

Supplementary Information

A machine learning route between band mapping and band structure

Contents

S1 Band structure reconstruction	28
S1.1 Physical foundations	28
S1.2 Markov random field modeling	29
S1.3 Optimization procedure	31
S1.4 Hyperparameter tuning	35
S1.5 Reconstructions using different theories as initializations	40
S2 Generation of and validation on synthetic data	40
S2.1 Generation of band structure data	40
S2.2 Initialization tuning	41
S2.3 Approximate generation of photoemission data	41
S2.4 Validation of the reconstruction algorithm	43
S2.5 Computational benchmarks	49
S3 Reconstruction for other experimental datasets	52
S3.1 Near-gap electronic bands of a topological insulator ($\text{Bi}_2\text{Te}_2\text{Se}$)	52
S3.2 Bulk electronic bands of gold (Au)	54
S4 Band structure calculations	54
S4.1 DFT calculations	54
S4.2 Brillouin zone tiling	57
S5 Band structure informatics	60
S5.1 Global scale structure descriptors	60
S5.2 Local scale structure descriptors	62

S1 Band structure reconstruction

S1.1 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, $\Sigma(\mathbf{k}, E) = \text{Re}\Sigma(\mathbf{k}, E) + i\text{Im}\Sigma(\mathbf{k}, E)$, 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 [5, 70], $A(\mathbf{k}, E)$. For a single energy band of a many-body electronic system,

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

Within this framework, the band loci of the photoemission (or quasiparticle) band structure (BS), $b(\mathbf{k}, E) = \epsilon_{\mathbf{k}} + \text{Re}\Sigma(\mathbf{k}, E)$, 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 [19], the Fermi-Dirac occupation function, $f_{\text{FD}}(E)$, 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,

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

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.

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

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

The condition $f_{\text{FD}}(E) \rightarrow 1$ applies to valence bands, while $|M_{f_j, i_j}(\mathbf{k}, E)| \rightarrow 1$ 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.

S1.2 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.

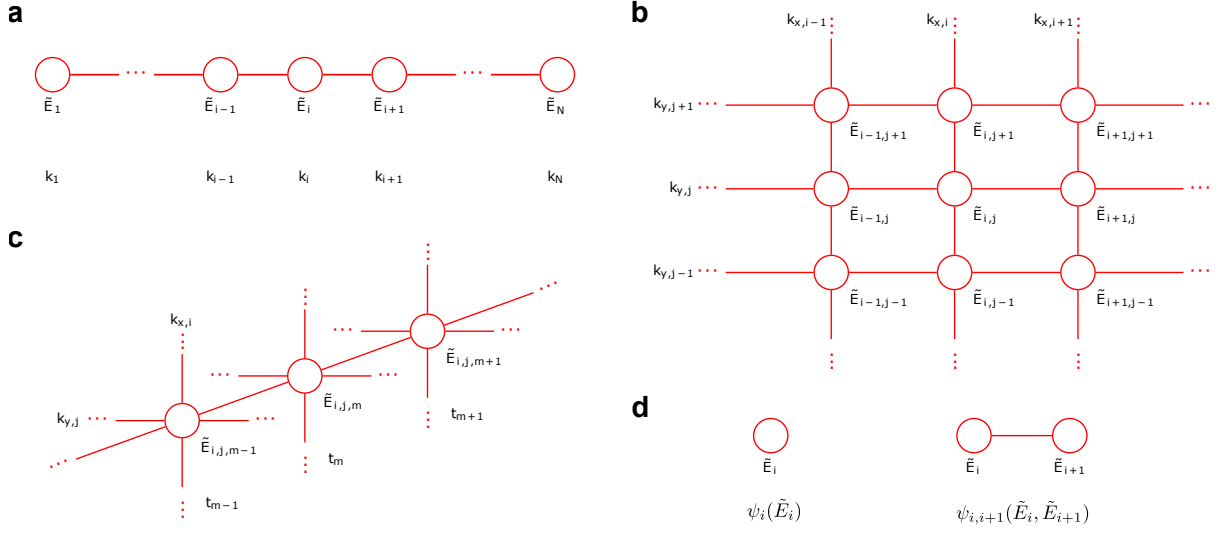
Deriving the MRF amounts to determining the joint distribution of the random variables associated with its graphical representation. In graphical model theory [71], a graph is constructed from the fundamental components called cliques. Each clique C of a graph is a subset of nodes that shares an edge 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 [71], 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 $\psi_C(\mathbf{X}_C) \geq 0$. According to the Hammersley-Clifford theorem [71–73], the joint distribution of a vector of random variables, \mathbf{X} , can be written in the factorized form,

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

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 neu-

tralize 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

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

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

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

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

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

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

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

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

S1.3 Optimization procedure

Optimization of the MRF model is a local minima-finding process [71]. 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) [57] 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.

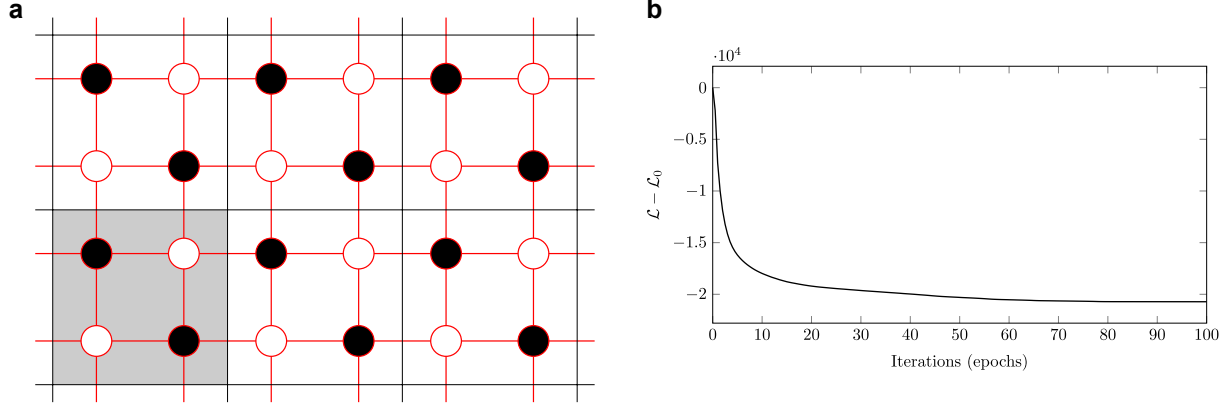
1. Iterated conditional mode: Originally developed for similar optimization problems arising in image denoising [71, 74, 75], ICM is applicable to optimizing MRF at any dimension. The ICM procedure includes (i) initialization of the random variables (e.g. $\{\tilde{E}_{i,j}\}$ in 2D MRF) and (ii) selection 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) [59] 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 [71, 76]. Analogously, blocking the black nodes d-separates the white nodes. Since the MRF models satisfy the Hammersley-Clifford theorem [72], 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 related to the coding method have been elaborated in [73, 77].

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:      $\log\_I_u[i, j, :, :] = \log(I[\text{ind\_x} + i, \text{ind\_y} + j, :])$ 
   # Iterative optimization of energy values
7: for n in range(N) do
   # Update white nodes
8:    $E_u[0, 0, :, :] = \text{update\_E}(0, 0, \log\_I_u, E_u, E)$ 
9:    $E_u[1, 1, :, :] = \text{update\_E}(1, 1, \log\_I_u, E_u, E)$ 
   # Update black nodes
10:   $E_u[0, 1, :, :] = \text{update\_E}(0, 1, \log\_I_u, E_u, E)$ 
11:   $E_u[1, 0, :, :] = \text{update\_E}(1, 0, \log\_I_u, E_u, E)$ 
   # Assemble reconstruction from all nodes in the units
12: for i in [0, 1] do
13:   for j in [0, 1] do
14:      $E_{\text{rec}}[\text{ind\_x} + i, \text{ind\_y} + j] = E_u[i, j, :, :]$ 

   # 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:    $\log\_p = \log\_I_u[i, j, :, :]$ 
   # Subtract by energy differences from nearest neighbor nodes within unit
18:    $\log\_p -= \text{squ\_diff}[(i + 1) \% 2, j, :, :]$ 
19:    $\log\_p -= \text{squ\_diff}[i, (j + 1) \% 2, :, :]$ 
   # Subtract by energy differences from nearest neighbor nodes of the neighbor-
   # ing unit
20:    $\log\_p -= \text{shift}(\text{squ\_diff}[(i + 1) \% 2, j, :, :], 2 * i - 1, \text{axis}=2)$ 
21:    $\log\_p -= \text{shift}(\text{squ\_diff}[i, (j + 1) \% 2, :, :], 2 * j - 1, \text{axis}=3)$ 
   # Return optimal energy values
22:   return  $E[\text{argmax}(\log\_p)]$ 
```



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.

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 of the original and the shifted grid. The procedure is implemented in the open-source **fuller** package [60] using Tensorflow [58]. 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 ~ 90 iterations.

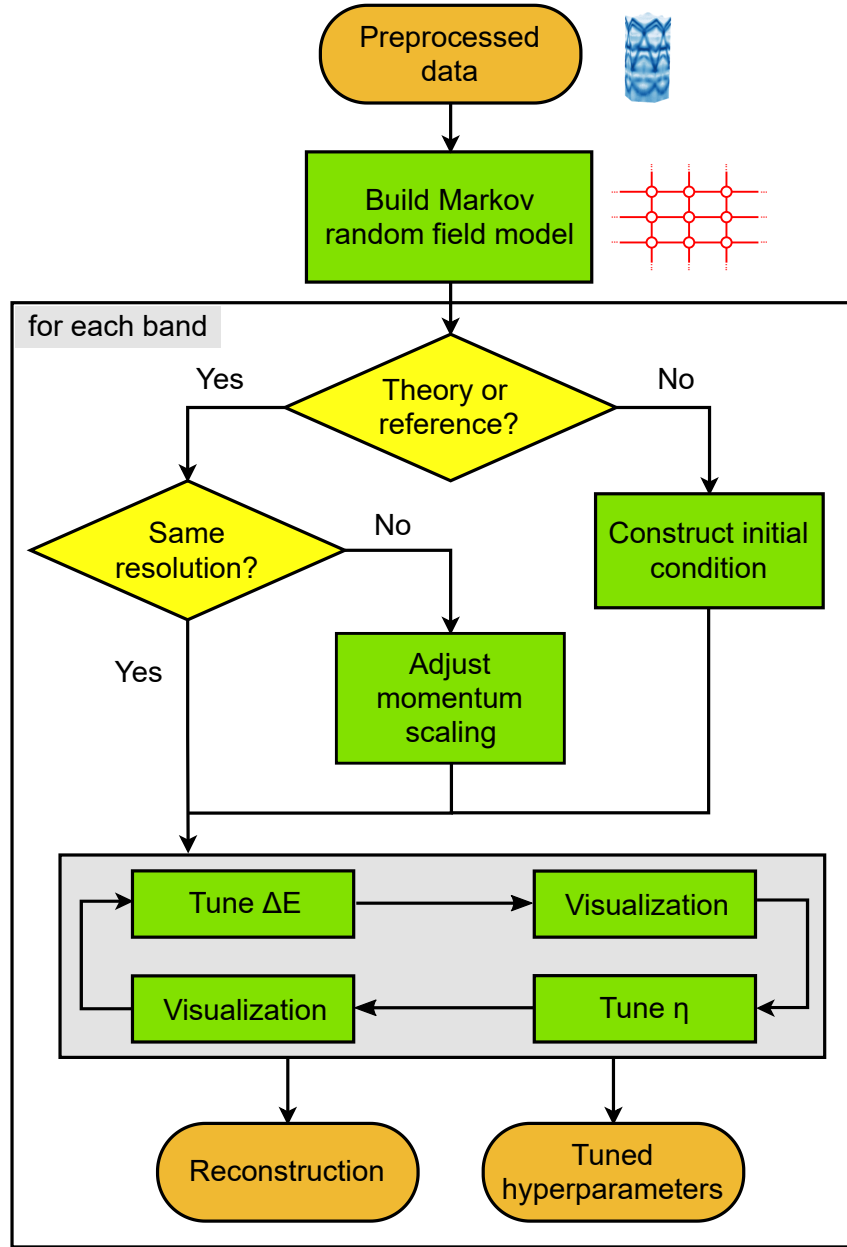
4. Robust initialization: Since the current MRF model doesn't include any explicit regularization on the outcome with respect to the initialization, the optimizer is free to explore a large range of values. In other words, the initial band dispersion is able to freely deform to fit to the band loci embedded 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. 6-9 (see Section S2).

S1.4 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. A flowchart presented in Supplementary Fig. 3 illustrates the general steps in obtaining a desirable reconstruction including where the tuning of each hyperparameter fits in.

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 3) 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



Supplementary Figure 3: **Flowchart for reconstruction tuning.** Illustration of the steps for tuning the reconstruction starting from preprocessed data (outcome from the procedure illustrated in the main text Fig. 1c-f). Tuning of the three hyperparameters – the momentum scaling, energy shift (ΔE) and nearest-neighbor Gaussian width (η), are placed in sequence within the workflow. The workflow outputs reconstruction of a single energy band with tuned hyperparameters at the end. For reconstructing the dispersion of multiple energy bands, the workflow is repeated over each band.

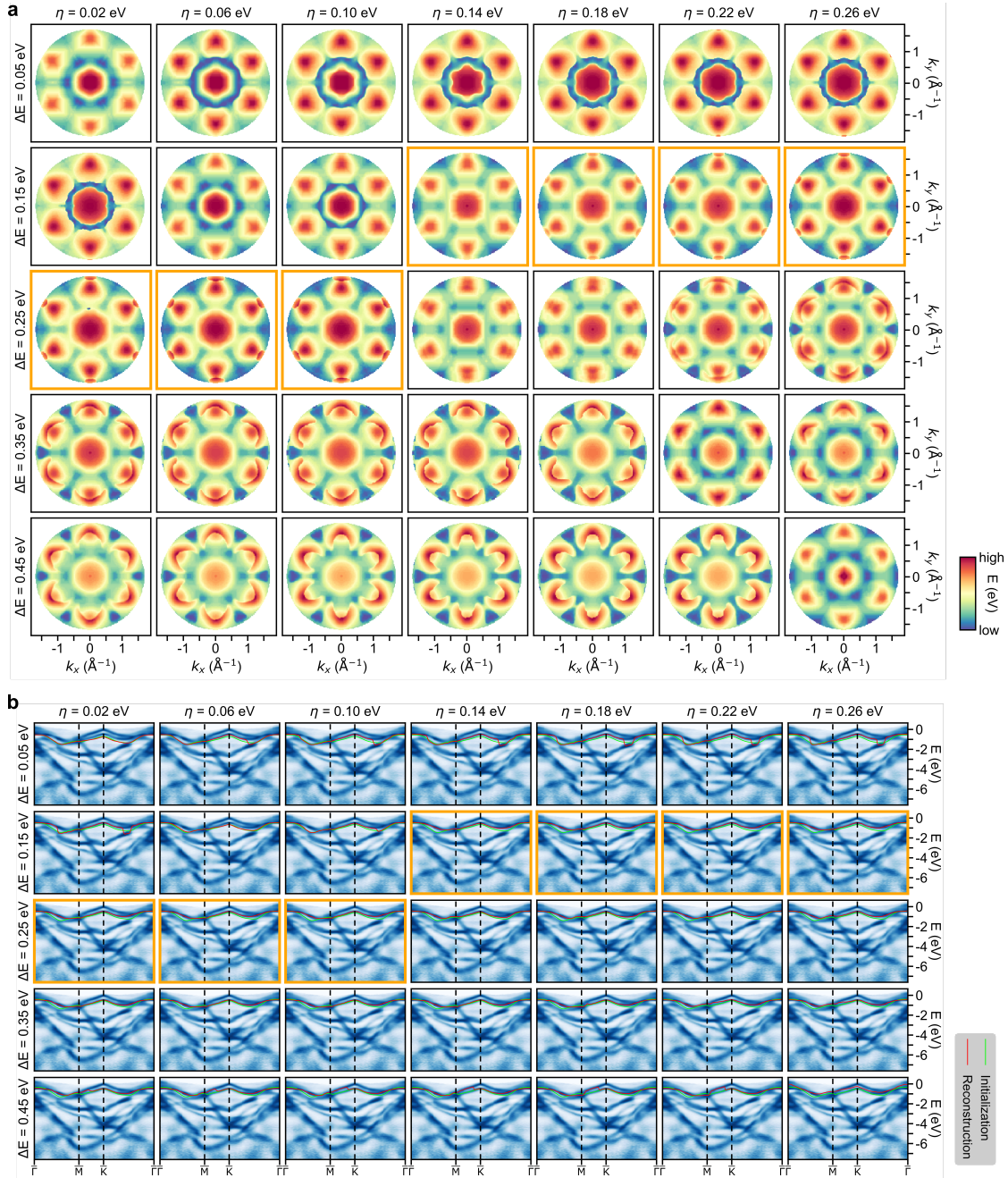
inspection of the theoretical band energies overplotted on photoemission data (usually in the energy-momentum slices). In practice, the necessary energy shifts vary between bands and also depend on level of approximation in the BS calculation used as initialization, as illustrated in Fig. 2a of the main 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 WSe₂, the momentum grid of the experimental data has a spacing of $\Delta k_x = \Delta k_y \approx 0.015 \text{ \AA}^{-1}$, we used $\eta \in [0.05, 0.2] \text{ eV}$. 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,

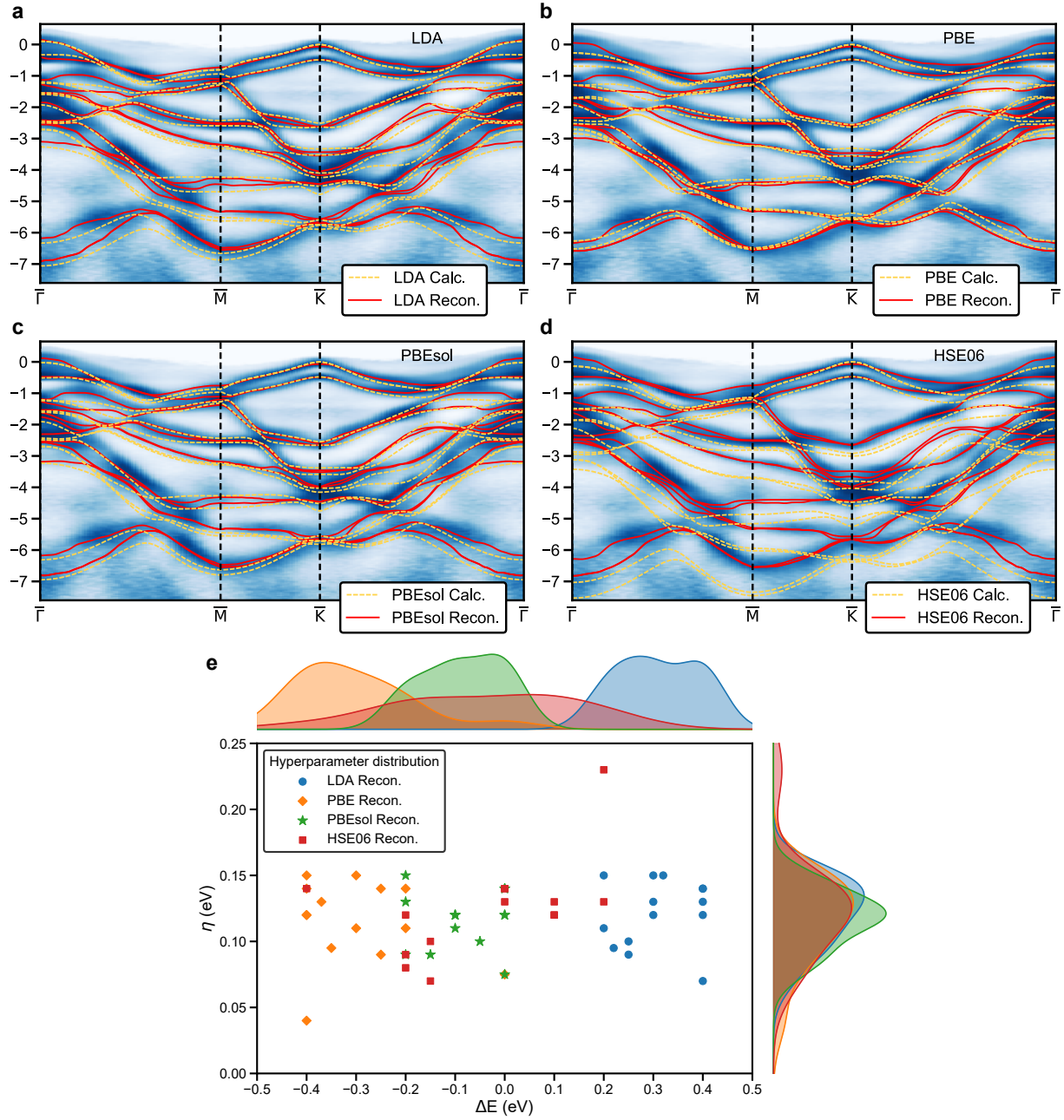
$$\text{const} \approx \frac{(\tilde{E}_{i+1,j} - \tilde{E}_{i,j})^2}{\eta^2} \approx \left(\frac{\partial E}{\partial k_x} \right)^2 \frac{\Delta k_x^2}{\eta^2}. \quad (18)$$

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 WSe₂, 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 WSe₂ (see Supplementary Fig. 4), 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, which represents a relatively broad acceptance range to yield a good reconstruction. Although this aspect is dependent on the data, in our experience, the hyperparameter tuning may be carried out in semi-automated fashion



Supplementary Figure 4: **Demonstration of hyperparameter tuning.** An example of tuning the hyperparameters, the rigid energy 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.



Supplementary Figure 5: **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. 12) is applied globally to all bands to align the energy zero at the \bar{K} point with the reconstruction. **e**, The distribution of hyperparameters used for the reconstruction in **a-d**.

guided by visualization and heuristics. Typically, 10-20 trials are sufficient to yield a good reconstruction, although a grid search may also be carried out for completeness. For a given dataset, the hyperparameters typically fall within a similar range, therefore, determining the range of hyperparameters need only be carried out once. The choice of hyperparameters is more flexible for reconstructing more isolated bands or those with fewer crossings, and *vice versa*. The band-wise reconstruction and the computational efficiency of the algorithm also enable further parallelization in hyperparameter tuning by distributing the optimization tasks in a high-performance computing infrastructure.

S1.5 Reconstructions using different theories as initializations

Comparison between reconstructed and theoretical band structures for 2H-WSe₂ are presented as a similarity matrix in the main 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. 5.

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.

S2.1 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. 6). 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 [78] and the tight-binding formulation of the two-band graphene band structure [79] as model band dispersion surfaces. The modified Griewank function takes the form,

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

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

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

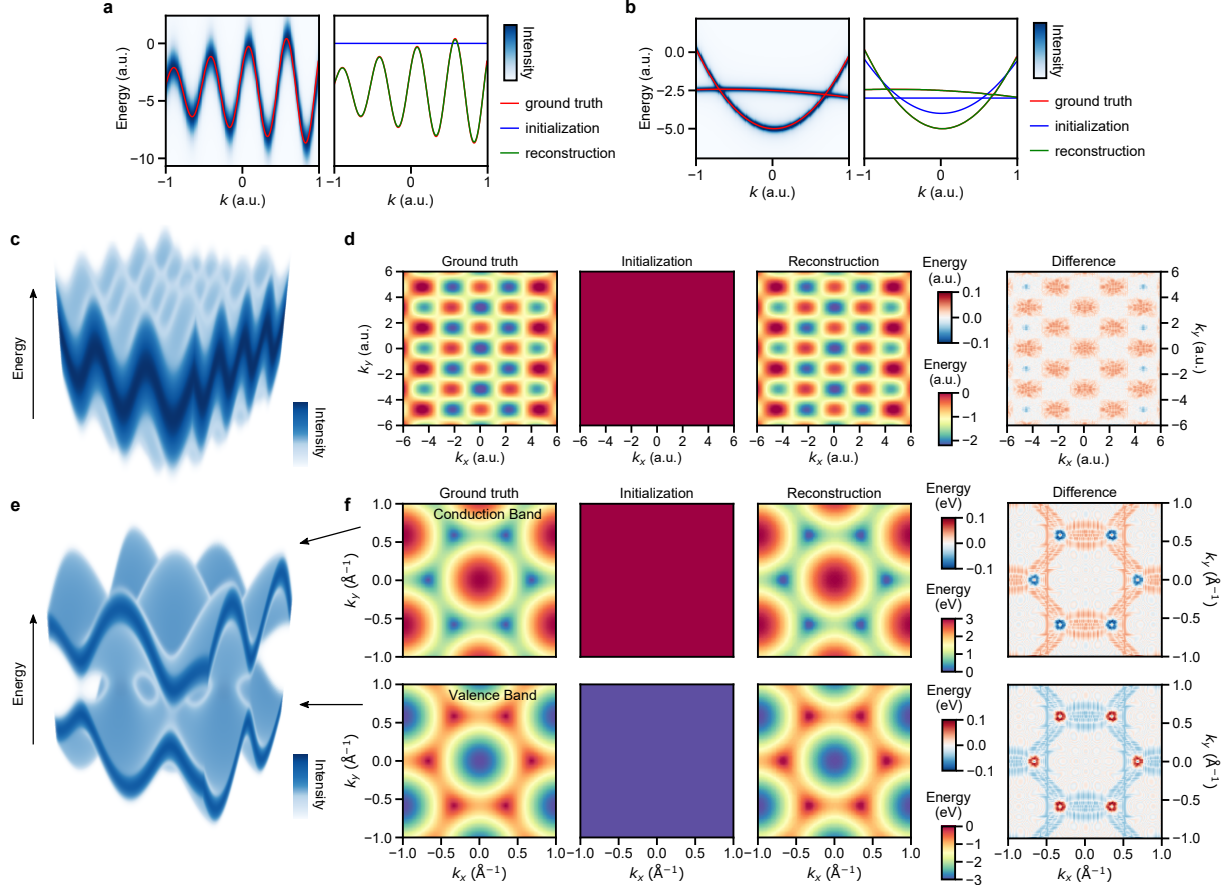
Here, E_+ and E_- refer to the conduction band and the valence band, respectively. Secondly, we used numerical band structures from DFT calculations with different exchange-correlation functionals (see section S4). They are more physically realistic, but also require more computation to obtain than generating bands from analytic functions.

S2.2 Initialization tuning

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 WSe₂, 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) [34, 80]. 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 tuning, we used DFT calculations with different exchange-correlation functionals (see section S4).

S2.3 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 [81] (with Gaussian and Lorentzian parameters σ and γ , and amplitude B) computed using the Faddeeva function W [82]. The Voigt profile approximates the convolution of a



Supplementary Figure 6: **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)) [78]. **e,f**, Reconstruction of the two bands of graphene nearby its Fermi level (**e,f**) formulated in the tight-binding model (see Eq. (20)) [79]. 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.

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).

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

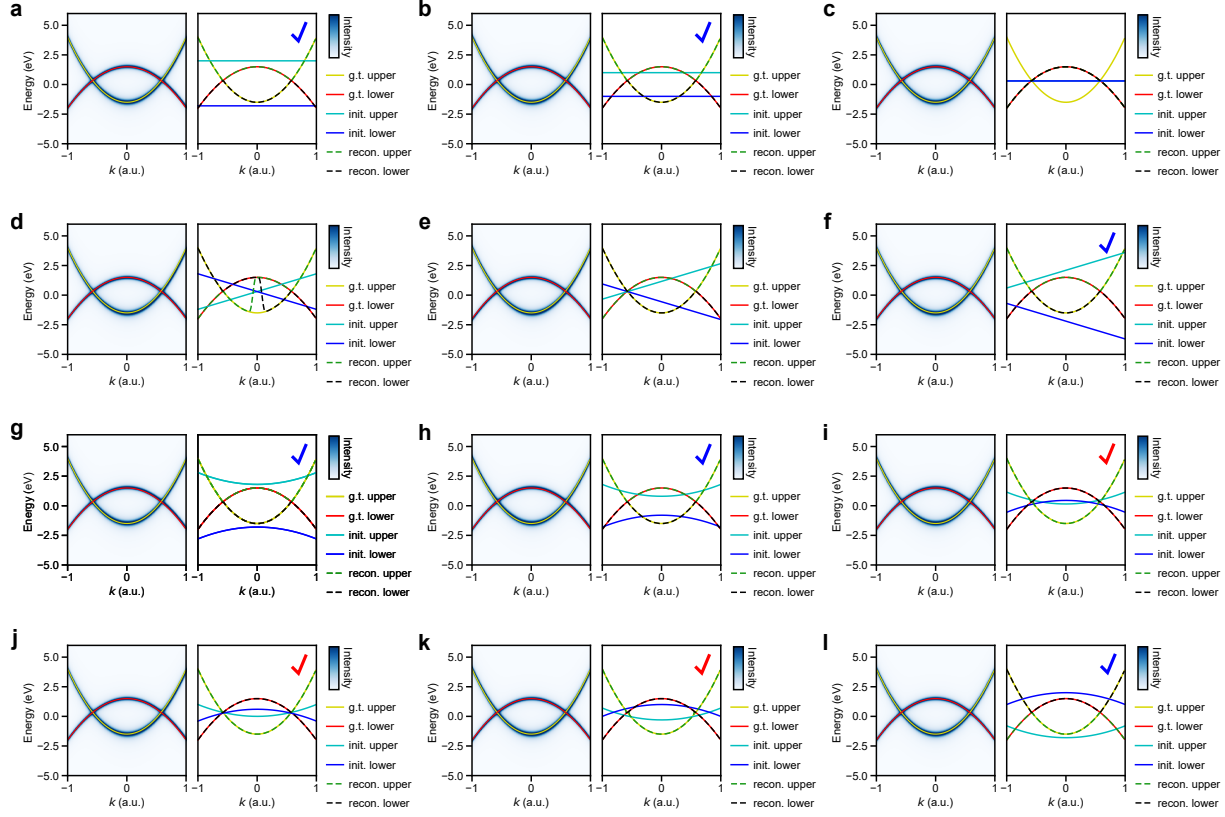
Without loss of generality, we assume the energy resolution in detection for all bands to be the same ($\sigma_j = \sigma$). For the cases shown in Supplementary Figs. 6-9, the linewidth parameter γ are set to a constant throughout the band. In all synthetic data, we omitted the inhomogeneous intensity modifications in realistic photoemission data due to experimental factors such as the experimental geometry, sample condition, matrix element effect, photon energy, etc. This omission relies on the assumption that the essential preprocessing step, such as symmetrization and contrast enhancement [25] 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 [7, 55]) has a higher momentum resolution than the momentum spacing used in data binning or generation.

S2.4 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. 6); Using synthetic multiband data generated from the LDA-level DFT (LDA-DFT) band structures of WSe₂ (see section S4), we tested out the sensitivity of reconstruction to the initialization (see Supplementary Fig. 9). In this case, to capture sufficient physical realism similar to the photoemission band mapping of WSe₂ presented in the main text, we set the energy resolution parameter of $\sigma = 100$ meV, the lineshape parameter $\gamma = 50$ meV [83], 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:

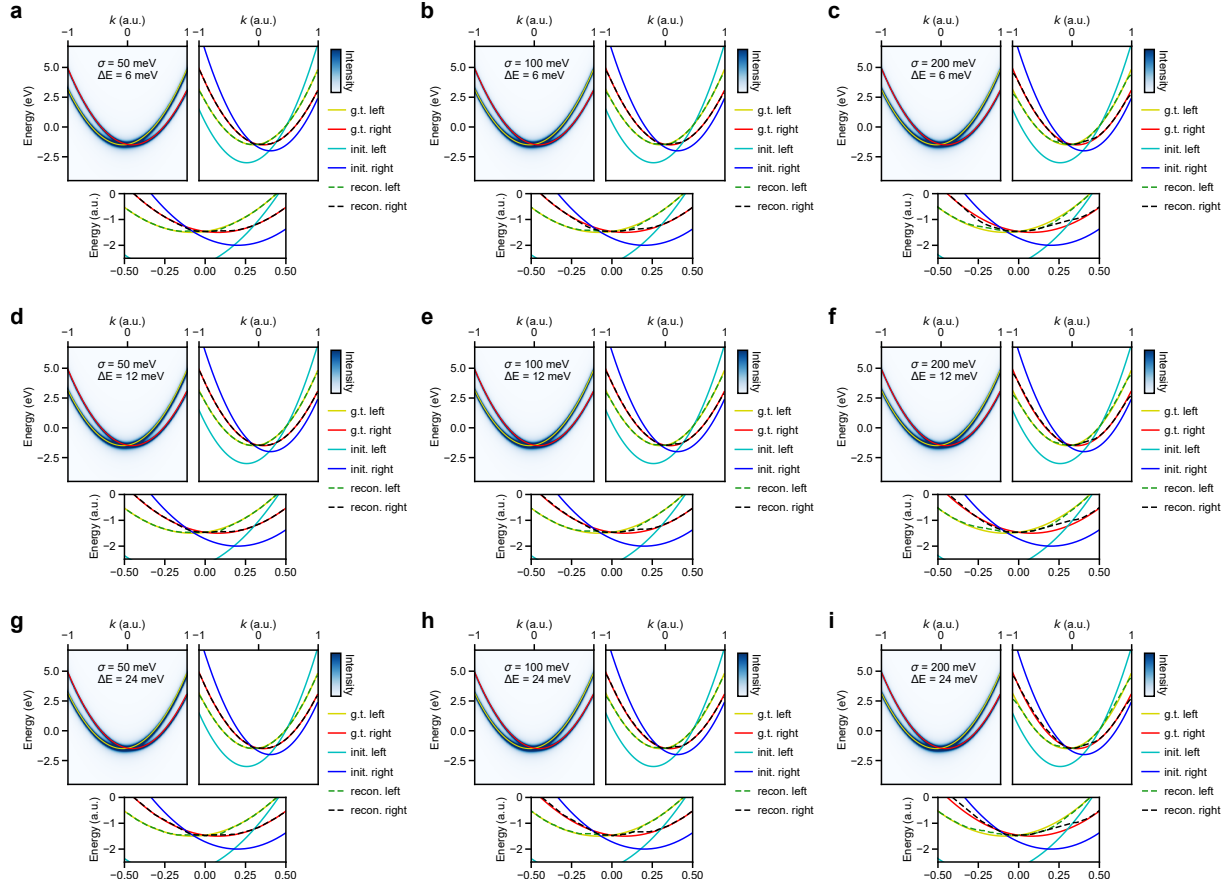
1. Reconstructing non-crossing bands: For isolated bands, we tested synthetic data constructed from a multi-sinusoidal band (Supplementary Fig. 6a), the band generated by the Griewank function (Supplementary Fig. 6c-d), and the two-band tight-binding model of graphene (Supplementary Fig. 6e-f). In these 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. 6b), a recurring pattern in band structures. To recover the dispersion without band index scrambling, the knowledge of



Supplementary Figure 7: Essential information in initialization for reconstruction. Results from a series of numerical experiments for demonstrating the effects of band-crossing information in the initialization. For clarity, the results are compared against the ground truth (g.t.) band dispersion – double-crossing parabolas – by overplotting in dashed lines. The tuning involves initializing the reconstruction with three sets of common curves: **a-c**, parallel straight lines, **d-f**, single-crossing straight lines, and **g-l**, double parabolas. The red check marks (✓) label the reconstructions with correct crossings, while the blue check marks (✓) label those with anti-crossings. All numerical experiments used the same simulated data from a toy model with double-crossing parabolas containing only the second and zeroth-order terms. For reconstruction experiments, the nearest-neighbor Gaussian width hyperparameter (η) in the MRF model is tuned, while the relative position of the initial conditions is shifted to each configuration.

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 double-crossing parabolas, the initializations that yield feasible outcomes are generated by slight tuning of the parabola parameters in the range that retains the crossing. A careful examination over possible scenarios largely confirms this



Supplementary Figure 8: The effects of data resolution on reconstruction. Results from a series of numerical experiments for demonstrating the effects of data resolution in either instrument resolution (σ) or energy spacing of data (ΔE) on the accuracy in reconstruction. The results are compared against the ground truth (g.t.) band dispersion – displaced parabolas – by overplotting in dashes lines. The σ parameter is tuned to 50 meV, 100 meV and 200 meV, while the ΔE parameter to 6 meV, 12 meV and 24 meV. In **a-i**, the synthetic data with ground truth dispersion is shown on the left, the reconstruction outcome is displayed on the right, along with a zoomed-in view near the crossing placed at the bottom. Quantitative values of the reconstruction error are given in Supplementary Table 1.

intuition: (1) Initialization with parallel straight lines (without any crossing) only results in non-crossing bands in the reconstruction (Supplementary Fig. 7a-c). When the initial straight line

contains the crossings in the ground truth, a symmetry breaking in the reconstruction takes place (Supplementary Fig. 7c), depending on the data and the Gaussian width hyperparameter (η). (2) Initialization with two straight lines containing a single crossing yields a reconstruction with at most a single crossing (Supplementary Fig. 7d-f). (3) Initialization with double parabolas yields a reconstruction with at most the same number of crossings within the range of the data (Supplementary Fig. 7g-l). When the reconstruction is successful, the crossings in the initialization are close to the intersection between the two parabolas. Besides, double-crossing parabolas with other parameters from those in Supplementary Fig. 7 are tested and similar outcomes are obtained.

Supplementary Table 1: **Reconstruction error in resolution tuning experiments.** For each band, the reconstruction error is the root-mean-square error per momentum spacing (unit in meV) between reconstruction and the ground truth, according to Eq. (8). In each numerical experiment, the tabulated reconstruction error is averaged over the corresponding two parabolic bands shown in Supplementary Fig. 8. The columns are the instrument resolution (σ) and the rows are the energy spacing (ΔE) used to generate the intensity data.

	$\sigma = 50$ meV	$\sigma = 100$ meV	$\sigma = 200$ meV
$\Delta E = 6$ meV	2.1	3.7	7.5
$\Delta E = 12$ meV	2.3	3.8	7.9
$\Delta E = 24$ meV	2.4	4.6	10.4

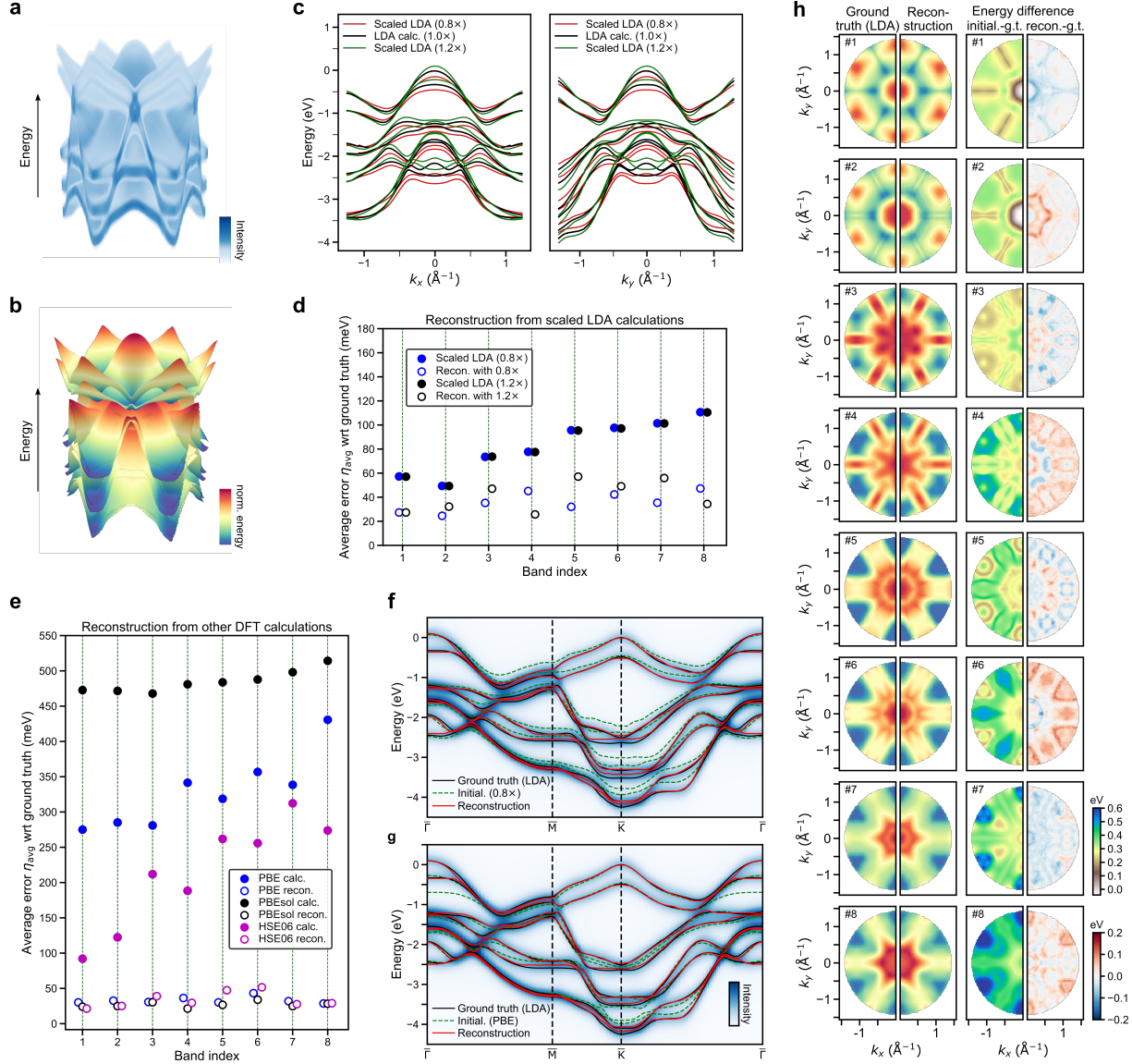
The crossing-band model is also an effective test case for resolution effects of the reconstruction algorithm. In this case, a momentum shift is introduced to two parabolic bands to produce the crossing, similar to the Rashba-split surface states of Au [84], which is often used to calibrate experimental resolution in photoemission studies. We conduct a series of numerical experiments using different width of the instrument resolution and energy spacing to simulate the resolution effect in the synthetic data, using reasonable parameter values. All numerical experiments use a nearest-neighbor Gaussian width hyperparameter η within $[0.08, 0.11]$ for the reconstruction and no rigid energy shift is introduced. We tabulate the outcomes visually in Supplementary Fig. 8 as gridded figures and quantitatively in Supplementary Table 1 using the reconstruction error (root-mean-square error between ground truth and reconstruction) with unit in meV. These results show that the reconstruction accuracy, as quantified by the error, has the same trend as the data resolution, which is determined by both the instrument resolution and energy sampling. The instrument resolution appears to have a larger effect on the reconstruction

than the energy spacing. In other words, the worse the data resolution ($\sigma = 200$ meV and $\Delta E = 24$ meV being the worst case), the higher the reconstruction error. From visual inspection of the reconstruction in Supplementary Fig. 8, including the zoomed-in region where the crossing is present, it appears that these changes in reconstruction accuracy create essentially no difference in the band dispersion away from the band crossing and only a marginal difference in the vicinity of the band crossing.

3. Sensitivity of reconstruction to scaled energies as initialization: We scaled the energies of the LDA-DFT band structure of WSe₂ (using the first 8 valence bands) around the mean energy of each band (see Supplementary Fig. 9c) 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. 9d,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 WSe₂ with PBE, PBEsol and HSE06 exchange-correlation functionals (see section S4) 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. 9e,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 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 [71] to formulate the reconstruction problem, albeit at the cost of much increased computational demand.

The results of the above numerical experiments demonstrate that the reconstruction by MAP



Supplementary Figure 9: **Validation on 3D synthetic multiband photoemission data.** **a**, Synthetic photoemission data with **b**, the underlying band structure obtained from LDA-level DFT calculation of WSe₂ (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 band.

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).

S2.5 Computational benchmarks

We used the available synthetic photoemission datasets based on the computed band structure of WSe_2 to construct benchmarks. The synthesis made use of the approach described in section S2.1. The two datasets used here, taken from [85], exhibit different characteristics, which may be qualitatively described using the energy range of an energy band. The more overlap in energy range between two energy bands, the more likely they have crossings (or anti-crossings). The dataset-specific information is as follows:

- The synthetic dataset of the WSe_2 K-point shows close proximity in energies between neighboring momentum locations. The energy ranges of all energy bands have no or up to a moderate degree of overlap. The dataset size is $30 \times 30 \times 500$ and contains 900 photoemission spectra.
- The synthetic high-symmetry line dataset of WSe_2 exhibits large dispersion. Since the high-symmetry line often represents the direction with the most dispersion in the band structure, the energy ranges of all energy bands are strongly overlapping. The dataset size is 186×500 and contains 186 photoemission spectra.

In both cases, the ground-truth band dispersions are taken from the LDA-DFT calculation, including all 14 valence bands, while the initializations for benchmarking both band reconstruction approaches are the PBE-DFT calculation (partial example see Supplementary Fig. 9g). Using these two datasets, we compare the reconstruction algorithms based on pointwise fitting (using the code in [39]) and MRF as introduced in this work. The hyperparameters for the pointwise fitting involves only the band-wise relative shifts applied in each band initialization (14 hyperparameters in total for 14 bands), which were tuned for each band sequentially from band #1 to band #14 using an expanding window approach (introduced in [39]). For the MRF reconstruction, the hyperparameters (including the band-wise shift and the width of the nearest-neighbor Gaussian prior, 28 hyperparameters in total for 14 bands) were tuned individually while reconstructing each band. The hyperparameter tuning made use of grid search through

a range of preset values, using the root-mean-squared (RMS) error for determining the final choice. All benchmarks were run on a on-premises computing server (Dell PowerEdge R840), equipped with four Intel Xeon Gold 6150 multicore CPUs.

Supplementary Table 2: **Algorithm comparison using benchmark datasets.** Two synthetic datasets with different number of spectra (N_{spec}) and the range of band indices (R_{band}) are used for benchmarking the algorithm performance. The per-band, per-spectrum reconstruction error (η_{bk}) is calculated using Eq. (22). The instability (r_{band}) quantifies the variation of the fitting residuals among all spectra within a dataset using the standard deviation of residuals, as in Eq. (23). The single-run time (t_{mono}) is the averaged elapsed time in a single execution of fitting, while the tuning time (t_{tune}) is the total time used for tuning the parameters to reach the final outcome. Both methods use DFT calculation as the initialization for the band positions.

Dataset ¹	N_{spec}	R_{band}	Pointwise line fitting ²				MRF reconstruction			
			t_{mono} (s)	t_{tune} (s)	η_{bk} (eV)	r_{band} (eV)	t_{mono} (s)	t_{tune} (s)	η_{bk} (eV)	r_{band} (eV)
WSe ₂ K point	900	1-2	42	421	4.6e-4	3.2e-4	6.6e-1	34	5.8e-3	4.1e-3
		3-4	168	2519	3.6e-1	1.4e-1	1.0	96	1.1e-2	5.5e-3
		5-8	412	11964	1.5e-1	5.3e-2	2.2	134	6.0e-2	2.1e-2
		9-14	2792	78181	3.6e-1	9.6e-2	3.8	236	7.8e-2	2.1e-2
WSe ₂ high-symmetry line	186	1-2	13	191	3.6e-1	2.1e-1	3.9e-1	32	1.1e-2	2.0e-2
		3-4	46	692	6.2e-1	2.9e-1	2.9e-1	31	1.9e-2	1.8e-2
		5-8	385	8858	5.5e-1	1.7e-1	8.6e-1	56	3.0e-2	1.4e-2
		9-14	872	27889	3.3	8.7e-1	1.6	109	4.1e-2	1.1e-2

¹ Datasets are obtained from [85].

² Executed using the software described in [39].

The computational performance of the two algorithms were evaluated using four different metrics as summarized in Supplementary Table 2. The timing metrics provided in the table include the average single-run computing time in each dataset as well as the total computing time of the hyperparameter tuning, which covers all grid search steps of the energy bands (indices described in R_{band}) in every benchmarking stage. The computing time for the single runs of each dataset show a clear advantage of the machine learning-based algorithm and the gap between the two algorithms only widens as the number of bands increases. The accuracy of the reconstruction is quantified by an RMS error averaged over all reconstructed bands and spectra,

following the expression for “band delta” in [67].

$$\eta_{\text{bk}}(E_{\text{gt}}, E_{\text{recon}}) = \sqrt{\frac{1}{N_b N_{\text{spec}}} \sum_{i=1}^{N_b} \sum_{\mathbf{k}} (E_{\text{gt},i,\mathbf{k}} - E_{\text{recon},i,\mathbf{k}})^2}, \quad (22)$$

where N_b is the number of bands and the subscript i is the band index. The instability is quantified by the standard deviation of the residual (difference between the ground truth and reconstructed energy dispersion), $\delta E = E_{\text{gt}} - E_{\text{recon}}$.

$$r_{\text{band}}(E_{\text{gt}}, E_{\text{recon}}) = \sqrt{\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{\mathbf{k}} (\overline{\delta E_{i,\mathbf{k}}^2} - \overline{\delta E_{i,\mathbf{k}}})^2}, \quad (23)$$

where the overline indicates the mean. This metric has been used in earth sciences to quantify surface roughness [41, 86], which may be interpreted similarly in our case. The difference is that the roughness in reconstructed surface is largely due to the instability of the optimization, besides the quality of the data, because band dispersions are generally smooth and continuous. In the main text Fig. 4, these tabulated metrics are normalized by the number of spectra to allow comparison between datasets, as is also adopted in [39]. We interpret the results in Supplementary Table 2 in the following two aspects:

- Computing time (t_{mono} and t_{tune}): For the same dataset, the single-run computing time of the MRF reconstruction is about 2-3 orders of magnitude faster than distributed pointwise fitting. Even with the hyperparameter tuning included, the MRF reconstruction still runs 1-2 orders of magnitude faster, although the MRF reconstruction requires tuning one more hyperparameter than the pointwise fitting approach for each band.
- Reconstruction quality (η_{band} and r_{band}): The significantly higher reconstruction error and instability for pointwise fitting is due to the lack of connectivity between neighbors come primarily from the (theoretical) initialization. Because for each band, even though a global energy shift hyperparameter is tuned, it cannot guarantee that everywhere locally the shift is optimal for band reconstruction, resulting in scrambled band indices in the local patches that the fitting fails. This scenario is a failure mode of the pointwise fitting-based reconstruction as illustrated in [39] for real-world experimental data. This limitation of pointwise fitting approach is less pronounced when the energy range overlap between bands is small, yet becomes more severe in the high-symmetry line dataset, where the strong energy range overlap and the multiple band

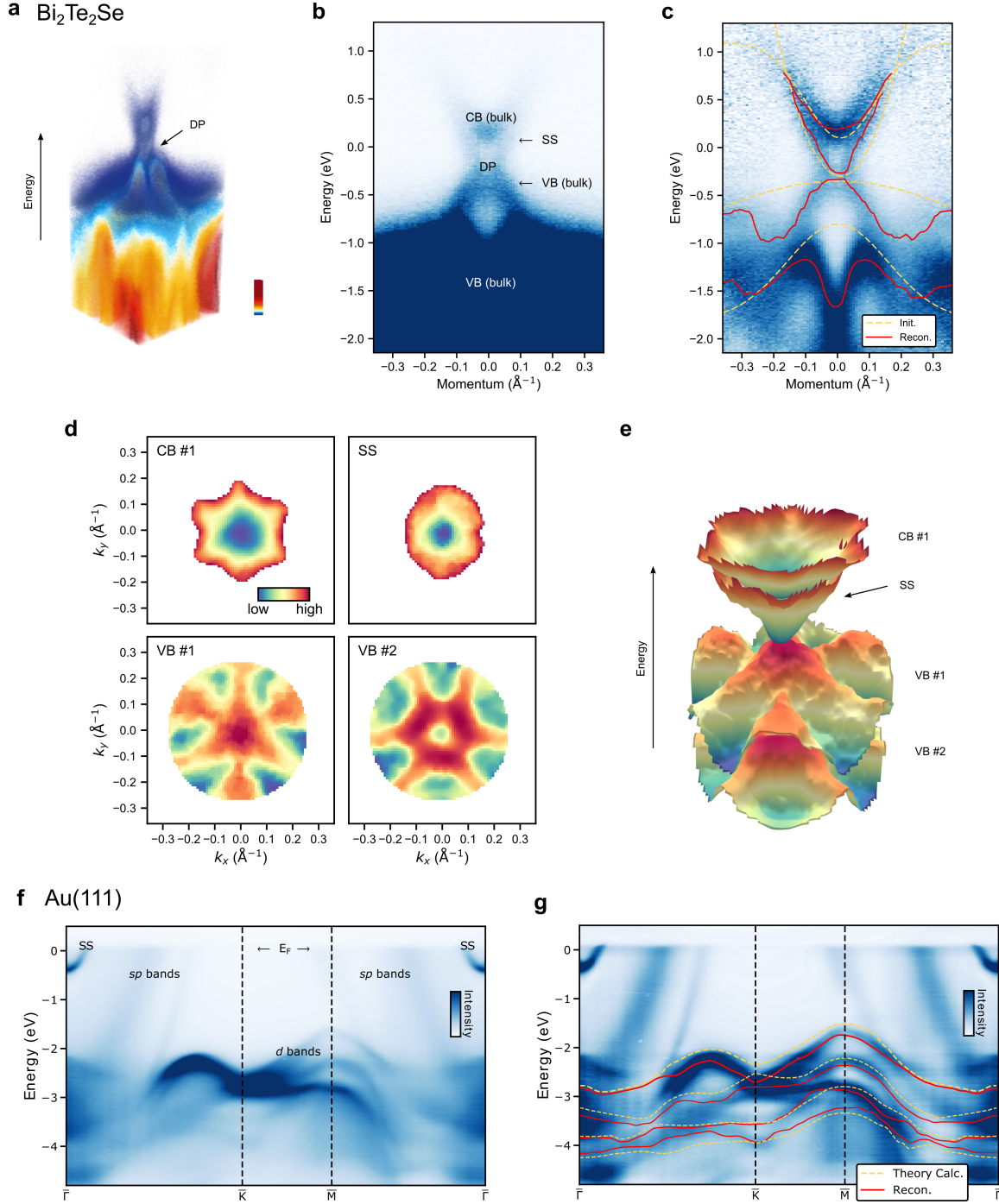
crossing (or anti-crossing) make the reconstruction harder to resolve by tuning a single energy shift parameter. The probabilistic framework of the MRF approach largely circumvents this limitation using a physical prior that accounts for the proximity of the neighboring energy values and achieves high stability in the reconstruction.

S3 Reconstruction for other experimental datasets

To test the functionality of our MRF reconstruction algorithm on other materials, we have acquired photoemission band mapping datasets from gold (Au), a metal, and bismuth tellurium selenide ($\text{Bi}_2\text{Te}_2\text{Se}$), a topological insulator. Due to the complexity of the electronic structure of these materials, we focus here on reconstructing a subset of the energy bands of these two materials that are pronounced within the measured energy range.

S3.1 Near-gap electronic bands of a topological insulator ($\text{Bi}_2\text{Te}_2\text{Se}$)

The dataset for $\text{Bi}_2\text{Te}_2\text{Se}$ was measured at room temperature at the Fritz Haber Institute in Berlin using a momentum microscope (SPECS METIS 1000). The sample growth method was previously described in [87]. A clean surface was prepared in vacuum by *in situ* cleaving with a Scotch tape. During the measurement, light excitation of 800 nm was used to examine ultra-fast dynamics. The temporal features were ignored here and averaged to improve the signal to noise ratio of the data. The photoemission spectra of $\text{Bi}_2\text{Te}_2\text{Se}$ near the Fermi level features a topologically-protected surface state (SS) that intersects at the Dirac point (DP) as shown in Supplementary Fig. 10b. The SS bridges the valence and conduction bands, an identifiable and prominent feature for this class of materials directly measurable via photoemission [35, 88]. Preprocessing of the 3D band mapping data follows the procedure for WSe_2 data described in the main text, except that the rotational symmetrization is only threefold, due to the symmetry of the material. We used numerical initializations from simple functions such as paraboloid and Gaussian in 2D, instead of any first-principles calculation. The reconstructed energy dispersions were smoothed using Chambolle’s total variation denoising algorithm [89] implemented in `scikit-image` [90], removing the high-frequency noise as a result of the Poisson statistics of the photoemission data. As shown in Supplementary Fig. 10c, the simple initializations we chose are sufficient to reconstruct the complex dispersion from the first two valence bands,



Supplementary Figure 10: **Band reconstruction for $\text{Bi}_2\text{Te}_2\text{Se}$ and $\text{Au}(111)$.** **a**, 3D view of the photoemission band mapping data of the topological insulator $\text{Bi}_2\text{Te}_2\text{Se}$ around the Dirac point (DP). The energy bands near the DP are labeled in **b** in a 2D slice through the DP. The outcome of reconstruction is superimposed on the preprocessed data in **c**. Momentum-resolved reconstruction is shown in 2D (**d**) and 3D (**e**), where the color map represents the energy values within each band. The experimental photoemission data for $\text{Au}(111)$ is shown in **f** with designations of the band structures labeled. Reconstruction of some of the d bands are shown in **g** along with the theoretical calculations used for initialization.

the SS and parts of the first conduction band occupied by excited electronic population. The appearance of the first conduction band for $\text{Bi}_2\text{Te}_2\text{Se}$ is a result of photoexcitation [91]. The reconstructed bands show sixfold symmetry and warping in agreement with previous theoretical investigations [35, 88], which is more straightforwardly visualized in 2D and 3D as in Supplementary Fig. 10d-e. For the dispersion surfaces of the SS and conduction band, we limited the band dispersion to realistic energy range not far from the photon energies of the excitation light pulses.

S3.2 Bulk electronic bands of gold (Au)

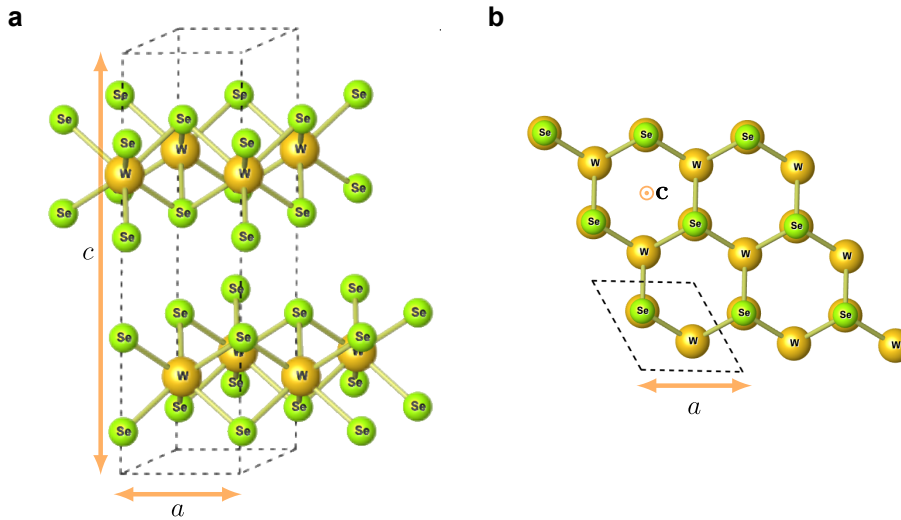
The Au dataset was measured at 100 K at the SGM-3 beamline [92] of the synchrotron radiation facility ASTRID2 in Aarhus, Denmark. The Au samples were purchased from MaTeCK GmbH with a (111) surface. The sample preparation procedure has been previously described [93]. The photoemission data were measured along the high-symmetry direction ($\Gamma\text{K}\text{M}\Gamma$) of Au(111), which exhibits a hexagonal symmetry in the surface Brillouin zone [94] (indicated with an overbar over each symmetry label) similar to WSe_2 . As shown in Supplementary Fig. 10f, the collection of energy bands present in the photoemission data for Au(111) include the surface state (SS), which have sufficient momentum and energy resolution, are composed of momentum-shifted parabolas [84]. The *sp* bands and the *d* bands are the low-energy bulk electronic bands of Au. The Au data has been preprocessed using contrast enhancement and intensity smoothing as described in the main text for the WSe_2 data before reconstruction. The reconstruction used existing DFT calculations, which feature a Au(111) slab containing five Au layers conducted according to [95], as initialization to retrieve parts of the *d* bands that are resolvable within the current dataset. The comparison between initialization and reconstruction is shown in Supplementary Fig. 10g.

S4 Band structure calculations

S4.1 DFT calculations

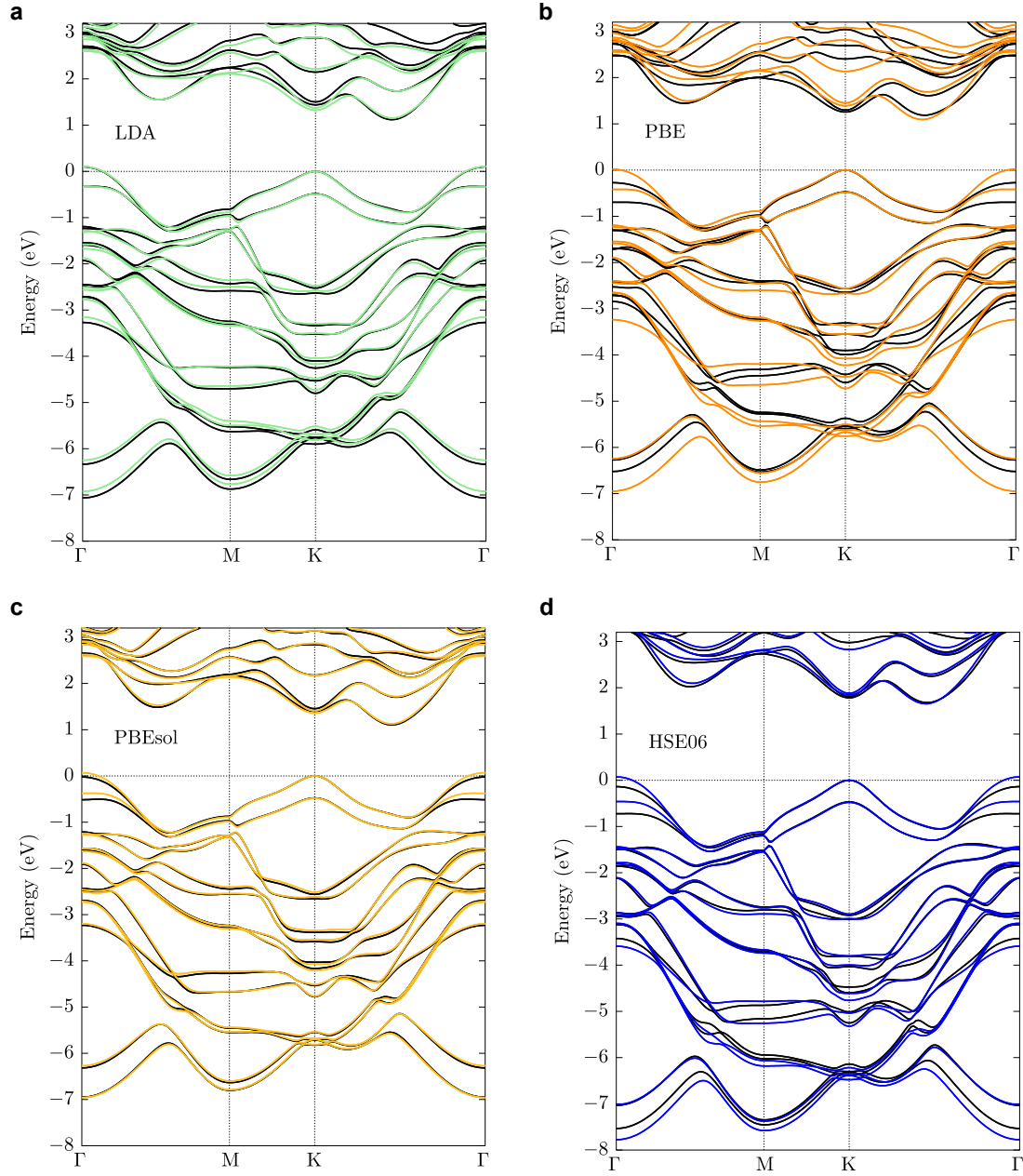
The crystal structure of bulk WSe_2 with 2H stacking (2H- WSe_2) belongs to the $\text{P6}_3/\text{mmc}$ space group and consists of two Se-W-Se triatomic layers as shown in Supplementary Fig. 11. The stacking order of the two hexagonal layers is $-\text{BAB}-\text{ABA}-$ and the long *c*-axis is oriented per-

pendicular 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 [66]. 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 $16 \times 16 \times 4$ uniform \mathbf{k} -grid 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 3 shows the optimized lattice constants, a and c , as obtained by the evaluation of the analytical stress tensor [96] 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-WSe₂ [97].



Supplementary Figure 11: **Crystal structure of bulk 2H-WSe₂.** **a**, Side view and **b**, top view of the crystal structure of 2H-WSe₂. 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.

The calculated BSs of bulk 2H-WSe₂ using different levels of approximation for the exchange-correlation (XC) functional are shown in Supplementary Fig. 12. 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



Supplementary Figure 12: **Bulk electronic band structure of 2H-WSe₂.** **a-d**, Band structure of bulk 2H-WSe₂ 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 3) 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 $k_z = 0.35 \text{ \AA}^{-1}$.

Supplementary Table 3: **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-WSe₂ 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 0.490 ²	0.473 ¹ 0.481 ²	0.476 ¹ 0.484 ²	0.467 ¹ 0.480 ²	0.5 ³
Band gap (eV)	1.022 ¹ 1.052 ²	1.186 ¹ 1.074 ²	1.105 ¹ 1.060 ²	1.679 ¹ 1.582 ²	1.219 ⁴

¹ Fully optimized structure.

² Optimized structure by fixing the lattice parameters to experimental values.

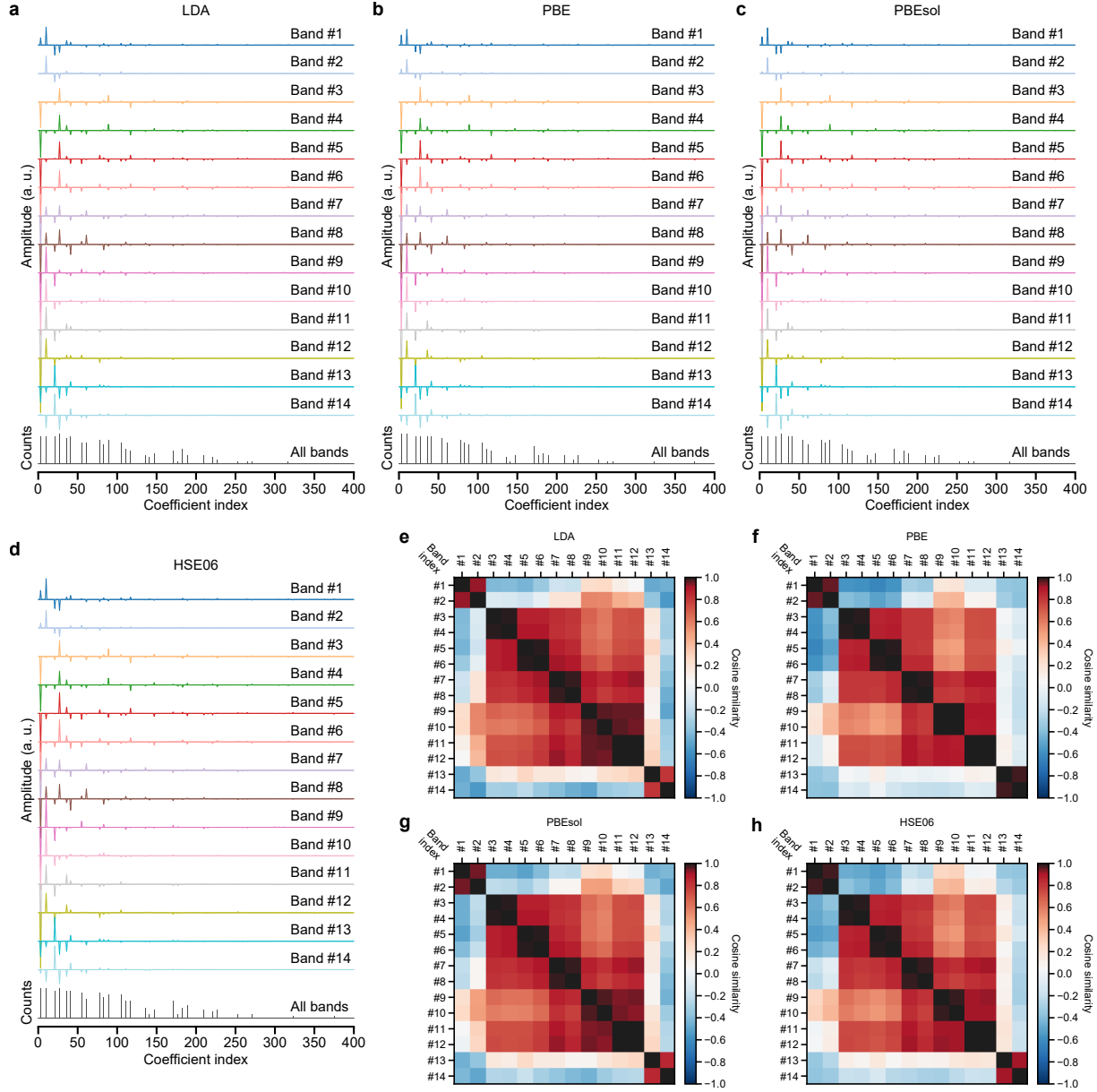
³ Ref. [10].

⁴ Ref. [98].

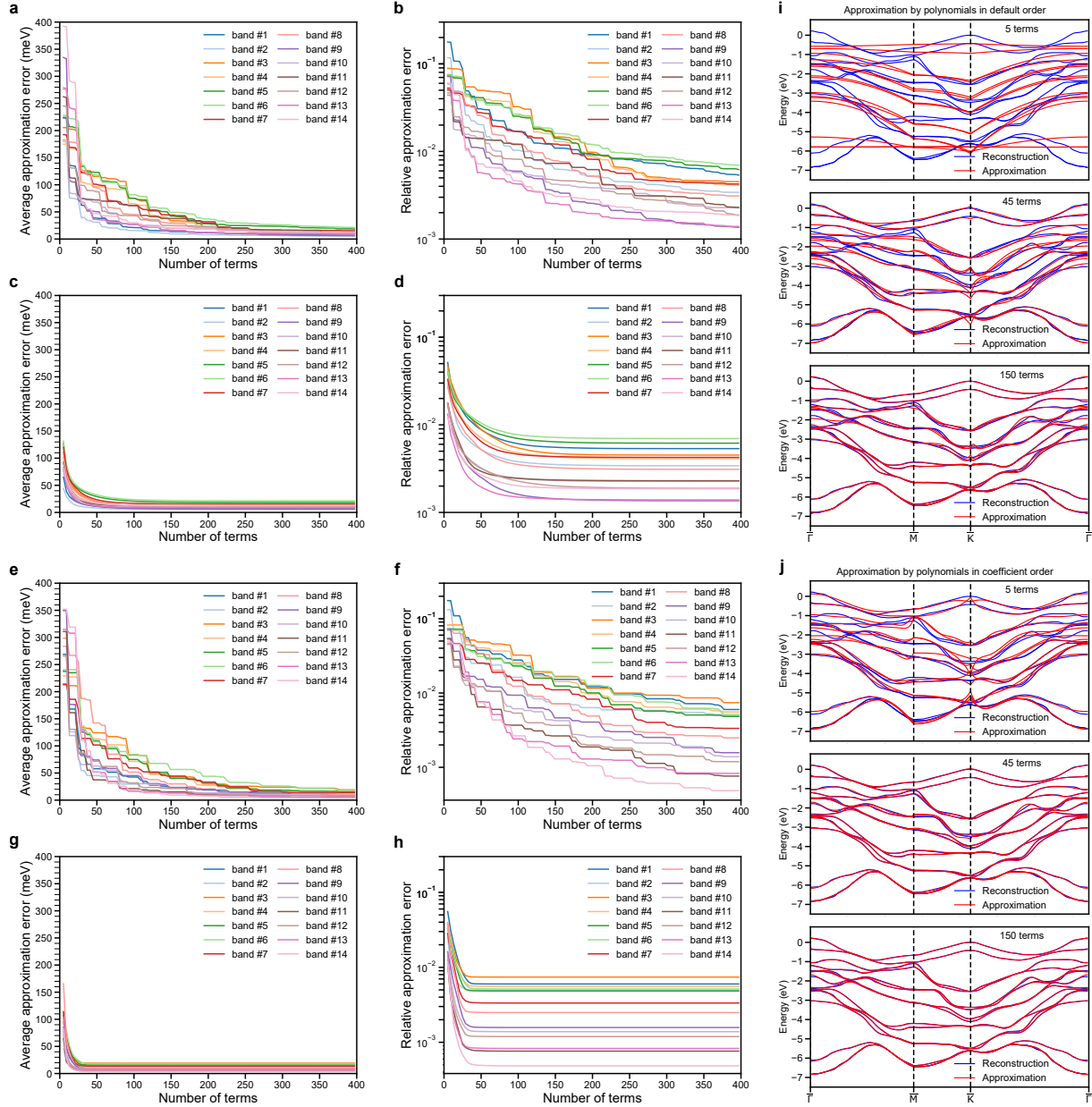
the conduction band minimum located along the Γ -K path (Γ and K being the bulk equivalents of the $\bar{\Gamma}$ and \bar{K} high-symmetry points). For both sets of optimized structures, the LDA results reveal a valence band maximum at the Γ point, compatible with experimental measurements, while the 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 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 [99]. 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 3.

S4.2 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-



Supplementary Figure 13: **Geometric featurization of the energy bands of WSe₂.** **a-d**, Decomposition of the 14 valence energy bands of WSe₂ 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 WSe₂ 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 14: **Approximation to the band structure of WSe₂ 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.

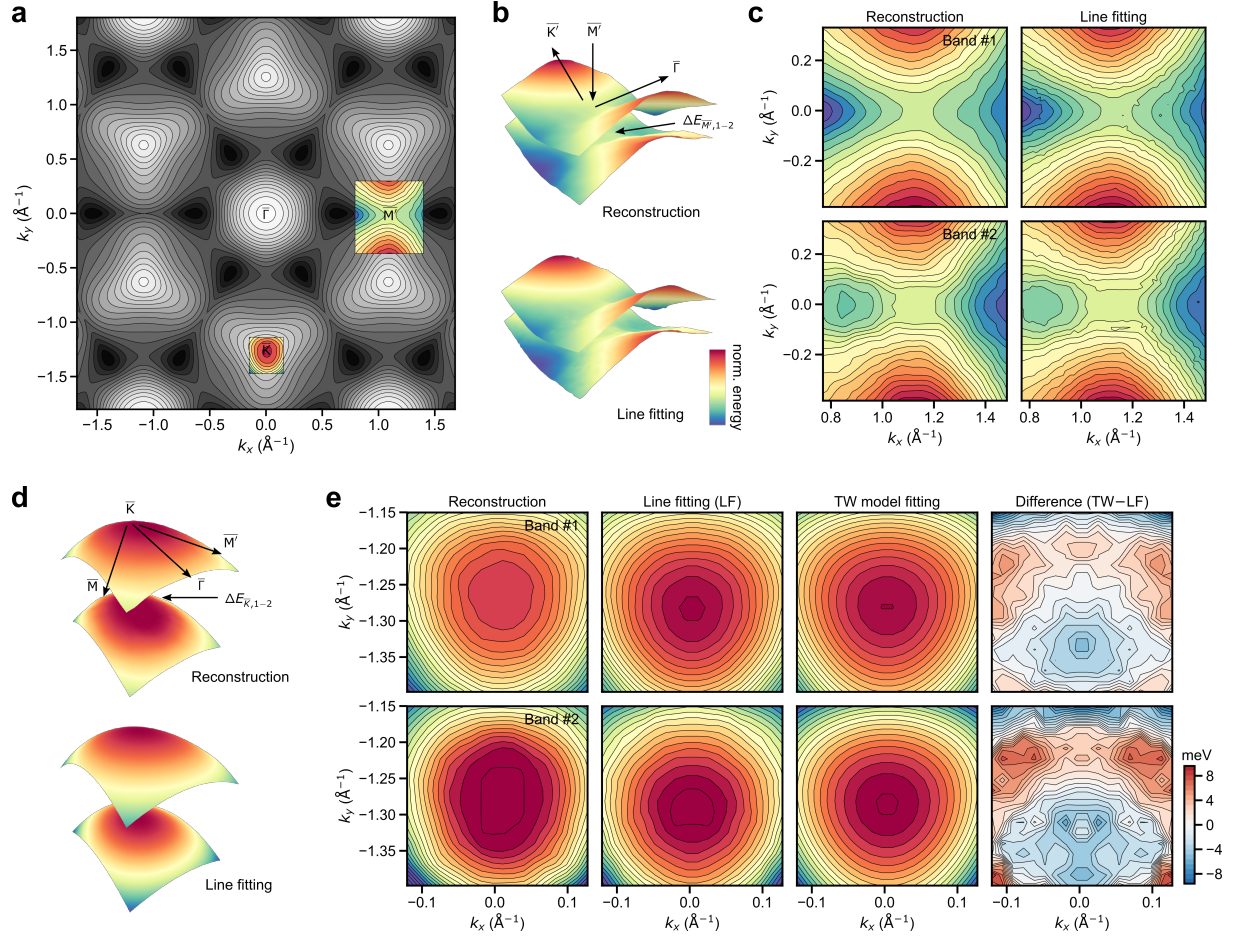
space rectangular patch that covers the Γ , K and M points of the Brillouin zone. The hexagonal Brillouin zone of WSe₂ has a sixfold rotation symmetry axis and two independent mirror planes in the (k_x, k_y) coordinates. The initial rectangular patch is first symmetrized about the two mirror planes in 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.

S5 Band structure informatics

S5.1 Global scale structure descriptors

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. 13 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 WSe₂ band structure.

In Supplementary Fig. 14, we demonstrate numerically the approximation capability of the hexagonal ZP basis set to all 14 valence bands of WSe₂. Despite the stark differences in energy dispersion, the approximation to reconstructed bands (Supplementary Fig. 14a-d) and theoretical band structure at the level of LDA-DFT (Supplementary Fig. 14e-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 the energy axis in the data (~ 18 meV) that results in the imperfect smoothness of the reconstructed 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. 14i-j in the momentum path



Supplementary Figure 15: **Local band structure parameters.** **a**, The first valence band of WSe₂ with constant-energy contours. The patches around high-symmetry points \bar{K} and \bar{M}' from reconstruction (with LDA-DFT as the initialization) are overlaid in color. **b,c**, Patch around the \bar{M}' -point, a saddle point in the dispersion surface, visualized in 3D (**b**) and 2D (**c**), respectively. The energy gap at \bar{M}' due to spin-orbit coupling (SOC) results in the energy difference $\Delta E_{\bar{M}',1-2}$. **d,e**, Patch around the \bar{K} -point, the energy maximum of the valence band, visualized in 3D (**d**) and 2D (**e**), respectively. The SOC results in the energy gap $\Delta E_{\bar{K},1-2}$. The outcome of fitting to a trigonal warping (TW) model around \bar{K} from $\mathbf{k}\cdot\mathbf{p}$ theory [28] is shown in **e**.

along high-symmetry lines of the reconstructed band structure.

S5.2 Local scale structure descriptors

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 (\bar{K} , \bar{M}' , and $\bar{\Gamma}$, see Supplementary Fig. 15a) and compiled the results in Supplementary Table 4. 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 \bar{K} , two spectral peaks corresponding to two spin-split bands were fit simultaneously, while around \bar{M}' and $\bar{\Gamma}$, four were fit simultaneously due to the spectral proximity of the first four valence bands (see Supplementary Fig. 5). The fitting is carried out using a linear superposition of Voigt lineshapes and the `lmfit` package [100] 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 \bar{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 4 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 \bar{K} and \bar{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 (~ 18 meV in energy).

The region extracted around \bar{K} (see Supplementary Fig. 15d-e) contains about 10% of the distance of $\bar{\Gamma} - \bar{K}$. Due to the strong trigonal warping (TW) effect in this class of materials, the effective masses and the TW parameters around \bar{K} were fit simultaneously in 2D using the momentum-space model derived from $\mathbf{k} \cdot \mathbf{p}$ theory [28].

$$E(\mathbf{q}) = \frac{\hbar^2 \mathbf{q}^2}{2m_{\bar{K}}} + C|\mathbf{q}|^3 \cos(3\varphi_{\mathbf{q}} + \theta) + E_0. \quad (24)$$

Here, \mathbf{q} is the momentum vector \mathbf{k} recentered on a particular \bar{K} (or \bar{K}') point by translation, $m_{\bar{K}}$ is the effective mass of the hole at \bar{K} point, C is the magnitude of the TW (named C_{3w} in [28]), $\varphi_{\mathbf{q}}$ is the polar angle in the coordinate system centered on a \bar{K} (or \bar{K}') point, θ is an auxiliary fitting parameter used to accommodate the orientation of the TW with respect to the

Supplementary Table 4: **Band structure parameters from experiment and theory.** Effective masses of holes ($m_{\bar{K}}$), trigonal warping parameters (C) are extract at \bar{K} point in the first two valence bands. Two directional effective masses at \bar{M}' ($m_{\bar{M}'}$), and one at $\bar{\Gamma}$ ($m_{\bar{\Gamma}}$), are obtained for the first valence band. The energy gaps (ΔE) between the first two valence bands are obtained at both \bar{K} and \bar{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
\bar{K}	$m_{\bar{K},1}/m_e$	−0.62	−0.60	−0.49	−0.42
\bar{K}	$m_{\bar{K},2}/m_e$	−0.74	−0.78	−0.64	−0.54
\bar{K}	$C_{\bar{K},1}$ (eV·Å ³)	5.3	5.8	6.2	4.5
\bar{K}	$C_{\bar{K},2}$ (eV·Å ³)	4.0	3.9	3.9	3.2
\bar{K}	$\Delta E_{\bar{K},1-2}$ (meV)	419	446	485	467
\bar{M}'	$m_{\bar{M}'-\bar{\Gamma},1}/m_e$	0.71	0.72	0.25	0.17
\bar{M}'	$m_{\bar{M}'-\bar{K},1}/m_e$	−1.6	−1.5	−1.1	−0.90
\bar{M}'	$\Delta E_{\bar{M}',1-2}$ (meV)	352	338	127	48
$\bar{\Gamma}$	$m_{\bar{\Gamma},1}/m_e$	−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 3.

pixel coordinates defined by the rectangular region of interest, E_0 accounts for the energy offset. The energy gaps at \bar{K} ($\Delta E_{\bar{K},1-2}$) and \bar{M}' ($\Delta E_{\bar{M}',1-2}$) are illustrated in Supplementary Fig. 15 (b and d), respectively. The \bar{M}' (or \bar{M}) point situates at a saddle point of the dispersion surface (first valence band), as shown in Supplementary Fig. 15b-c. Its lower symmetry (compared with \bar{K} , \bar{K}' and $\bar{\Gamma}$) means that the effective masses exhibits anisotropy, with opposite signs and magnitude along the $\bar{M}' - \bar{\Gamma}$ and $\bar{M}' - \bar{K}'$ 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 [28]). The second valence band is not fitted at \bar{M}' due to the pronounced dispersion modulation by interband coupling unaccounted for in the existing saddle-shaped model. At around $\bar{\Gamma}$, a single effective mass is extracted by fitting a paraboloid to a local patch of the dispersion surface.

Supplementary Video 1

Left side shows the position of the cut viewed from the first projected Brillouin zone of WSe_2 . Right side shows the corresponding 2D cut in (k_y, E) coordinates from volumetric band mapping data, overlaid with the DFT calculation performed at the LDA level (LDA-DFT), used to initialize the reconstruction, and the resulting 14 reconstructed valence bands.

Supplementary Video 2

Left side shows the position of the cut viewed from the first projected Brillouin zone of WSe_2 . Right side shows the corresponding 2D cut in (k_x, E) coordinates from volumetric band mapping data, overlaid with the DFT calculation performed at the LDA level (LDA-DFT), used to initialize the reconstruction, and the resulting 14 reconstructed valence bands.

Supplementary Video 3

The movie explores the reconstructed valence bands from photoemission band mapping data on WSe_2 using LDA-level DFT calculation as the initialization. It illustrates the generation of an exploded view of the bands from the original reconstruction, the bands viewed collectively from different angles and the individual view of each band.

References

70. Hüfner, S. *Photoelectron Spectroscopy* 3rd Edition, 662 (Springer Berlin Heidelberg, 2003).
71. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
72. Hammersley, J. M. & Clifford, P. *Markov Fields on Finite Graphs and Lattices* Unpublished. 1971.
73. Besag, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 192–236 (1974).
74. Geman, S. & Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741 (1984).
75. Besag, J. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society: Series B (Methodological)* **48**, 259–279 (1986).
76. Pearl, J. *Probabilistic Reasoning in Intelligent Systems* 552 (Morgan Kaufmann, 1988).
77. Besag, J. *On the statistical analysis of nearest-neighbours* in *Proceedings of the European Meeting of Statisticians, Budapest* (1972).
78. Locatelli, M. A note on the Griewank test function. *Journal of Global Optimization* **25**, 169–174 (2003).
79. Bena, C. & Montambaux, G. Remarks on the tight-binding model of graphene. *New Journal of Physics* **11**, 095003 (2009).
80. Mahajan, V. N. & Dai, G.-m. Orthonormal polynomials for hexagonal pupils. *Optics Letters* **31**, 2462 (2006).
81. Van de Hulst, H. C. & Reesinck, J. J. M. Line Breadths and Voigt Profiles. *The Astrophysical Journal* **106**, 121 (1947).
82. Zaghoul, M. R. & Ali, A. N. Algorithm 916: Computing the Faddeyeva and Voigt Functions. *ACM Transactions on Mathematical Software* **38**, 1–22 (2011).
83. Dong, S. *et al.* Direct measurement of key exciton properties: Energy, dynamics, and spatial distribution of the wave function. *Natural Sciences* **1**, 1:e10010 (2021).
84. LaShell, S., McDougall, B. A. & Jensen, E. Spin Splitting of an Au(111) Surface State Band Observed with Angle Resolved Photoelectron Spectroscopy. *Physical Review Letters* **77**, 3419–3422 (1996).
85. Xian, R. P. *pesarxiv* <https://github.com/mpes-kit/pesarxiv>.
87. Mi, J.-L. *et al.* Phase Separation and Bulk p-n Transition in Single Crystals of Bi₂Te₂Se Topological Insulator. *Advanced Materials* **25**, 889–893 (2013).
88. Michiardi, M. *et al.* Strongly anisotropic spin-orbit splitting in a two-dimensional electron gas. *Physical Review B* **91**, 035445 (2015).

89. Chambolle, A. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision* **20**, 89–97 (2004).
90. Van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
91. Papalazarou, E. *et al.* Unraveling the Dirac fermion dynamics of the bulk-insulating topological system Bi₂Te₂Se. *Physical Review Materials* **2**, 104202 (2018).
92. Hoffmann, S., S ndergaard, C., Schultz, C., Li, Z. & Hofmann, P. An undulator-based spherical grating monochromator beamline for angle-resolved photoemission spectroscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **523**, 441–453 (2004).
93. Dendzik, M., Bianchi, M., Michiardi, M., Sanders, C. E. & Hofmann, P. Reconstruction-induced trefoil knot Fermi contour of Au(111). *Physical Review B* **94**, 201401 (2016).
94. Tusche, C., Krasnyuk, A. & Kirschner, J. Spin resolved bandstructure imaging with a high resolution momentum microscope. *Ultramicroscopy* **159**, 520–529 (2015).
95. Lin, I.-B., Sheu, T. W.-H. & Li, J.-H. Effects of exchange correlation functional on optical permittivity of gold and electromagnetic responses. *Optics Express* **22**, 30725 (2014).
96. Knuth, F., Carbogno, C., Atalla, V., Blum, V. & Scheffler, M. All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals. *Computer Physics Communications* **190**, 33–50 (2015).
97. Vo , D., Kr ger, P., Mazur, A. & Pollmann, J. Atomic and electronic structure of WSe₂ from *ab initio* theory: Bulk crystal and thin film systems. *Phys. Rev. B* **60**, 14311–14317 (20 1999).
98. Kam, K. K., Chang, C. L. & Lynch, D. W. Fundamental absorption edges and indirect band gaps in W_{1-x}Mo_xSe₂ (0 ≤ x ≤ 1). *Journal of Physics C: Solid State Physics* **17**, 4031–4040 (1984).
99. Desai, S. B. *et al.* Strain-Induced Indirect to Direct Bandgap Transition in Multilayer WSe₂. *Nano Letters* **14**, 4592–4597 (2014).
100. Newville, M. *et al.* lmfit/lmfit-py 1.0.0 <https://doi.org/10.5281/zenodo.3588521>. 2019.