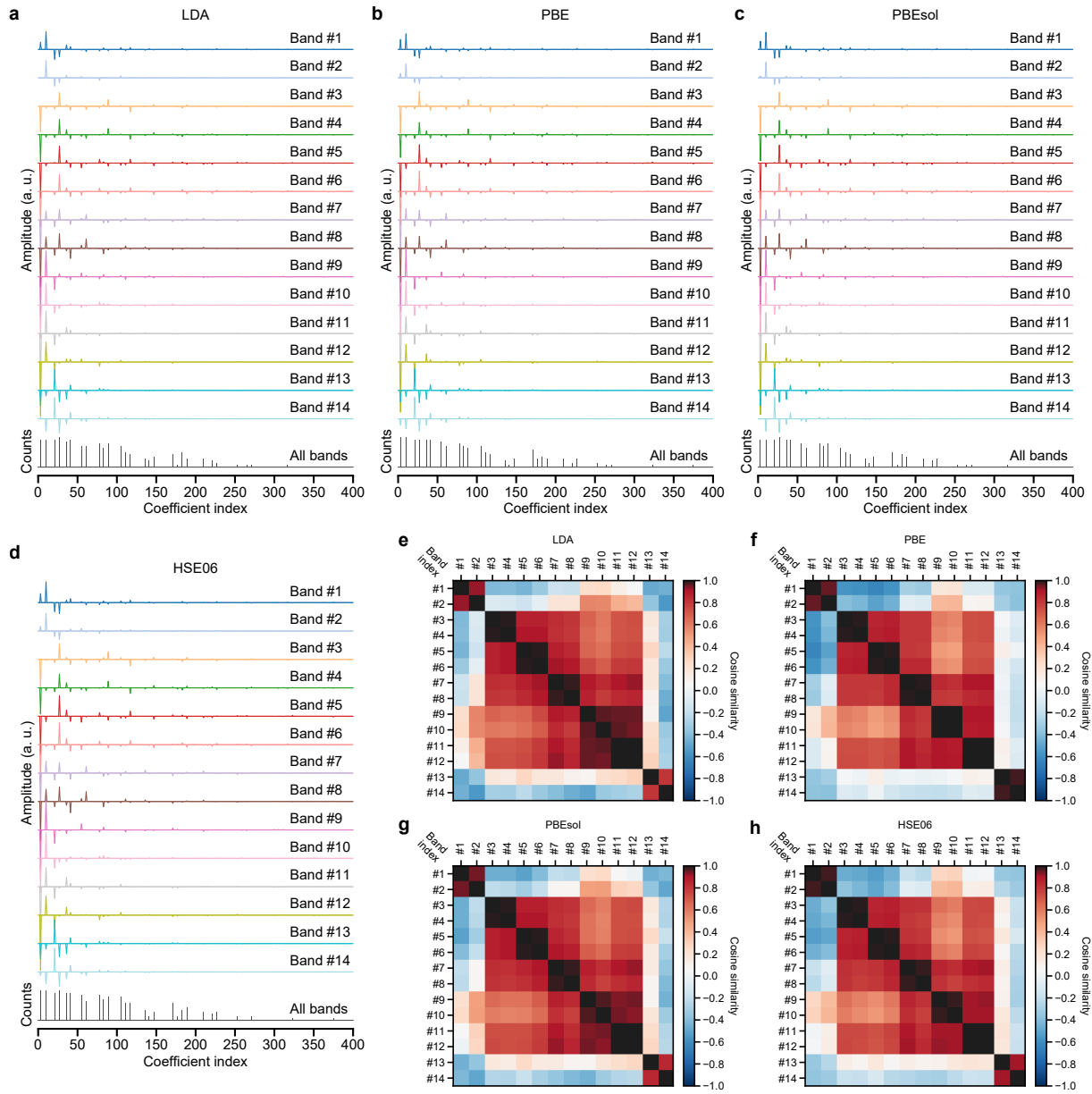
Supplementary Figure 9: **Geometric featurization of the energy bands of WSe$_2$. a-d**, Decomposition of the 14 valence energy bands of WSe$_2$ into hexagonal Zernike polynomials for the DFT band structure calculations carried out at the levels of LDA (**a**), PBE (**b**), PBEsol (**c**), and HSE06 (**d**), respectively. Similar characteristics are seen compared with the reconstructed band structure shown in Fig. 3a in the main text, including the sparse distribution of significant basis terms and the decreasing dependence on higher-order basis terms. **e-h**, Cosine similarity matrices between the 14 energy bands of WSe$_2$ for the DFT band structure calculations carried out at the levels of LDA (**e**), PBE (**f**), PBEsol (**g**), and HSE06 (**h**), respectively. The characteristics of these matrices resemble that calculated for the reconstructed band structure as shown in Fig. 3c in the main text.
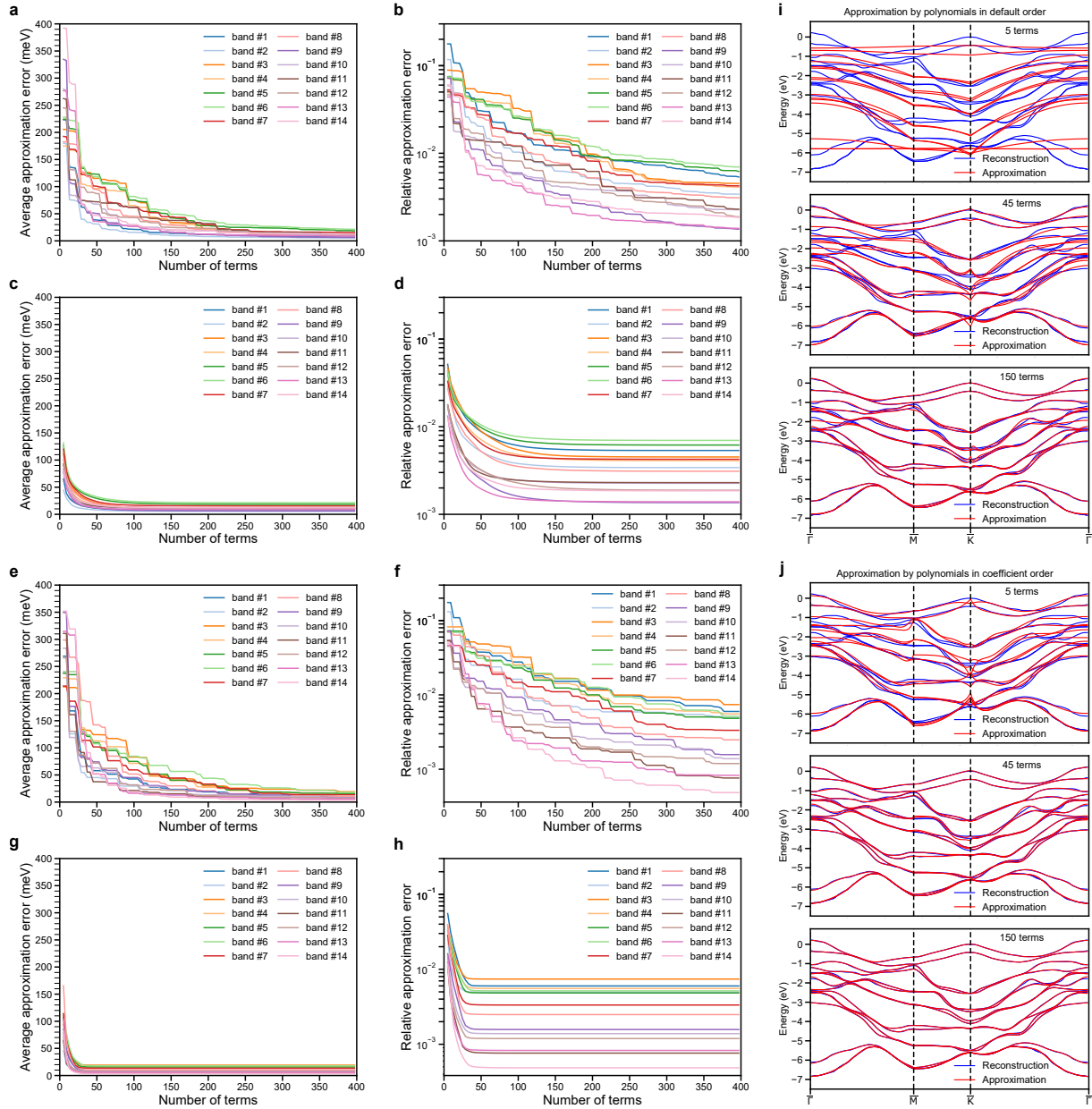
Supplementary Figure 10: **Approximation to the band structure of WSe$_2$ by a polynomial basis**. **a-j**, Demonstration of the convergence properties of the polynomial approximation using reconstructed photoemission band structure (**a-d**) and DFT band structure calculated at the LDA level (**e-h**). When summing the hexagonal Zernike polynomial in the default order, the average and relative approximation errors for the reconstructed (**a,b**) and theoretical (**e,f**) energy bands converge much slower than summing the polynomials in an ordering ranked by the magnitude of their coefficients (coefficient order). This observation is similar for reconstructed (**c,d**) and theoretical (**g,h**) energy bands. **i-j**, Visualization of the difference in convergence rates using the reconstructed band structure along the high-symmetry lines. The naturally-ordered polynomial basis has not yet converged with 150 terms (**i**), while the coefficient-ranked polynomials (**j**) produces an accurate approximation well within that limit.

theoretical band structure at the level of LDA-DFT (Supplementary Fig. 10e-h) show comparable convergence rates. Quantitatively, the approximation using hexagonal ZPs ordered by the magnitude of the corresponding coefficients (i.e. coefficient order) converges to within 10-30 meV/band within 50 polynomial basis terms, significantly faster than using the default order (see also Fig. 3b for reference). The remaining errors are on par with the finite step size along
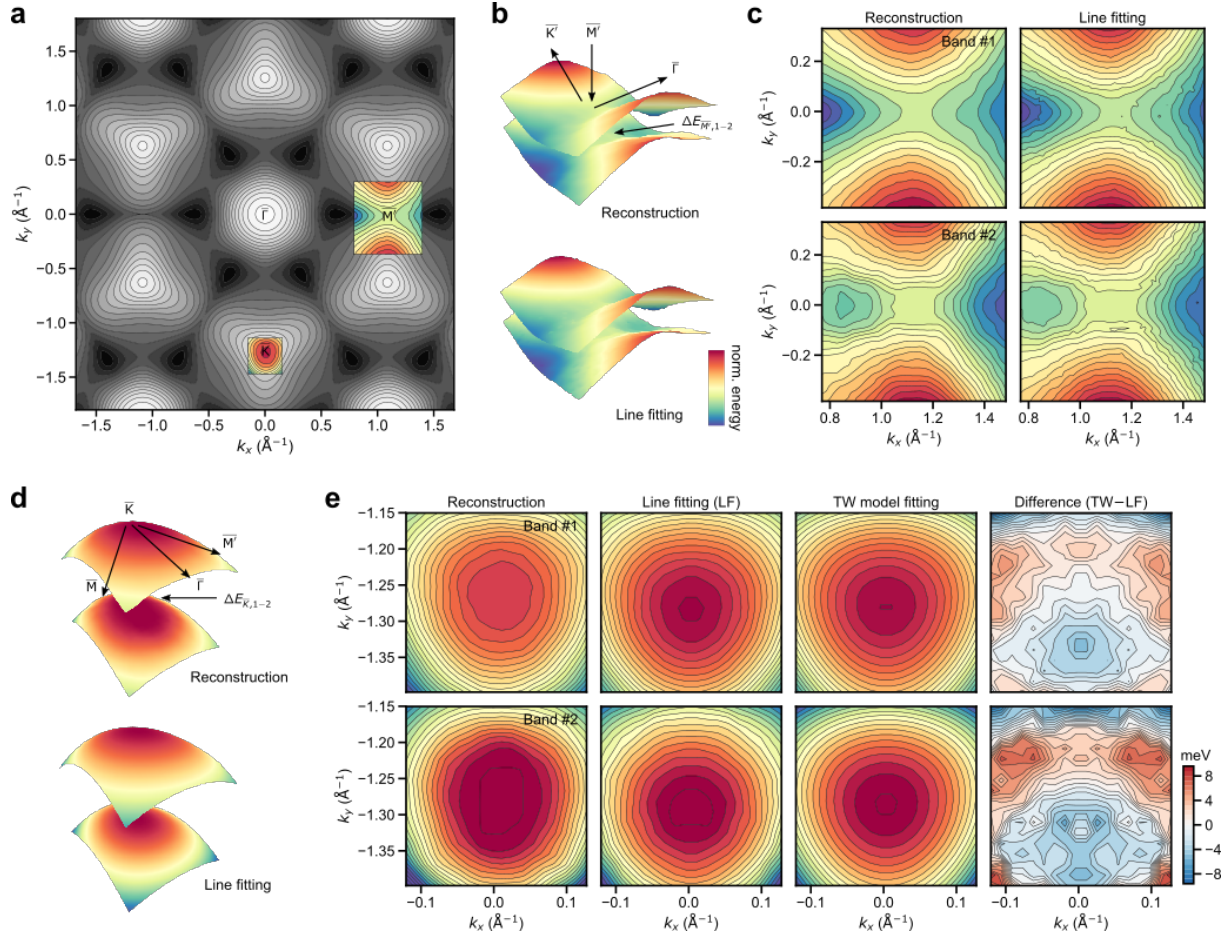
Supplementary Figure 11: **Local band structure parameters**. **a**, The first valence band of WSe$_2$ with constant-energy contours. The patches around high-symmetry points $\overline{K}$ and $\overline{M}'$ from reconstruction (with LDA-DFT as the initialization) are overlaid in color. **b,c**, Patch around the $\overline{M}'$-point, a saddle point in the dispersion surface, visualized in 3D (**b**) and 2D (**c**), respectively. The energy gap at $\overline{M}'$ due to spin-orbit coupling (SOC) results in the energy difference $\Delta E_{\overline{M}',1-2}$. **d,e**, Patch around the $\overline{K}$-point, the energy maximum of the valence band, visualized in 3D (**d**) and 2D (**e**), respectively. The SOC results in the energy gap $\Delta E_{\overline{K},1-2}$. The outcome of fitting to a trigonal warping (TW) model around $\overline{K}$ from **k·p** theory [22] is shown in **e**.

the energy axis in the data ($\sim$ 18 meV) that results in the imperfect smoothness of the recon-

structed bands. This further proves that the hexagonal ZPs can provide an accurate and sparse approximation for the band structure data. The trend of convergence between these two types of polynomial ordering is further illustrated in Supplementary Fig. 10i-j in the momentum path along high-symmetry lines of the reconstructed band structure.

Local scale structure. Local structural information includes energy gaps, effective masses, warpings, (avoided) crossings, etc. We extracted some of their associated parameters at and around three high-symmetry points ($\overline{K}$, $\overline{M'}$, and $\overline{\Gamma}$, see Supplementary Fig. 11a) and compiled the results in Supplementary Table 2. The dispersions and band structure parameters from the MAP reconstruction are compared with those extracted by line-by-line fitting of the EDCs, which used the band energies from the reconstruction as initialization to improve robustness. Around $\overline{K}$, two spectral peaks corresponding to two spin-split bands were fit simultaneously, while around $\overline{M'}$ and $\overline{\Gamma}$, four were fit simultaneously due to the spectral proximity of the first four valence bands (see Supplementary Fig. 4). The fitting is carried out using a linear superposition of Voigt lineshapes and the lmfit package [70] with the reconstructed band energy as initialization (but not fixed). The fitting procedure iterates over the EDCs (e.g. a total of 50×50 EDCs for the patch around $\overline{M'}$). Unstable fits yielding erratic results (e.g. if differing significantly from neighboring values) are re-fit with either algorithmically or manually adjusted initialization. Supplementary Table 2 shows that the local structural information from reconstruction is generally consistent with those obtained by iterative pointwise fitting, while differing from DFT calculations. The deviations in the size of energy gaps at $\overline{K}$ and $\overline{M'}$ between reconstruction and pointwise fitting lie in the same range as the momentum-averaged reconstruction errors (see section S2), which are due to the finite coordinate spacing in the data ($\sim$ 18 meV in energy).

The region extracted around $\overline{K}$ (see Supplementary Fig. 11d-e) contains about 10% of the distance of $\overline{\Gamma} - \overline{K}$. Due to the strong trigonal warping (TW) effect in this class of materials, the effective masses and the TW parameters around $\overline{K}$ were fit simultaneously in 2D using the momentum-space model derived from $\mathbf{k}\cdot\mathbf{p}$ theory [22].

$$E(\mathbf{q}) = \frac{\hbar^2\mathbf{q}^2}{2m_{\overline{K}}} + C|\mathbf{q}|^3 \cos(3\varphi_{\mathbf{q}} + \theta) + E_0. \tag{22}$$

Here, $\mathbf{q}$ is the momentum vector $\mathbf{k}$ recentered on a particular $\overline{K}$ (or $\overline{K'}$) point by translation, $m_{\overline{K}}$ is the effective mass of the hole at $\overline{K}$ point, C is the magnitude of the TW (named $C_{3\mathrm{w}}$

41

Supplementary Table 2: **Band structure parameters from experiment and theory.**
Effective masses of holes ($m_{\overline{K}}$), trigonal warping parameters (C) are extract at $\overline{K}$ point in the first two valence bands. Two directional effective masses at $\overline{M}'$ ($m_{\overline{M}'}$), and one at $\overline{\Gamma}$ ($m_{\overline{\Gamma}}$), are obtained for the first valence band. The energy gaps (ΔE) between the first two valence bands are obtained at both $\overline{K}$ and $\overline{M}'$ points. The number (1 or 2) in the subscript of the parameter symbols denotes the valence band index, m_e is the mass of an isolated electron.

Symmetry point	Parameter	LDA recon. [1]	Line fitting [2]	LDA [3]	HSE06 [3]
$\overline{K}$	$m_{\overline{K},1}/m_e$	-0.62	-0.60	-0.49	-0.42
$\overline{K}$	$m_{\overline{K},2}/m_e$	-0.74	-0.78	-0.64	-0.54
$\overline{K}$	$C_{\overline{K},1}$ (eV·Å^3)	5.3	5.8	6.2	4.5
$\overline{K}$	$C_{\overline{K},2}$ (eV·Å^3)	4.0	3.9	3.9	3.2
$\overline{K}$	$\Delta E_{\overline{K},1-2}$ (meV)	419	446	485	467
$\overline{M}'$	$m_{\overline{M}'-\overline{\Gamma},1}/m_e$	0.71	0.72	0.25	0.17
$\overline{M}'$	$m_{\overline{M}'-\overline{K}',1}/m_e$	-1.6	-1.5	-1.1	-0.90
$\overline{M}'$	$\Delta E_{\overline{M}',1-2}$ (meV)	352	338	127	48
$\overline{\Gamma}$	$m_{\overline{\Gamma},1}/m_e$	-0.82	-1.1	-0.81	-1.0

[1] Using band dispersion reconstructed globally by the proposed probabilistic machine learning algorithm with DFT calculation at the LDA level as the initialization.

[2] Using band dispersion from iterative lineshape fitting of the energy distribution curves (in an region around the corresponding high-symmetry points).

[3] With fully optimized structure, see Supplementary Table 1.

in [22]), $\varphi_{\mathbf{q}}$ is the polar angle in the coordinate system centered on a $\overline{K}$ (or $\overline{K}'$) point, θ is an auxiliary fitting parameter used to accommodate the orientation of the TW with respect to the pixel coordinates defined by the rectangular region of interest, E_0 accounts for the energy offset. The energy gaps at $\overline{K}$ ($\Delta E_{\overline{K},1-2}$) and $\overline{M}'$ ($\Delta E_{\overline{M}',1-2}$) are illustrated in Supplementary Fig. 11 (b and d), respectively. The $\overline{M}'$ (or $\overline{M}$) point situates at a saddle point of the dispersion surface (first valence band), as shown in Supplementary Fig. 11b-c. Its lower symmetry (compared with $\overline{K}$, $\overline{K}'$ and $\overline{\Gamma}$) means that the effective masses exhibits anisotropy, with opposite signs and magnitude along the $\overline{M}' - \overline{\Gamma}$ and $\overline{M}' - \overline{K}'$ directions. We fit the dispersion locally using a model that also accounts for the spin-orbit interaction involving a linear momentum-dependent shift (Eq. 14 in [22]). The second valence band is not fitted at $\overline{M}'$ due to the pronounced dispersion modulation by interband coupling unaccounted for in the existing saddle-shaped model. At around $\overline{\Gamma}$, a single effective mass is extracted by fitting a paraboloid to a local patch of the dispersion surface.

References

[1] L. P. Bouckaert, R. Smoluchowski, and E. Wigner. Theory of Brillouin Zones and Symmetry Properties of Wave Functions in Crystals. *Physical Review*, 50(1):58–67, 7 1936.

[2] T.-C. Chiang and F. Seitz. Photoemission spectroscopy in solids. *Annalen der Physik*, 10(1-2):61–74, 2 2001.

[3] T. Valla, A. V. Fedorov, P. D. Johnson, B. O. Wells, S. L. Hulbert, Q. Li, G. D. Gu, and N. Koshizuka. Evidence for Quantum Critical Behavior in the Optimally Doped Cuprate Bi2Sr2CaCu2O8+. *Science*, 285(5436):2110–2113, 9 1999.

[4] G. Levy, W. Nettke, B. M. Ludbrook, C. N. Veenstra, and A. Damascelli. Deconstruction of resolution effects in angle-resolved photoemission. *Physical Review B*, 90(4):045150, 7 2014.

[5] Gerd Schönhense, Katerina Medjanik, and Hans-Joachim Elmers. Space-, time- and spin-resolved photoemission. *Journal of Electron Spectroscopy and Related Phenomena*, 200:94–118, 4 2015.

[6] K. Medjanik, O. Fedchenko, S. Chernov, D. Kutnyakhov, M. Ellguth, A. Oelsner, B. Schönhense, T. R. F. Peixoto, P. Lutz, C.-H. Min, F. Reinert, S. Däster, Y. Acremann, J. Viefhaus, W. Wurth, H. J. Elmers, and G. Schönhense. Direct 3D mapping of the Fermi surface and Fermi velocity. *Nature Materials*, 16(6):615–621, 3 2017.

[7] Eric B. Isaacs and Chris Wolverton. Inverse Band Structure Design via Materials Database Screening: Application to Square Planar Thermoelectrics. *Chemistry of Materials*, 30(5):1540–1546, 3 2018.

[8] E. G. Marin, M. Perucchini, D. Marian, G. Iannaccone, and G. Fiori. Modeling of Electron Devices Based on 2-D Materials. *IEEE Transactions on Electron Devices*, 65(10):4167–4179, 10 2018.

[9] Andrea Damascelli, Zahid Hussain, and Zhi-Xun Shen. Angle-resolved photoemission studies of the cuprate superconductors. *Reviews of Modern Physics*, 75(2):473–541, 4 2003.

[10] Christopher Corder, Peng Zhao, Jin Bakalis, Xinlong Li, Matthew D. Kershis, Amanda R. Muraca, Michael G. White, and Thomas K. Allison. Ultrafast extreme ultraviolet photoemission without space charge. *Structural Dynamics*, 5(5):054301, 9 2018.

[11] M. Puppin, Y. Deng, C. W. Nicholson, J. Feldl, N. B. M. Schröter, H. Vita, P. S. Kirchmann, C. Monney, L. Rettig, M. Wolf, and R. Ernstorfer. Time- and angle-resolved photoemission spectroscopy of solids in the extreme ultraviolet at 500 kHz repetition rate. *Review of Scientific Instruments*, 90(2):023104, 2 2019.

[12] J. M. Riley, F. Mazzola, M. Dendzik, M. Michiardi, T. Takayama, L. Bawden, C. Granerød, M. Leandersson, T. Balasubramanian, M. Hoesch, T. K. Kim, H. Takagi, W. Meevasana, Ph. Hofmann, M. S. Bahramy, J. W. Wells, and P. D. C. King. Direct observation of spin-polarized bulk bands in an inversion-symmetric semiconductor. *Nature Physics*, 10(11):835–839, 11 2014.

[13] M. S. Bahramy, O. J. Clark, B.-J. Yang, J. Feng, L. Bawden, J. M. Riley, I. Marković, F. Mazzola, V. Sunko, D. Biswas, S. P. Cooil, M. Jorge, J. W. Wells, M. Leandersson, T. Balasubramanian, J. Fujii, I. Vobornik, J. E. Rault, T. K. Kim, M. Hoesch, K. Okawa, M. Asakawa, T. Sasagawa, T. Eknapakul, W. Meevasana, and P. D. C. King. Ubiquitous formation of bulk Dirac cones and topological surface states from a single orbital manifold in transition-metal dichalcogenides. *Nature Materials*, 17(1):21–28, 1 2018.

[14] Niels B. M. Schröter, Ding Pei, Maia G. Vergniory, Yan Sun, Kaustuv Manna, Fernando de Juan, Jonas. A. Krieger, Vicky Süss, Marcus Schmidt, Pavel Dudin, Barry Bradlyn, Timur K. Kim, Thorsten Schmitt, Cephise Cacho, Claudia Felser, Vladimir N. Strocov, and Yulin Chen. Chiral topological semimetal with multifold band crossings and long Fermi arcs. *Nature Physics*, 5 2019.

[15] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[16] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 5 2015.

[17] Chaohui Wang, Nikos Komodakis, and Nikos Paragios. Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, 117(11):1610 – 1627, 2013.

[18] Kevin Kaufmann, Chaoyi Zhu, Alexander S. Rosengarten, Daniel Maryanovsky, Tyler J. Harrington, Eduardo Marin, and Kenneth S. Vecchio. Crystal symmetry determination in electron diffraction using machine learning. *Science*, 367(6477):564–568, jan 2020.

[19] Vincent Stimper, Stefan Bauer, Ralph Ernstorfer, Bernhard Scholkopf, and Rui Patrick Xian. Multidimensional Contrast Limited Adaptive Histogram Equalization. *IEEE Access*, 7:165437–165447, 6 2019.

[20] M. Traving, M. Boehme, L. Kipp, M. Skibowski, F. Starrost, E. E. Krasovskii, A. Perlov, and W. Schattke. Electronic structure of WSe_2: A combined photoemission and inverse photoemission study. *Physical Review B*, 55(16):10392–10399, 4 1997.

[21] Th. Finteis, M. Hengsberger, Th. Straub, K. Fauth, R. Claessen, P. Auer, P. Steiner, S. Hüfner, P. Blaha, M. Vögt, M. Lux-Steiner, and E. Bucher. Occupied and unoccupied electronic band structure of WSe_2. *Physical Review B*, 55(16):10400–10411, 4 1997.

[22] Andor Kormányos, Guido Burkard, Martin Gmitra, Jaroslav Fabian, Viktor Zólyomi, Neil D Drummond, and Vladimir Fal'ko. $k \cdot p$ theory for two-dimensional transition metal dichalcogenide semiconductors. *2D Materials*, 2(2):022001, 4 2015.

[23] Simon Moser. An experimentalist's guide to the matrix element in angle resolved photoemission. *Journal of Electron Spectroscopy and Related Phenomena*, 214:29–52, 1 2017.

[24] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pages 1–20. AIP, 2001.

[25] Dengsheng Zhang and Guojun Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 1 2004.

[26] A. Khotanzad and Y.H. Hong. Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):489–497, 5 1990.

[27] Virendra N. Mahajan and Guang-ming Dai. Orthonormal polynomials in wavefront analysis: analytical solution. *Journal of the Optical Society of America A*, 24(9):2994, 2007.

[28] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 4 2008.

[29] Dorothea Golze, Marc Dvorak, and Patrick Rinke. The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy. *Frontiers in Chemistry*, 7:377, 7 2019.

[30] M Zacharias, M Scheffler, and C Carbogno. Fully anharmonic, non-perturbative theory of vibronically renormalized electronic band structures. *arXiv*, 2003.10417, 2020.

[31] Matthias Ehrhardt and Thomas Koprucki, editors. *Multi-Band Effective Mass Approximations*, volume 94 of *Lecture Notes in Computational Science and Engineering*. Springer, 2014.

[32] R.A. Ewings, A. Buts, M.D. Le, J. van Duijn, I. Bustinduy, and T.G. Perring. Horace : Software for the analysis of data from single crystal spectroscopy experiments at time-of-flight neutron instruments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 834:132–142, 10 2016.

[33] Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, Michal Walczak, Jochen Garcke, Christian Bauckhage, and Jannis Schuecker. Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems. *arXiv*, 1903.12394, 3 2019.

[34] A. Oelsner, O. Schmidt, M. Schicketanz, M. Klais, G. Schönhense, V. Mergel, O. Jagutzki, and H. Schmidt-Böcking. Microspectroscopy and imaging using a delay line detector in time-of-flight photoemission microscopy. *Review of Scientific Instruments*, 72(10):3968–3974, 10 2001.

[35] Rui Patrick Xian, Yves Acremann, Steinn Ymir Agustsson, Maciej Dendzik, Kevin Bühlmann, Davide Curcio, Dmytro Kutnyakhov, Frederico Pressacco, Michael Heber,

Shuo Dong, Jure Demsar, Wilfried Wurth, Philip Hofmann, Martin Wolf, Laurenz Rettig, and Ralph Ernstorfer. An open-source, distributed workflow for band mapping data in multidimensional photoemission spectroscopy. *arXiv*, 1909.07714, 9 2019.

[36] SPECS GmbH. METIS 1000 Brochure. https://www.specs-group.com/fileadmin/user_upload/products/brochures/SPECS_Brochure-METIS_RZ_web.pdf, 2019.

[37] Rui Patrick Xian, Laurenz Rettig, and Ralph Ernstorfer. Symmetry-guided nonrigid registration: The case for distortion correction in multidimensional photoemission spectroscopy. *Ultramicroscopy*, 202:133–139, 7 2019.

[38] J Kittler and J Fglein. Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2(1):13 – 29, 1984.

[39] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[40] Stan Li. *Markov Random Field Modeling in Image Analysis*. Advances in Pattern Recognition. Springer, 3 edition, 2009.

[41] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer New York, 2nd edition, 2006.

[42] Vincent Stimper and Rui Patrick Xian. fuller. https://github.com/mpes-kit/fuller.

[43] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, 8 1980.

[44] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45:13244–13249, 6 1992.

[45] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, 10 1996.

[46] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Physical Review Letters*, 100(13):136406, 4 2008.

[47] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 5 2003.

[48] Volker Blum, Ralf Gehrke, Felix Hanke, Paula Havu, Ville Havu, Xinguo Ren, Karsten Reuter, and Matthias Scheffler. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications*, 180(11):2175–2196, 11 2009.

[49] William P. Huhn and Volker Blum. One-hundred-three compound band-structure benchmark of post-self-consistent spin-orbit coupling treatments in density functional theory. *Physical Review Materials*, 1(3):033803, 8 2017.

[50] James C. Wyant and Katherine Creath. Basic wavefront aberration theory. In *Applied Optics and Optical Engineering*, volume Xl, pages 1–53. Academic Press, Inc., 1992.

[51] David S. Watkins. *Fundamentals of matrix computations*. Wiley, 3rd edition, 2010.

[52] Stefan Hüfner. *Photoelectron Spectroscopy*. Advanced Texts in Physics. Springer Berlin Heidelberg, Berlin, Heidelberg, 3rd editio edition, 2003.

[53] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[54] J. M. Hammersley and P. Clifford. Markov Fields on Finite Graphs and Lattices. Unpublished, 1971.

[55] Julian Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

[56] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 11 1984.

[57] Julian Besag. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3):259–279, 7 1986.

[58] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[59] J. Besag. On the statistical analysis of nearest-neighbours. In *Proceedings of the European Meeting of Statisticians, Budapest*, 1972.

[60] M. Locatelli. A note on the Griewank test function. *Journal of Global Optimization*, 25(2):169–174, 2 2003.

[61] Cristina Bena and Gilles Montambaux. Remarks on the tight-binding model of graphene. *New Journal of Physics*, 11(9):095003, 9 2009.

[62] Virendra N. Mahajan and Guang-ming Dai. Orthonormal polynomials for hexagonal pupils. *Optics Letters*, 31(16):2462, 8 2006.

[63] H. C. van de Hulst and J. J. M. Reesinck. Line Breadths and Voigt Profiles. *The Astrophysical Journal*, 106:121, 7 1947.

[64] Mofreh R. Zaghloul and Ahmed N. Ali. Algorithm 916: Computing the Faddeyeva and Voigt Functions. *ACM Transactions on Mathematical Software*, 38(2):1–22, 12 2011.

[65] Hannes Huebener. Calculations of the momentum-resolved electron self-energy of 2H-WSe$_2$. Unpublished, 2016.

[66] Franz Knuth, Christian Carbogno, Viktor Atalla, Volker Blum, and Matthias Scheffler. All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals. *Computer Physics Communications*, 190:33 – 50, 2015.

[67] D. Voß, P. Krüger, A. Mazur, and J. Pollmann. Atomic and electronic structure of WSe_2 from *ab initio* theory: Bulk crystal and thin film systems. *Phys. Rev. B*, 60:14311–14317, 11 1999.

[68] K K Kam, C L Chang, and D W Lynch. Fundamental absorption edges and indirect band gaps in $W_{1-x}Mo_xSe_2$ ($0 \leqslant x \leqslant 1$). *Journal of Physics C: Solid State Physics*, 17(22):4031–4040, 8 1984.

[69] Sujay B. Desai, Gyungseon Seol, Jeong Seuk Kang, Hui Fang, Corsin Battaglia, Rehan Kapadia, Joel W. Ager, Jing Guo, and Ali Javey. Strain-induced indirect to direct bandgap transition in multilayer WSe_2. *Nano Letters*, 14(8):4592–4597, 2014. PMID: 24988370.

[70] Matt Newville, Renee Otten, Andrew Nelson, Antonino Ingargiola, Till Stensitzki, Dan Allan, Austin Fox, Faustin Carter, Michał, Dima Pustakhod, Yoav Ram, Glenn, Christoph Deil, Stuermer, Alexandre Beelen, Oliver Frost, Nicholas Zobrist, Gustavo Pasquevich, Allan L. R. Hansen, Tim Spillane, Shane Caldwell, Anthony Polloreno, Andrewhannum, Julius Zimmermann, Jose Borreguero, Jonathan Fraine, Deep-42-thought, Benjamin F. Maier, Ben Gamari, and Anthony Almarza. lmfit/lmfit-py 1.0.0. https://doi.org/10.5281/zenodo.3588521, dec 2019.