Supporting Information

Polarization-controlled ultrafast carrier dynamics in sliding ferroelectrics revealed by machine learning-accelerated nonadiabatic dynamics simulations

Dong-Hui Xu[1], Shuangli Yue[1], Yitong Wang[1], Ronghui Tang[1], Xiang-Yang Liu[2], Li Zhang[1], Yunpeng Lu[3], Xian Wang[*1,3], Mingli Yang[4]

1. Institute of Atomic and Molecular Physics, Key Laboratory of High Energy Density Physics of Ministry of Education, Sichuan University, Chengdu 610065, China.

2. College of Chemistry and Material Science, Sichuan Normal University, Chengdu 610068, China.

3. School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 637371, Singapore.

4. Research Center for Materials Genome Engineering, Sichuan University, Chengdu 610065, China.

Contents of the Supporting Information

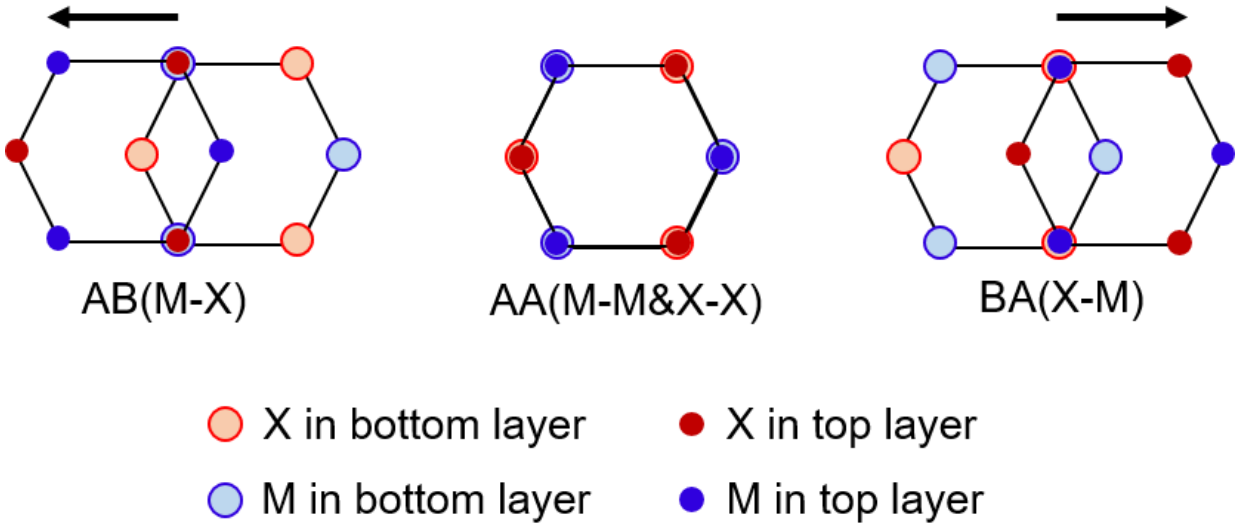

Figure S1. **Schematic illustration of the interlayer sliding pathway in the bilayer heterostructures.** The relative in-plane displacement between the two layers transforms the stacking sequence from AB (M-X) through the intermediate AA (M-M and X-X) configuration to BA (X-M). Here, M and X denote the metal and chalcogen atoms, respectively. Light-blue and dark-blue circles represent M atoms in the bottom and top layers, respectively, whereas light-red and red circles represent X atoms in the bottom and top layers, respectively. Black arrows indicate the direction of relative interlayer sliding.

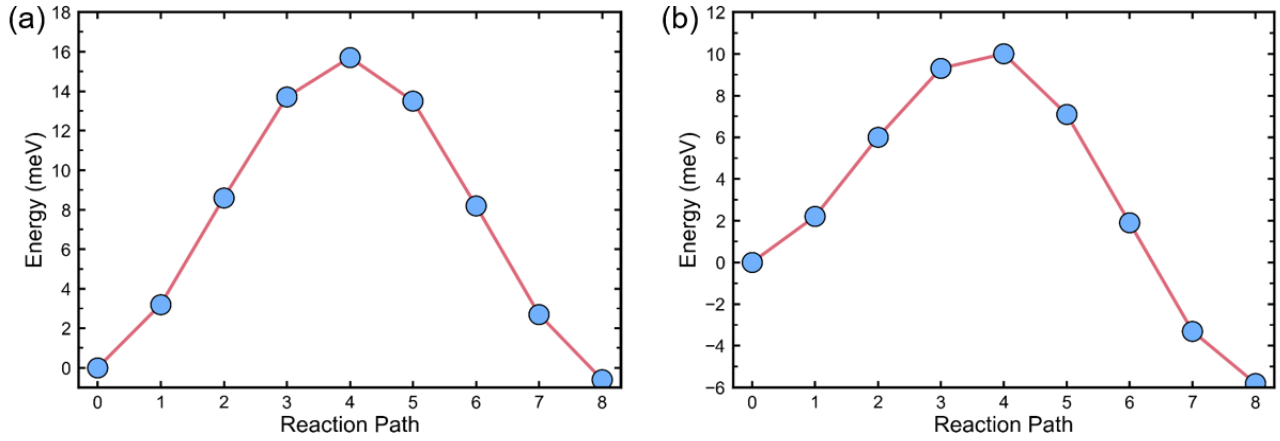

Figure S2. **Minimum-energy interlayer sliding pathways between the AB and BA stacking configurations.** Nudged elastic band (NEB) results are shown for **a**) $ZrTe_2/HfTe_2$ and **b**) $TiSe_2/HfSe_2$ along the AB-to-BA sliding pathway. Blue circles denote the calculated intermediate images along the reaction path, while the red lines connect the corresponding relative energies to guide the eye. The energy is referenced to the initial AB-stacked configuration, and the sliding barriers are reported in meV.

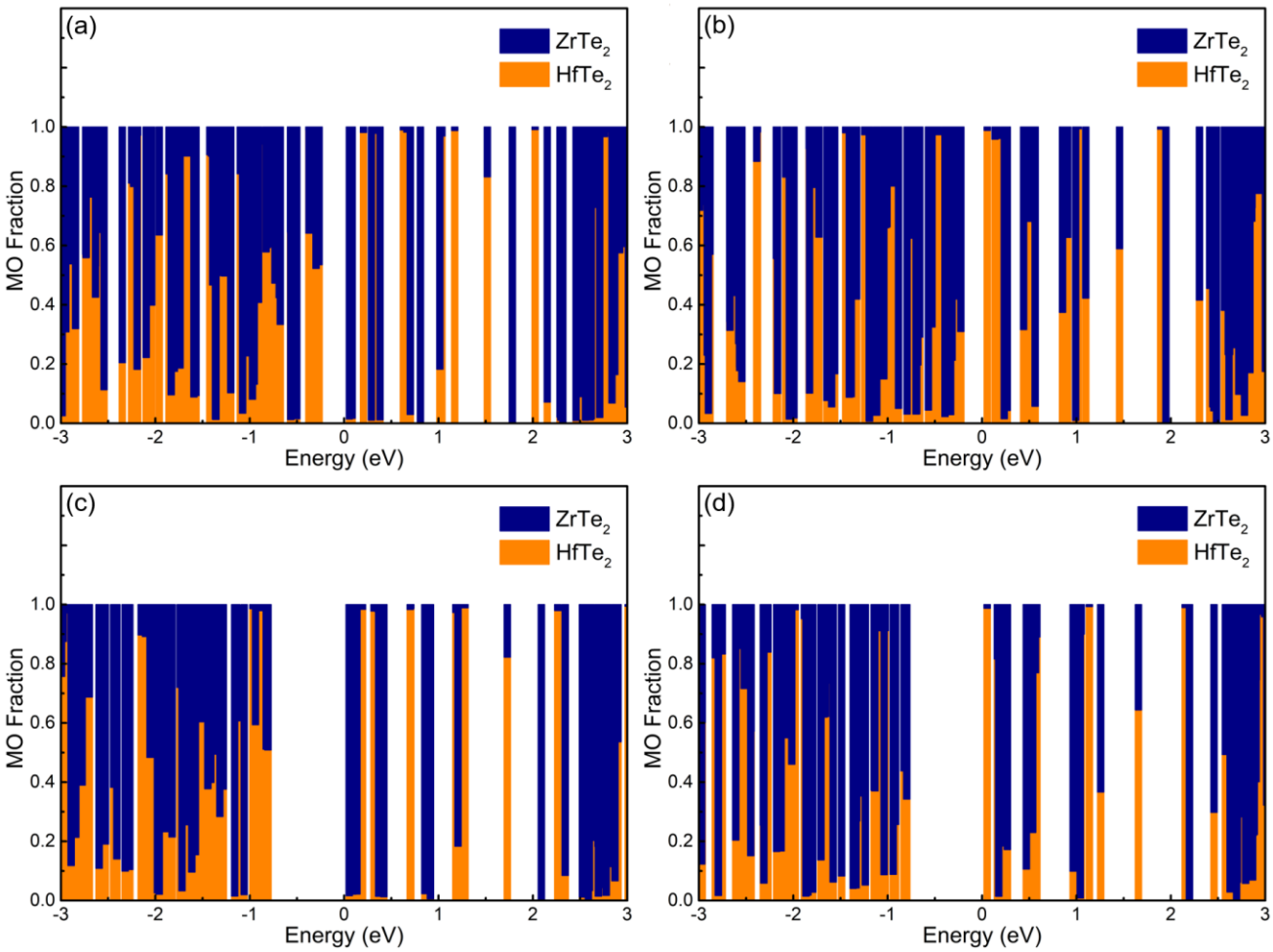

Figure S3. **Layer-resolved molecular orbital fractions of $ZrTe_2$/$HfTe_2$ under different stacking configurations and exchange-correlation treatments.** Molecular orbital (MO) fractions are shown for **a**) AB-$ZrTe_2$/$HfTe_2$ calculated using the Perdew-Burke-Ernzerhof functional with D3 dispersion correction (PBE+D3), **b**) BA-$ZrTe_2$/$HfTe_2$ using PBE+D3, **c**) AB-$ZrTe_2$/$HfTe_2$ using the HSE06 hybrid functional with D3 dispersion correction (HSE06+D3), and **d**) BA-$ZrTe_2$/$HfTe_2$ using HSE06+D3. Blue bars represent the fractional contribution of the $ZrTe_2$ layer to each molecular orbital, whereas orange bars represent the corresponding contribution of the $HfTe_2$ layer. The horizontal axis gives the orbital energy in eV, and the vertical axis gives the layer-resolved MO fraction.

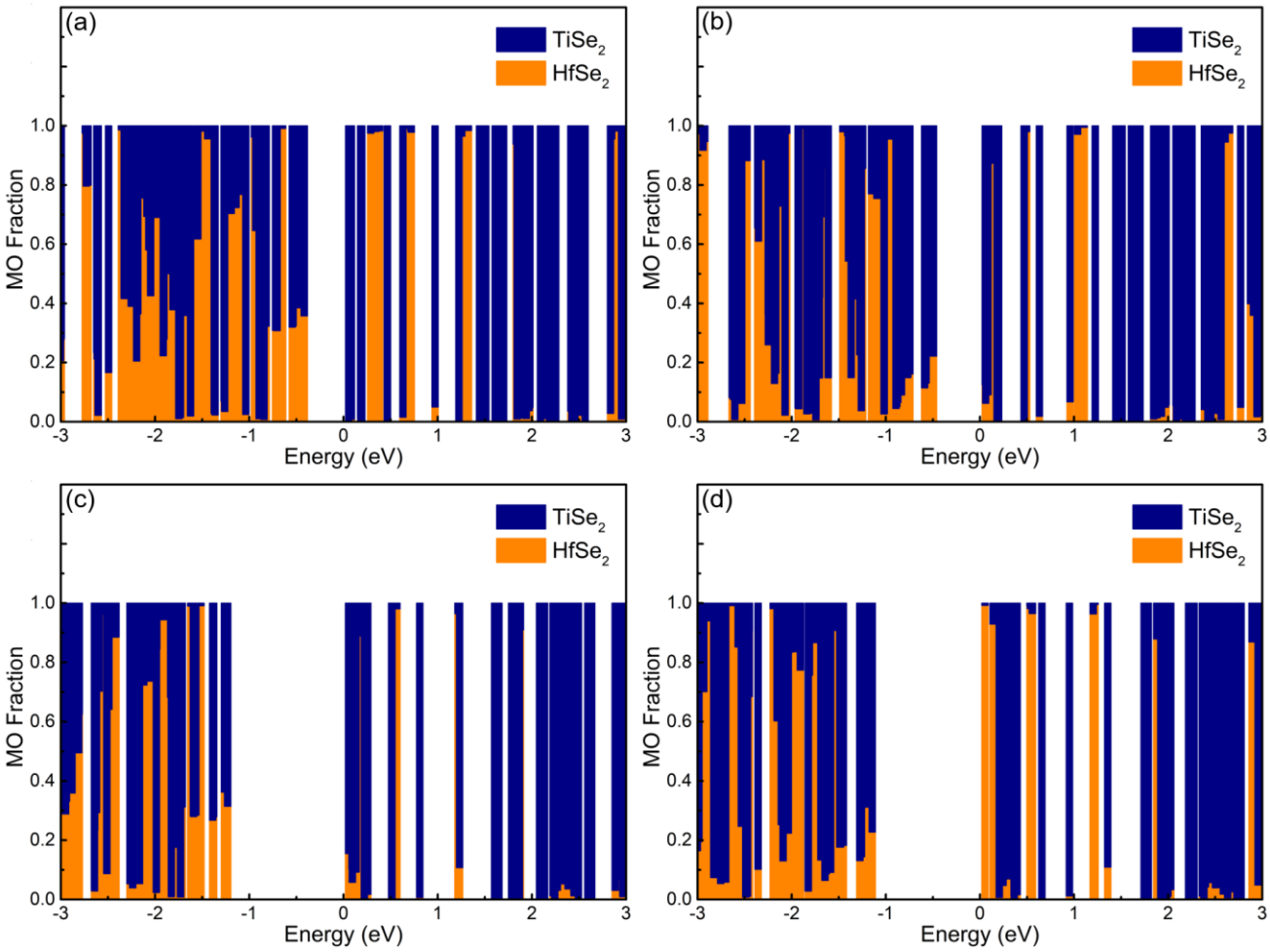

Figure S4. Layer-resolved molecular orbital fractions of $TiSe_2/HfSe_2$ under different stacking configurations and exchange-correlation treatments. Molecular orbital (MO) fractions are shown for **a**) AB-$TiSe_2/HfSe_2$ calculated using the Perdew-Burke-Ernzerhof functional with D3 dispersion correction (PBE+D3), **b**) BA-$TiSe_2/HfSe_2$ using PBE+D3, **c**) AB-$TiSe_2/HfSe_2$ using the HSE06 hybrid functional with D3 dispersion correction (HSE06+D3), and **d**) BA-$TiSe_2/HfSe_2$ using HSE06+D3. Blue bars represent the fractional contribution of the $TiSe_2$ layer to each molecular orbital, whereas orange bars represent the corresponding contribution of the $HfSe_2$ layer. The horizontal axis gives the orbital energy in eV, and the vertical axis gives the layer-resolved MO fraction.

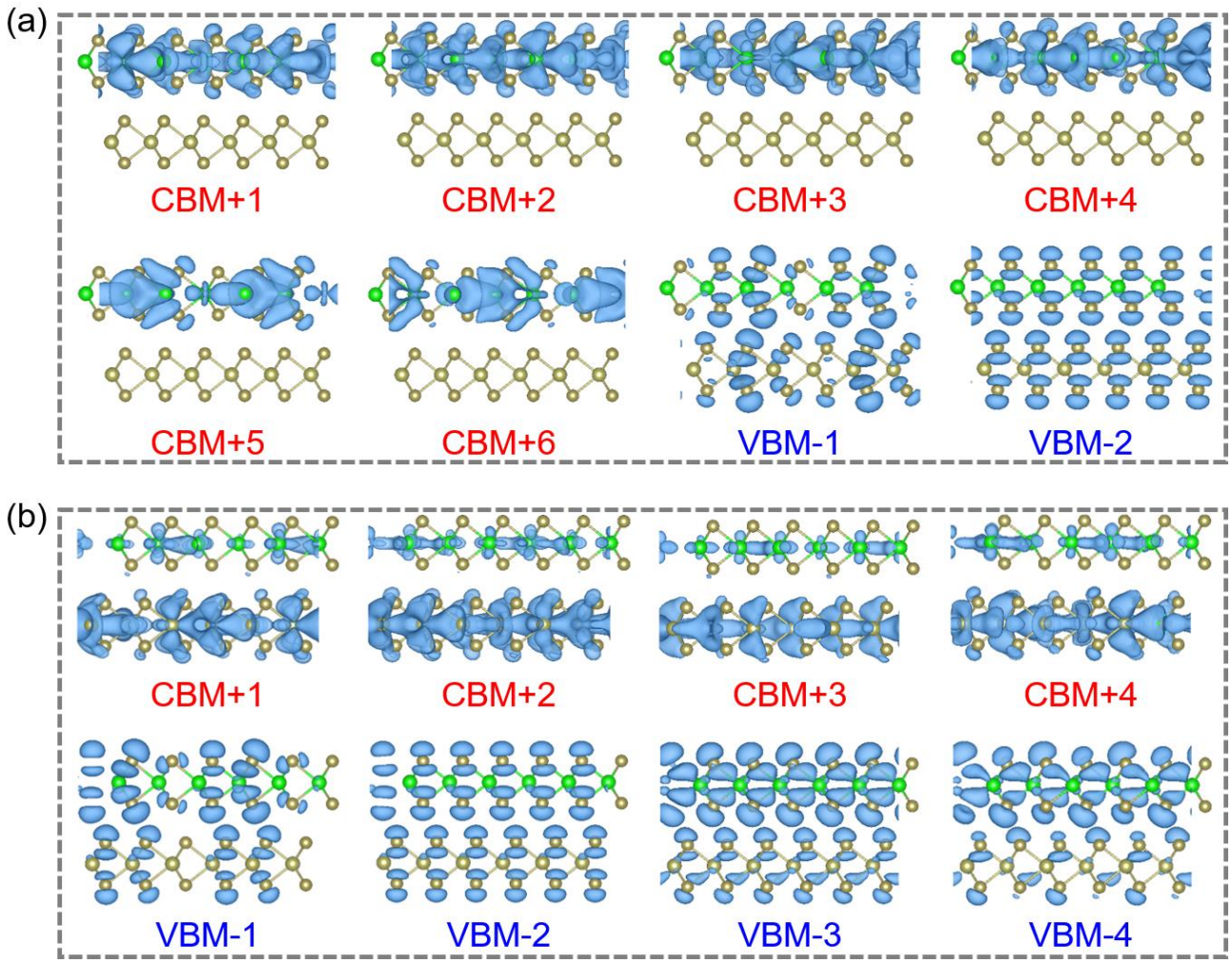

Figure S5. **Intermediate electronic states associated with electron localization in $ZrTe_2$/$HfTe_2$.** Spatial distributions of selected intermediate electronic states are shown for **a**) AB-$ZrTe_2$/$HfTe_2$ and **b**) BA-$ZrTe_2$/$HfTe_2$. In both stacking configurations, the upper and lower layers correspond to $ZrTe_2$ and $HfTe_2$, respectively. Red labels denote conduction-band states referenced to the conduction-band minimum (CBM), whereas blue labels denote valence-band states referenced to the valence-band maximum (VBM). The blue isosurfaces represent the spatial distributions of the selected electronic states.

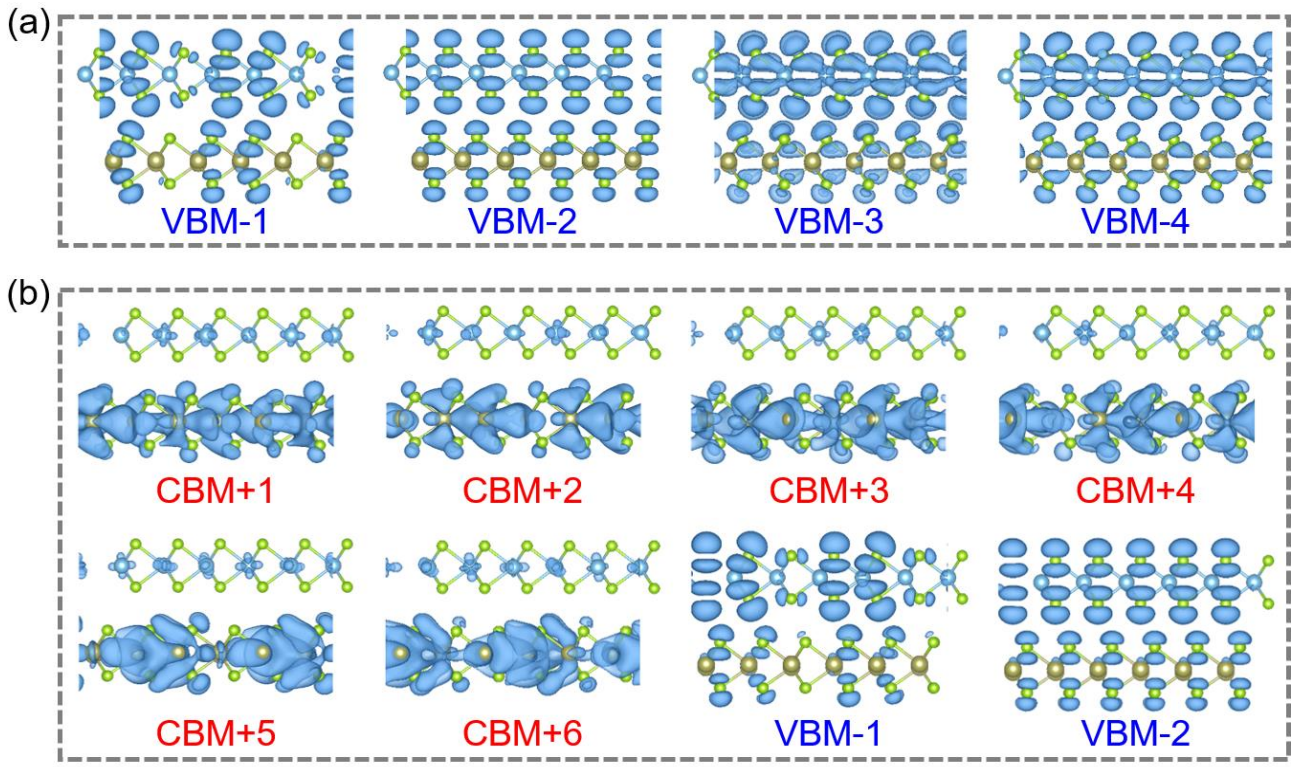

Figure S6. **Intermediate electronic states associated with electron localization in $TiSe_2/HfSe_2$.** Spatial distributions of selected intermediate electronic states are shown for **a**) AB-$TiSe_2/HfSe_2$ and **b**) BA-$TiSe_2/HfSe_2$. In both stacking configurations, the upper and lower layers correspond to $TiSe_2$ and $HfSe_2$, respectively. Red labels denote conduction-band states referenced to the conduction-band minimum (CBM), whereas blue labels denote valence-band states referenced to the valence-band maximum (VBM). The blue isosurfaces represent the spatial distributions of the selected electronic states.

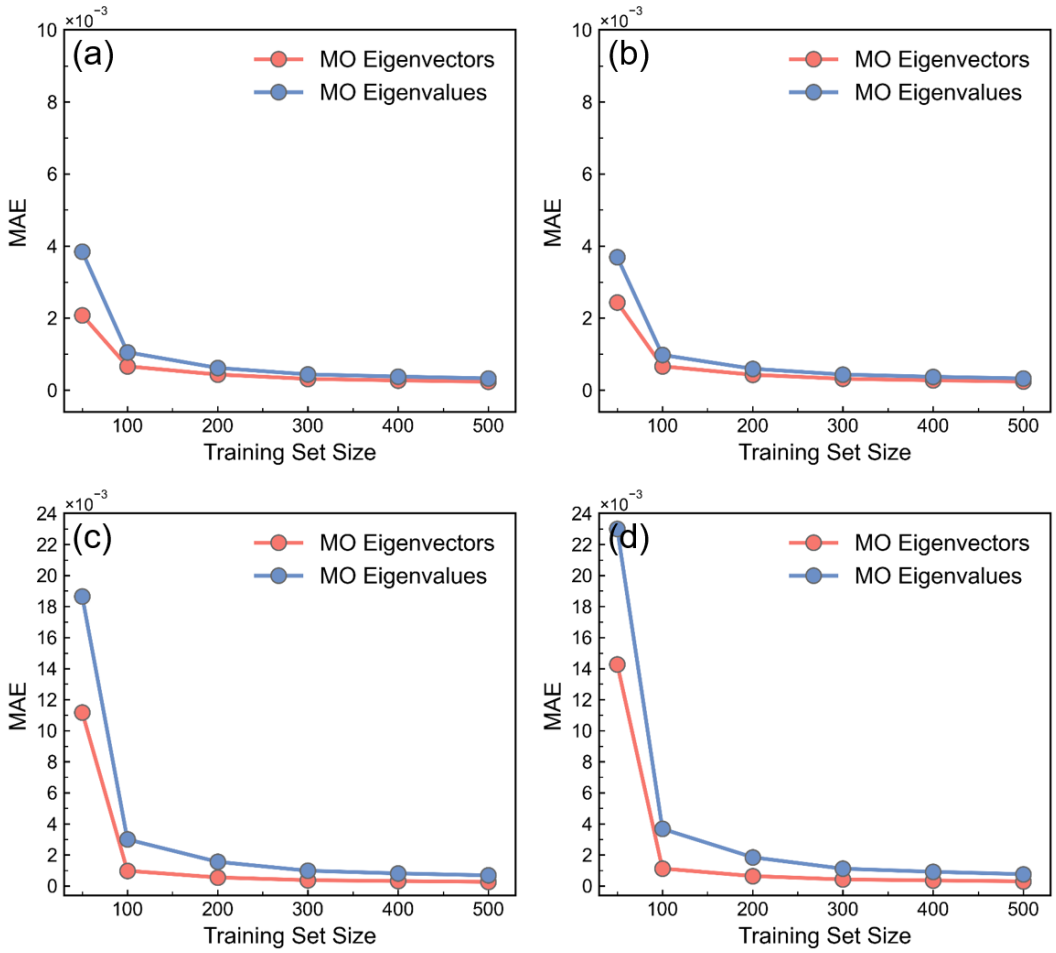

Figure S7. **Dependence of molecular orbital prediction errors on the KRR training set size.** Mean absolute errors (MAEs) of molecular orbital (MO) eigenvectors and eigenvalues predicted using kernel ridge regression (KRR) models with different training set sizes are shown for **a**) AB-$ZrTe_2$/$HfTe_2$, **b**) BA-$ZrTe_2$/$HfTe_2$, **c**) AB-$TiSe_2$/$HfSe_2$, and **d**) BA-$TiSe_2$/$HfSe_2$. Coral red circles and lines represent the MAEs of the MO eigenvectors, whereas cornflower blue circles and lines represent the MAEs of the MO eigenvalues. The horizontal axis denotes the number of structures used for model training, and the vertical axis shows the corresponding MAE values, with the indicated scaling factor of $\times 10^{-3}$.

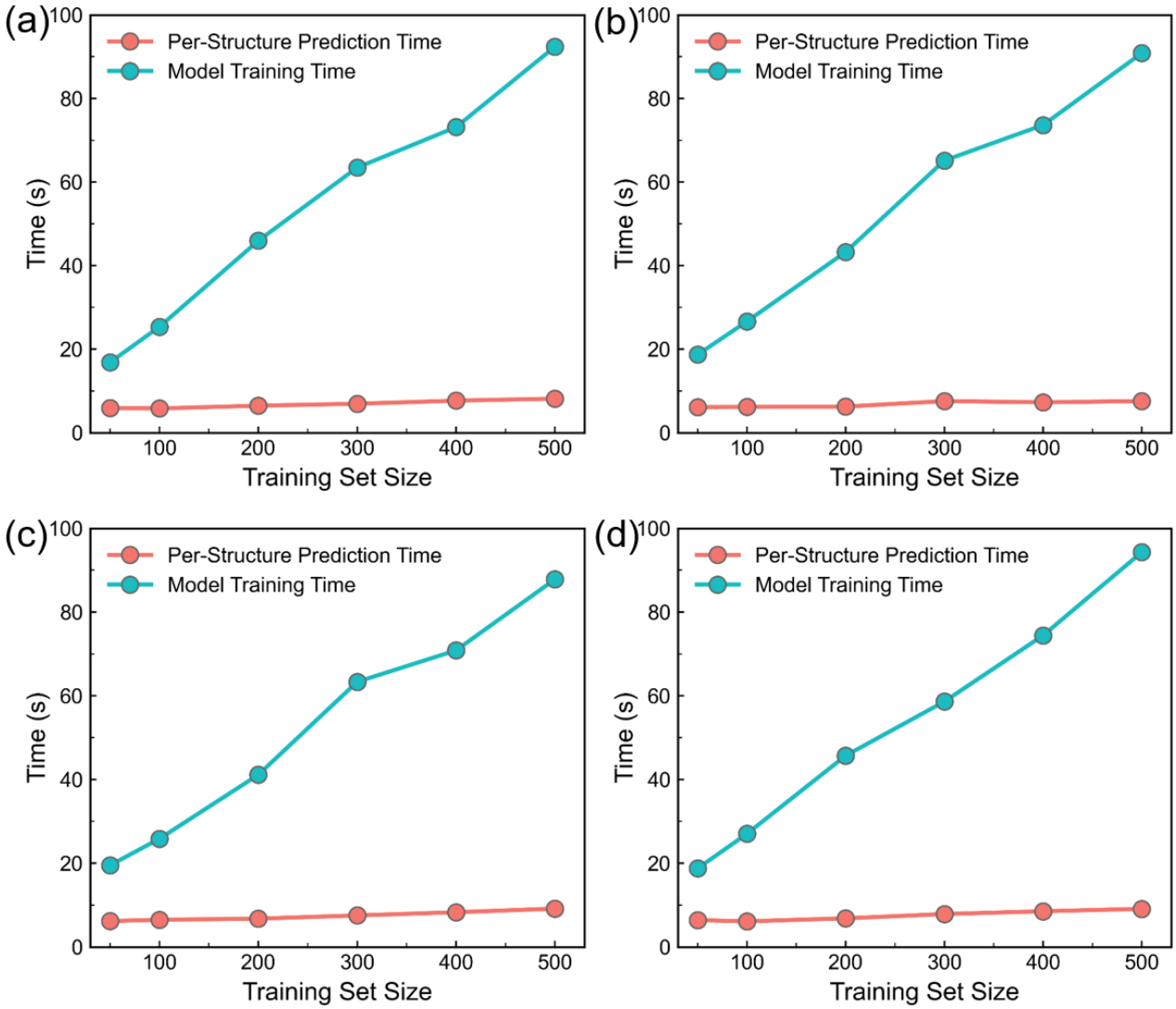

Figure S8. **Computational cost of KRR model training and Hamiltonian prediction as a function of training set size.** Model training times and per-structure Hamiltonian prediction times for kernel ridge regression (KRR) models with different training set sizes are shown for **a**) AB-$ZrTe_2$/$HfTe_2$, **b**) BA-$ZrTe_2$/$HfTe_2$, **c**) AB-$TiSe_2$/$HfSe_2$, and **d**) BA-$TiSe_2$/$HfSe_2$. Teal circles and lines represent the total KRR model training time, whereas coral red circles and lines represent the prediction time required to obtain the Hamiltonian for a single structure. The horizontal axis denotes the number of structures used for model training, and the vertical axis gives the computational time in seconds (s).

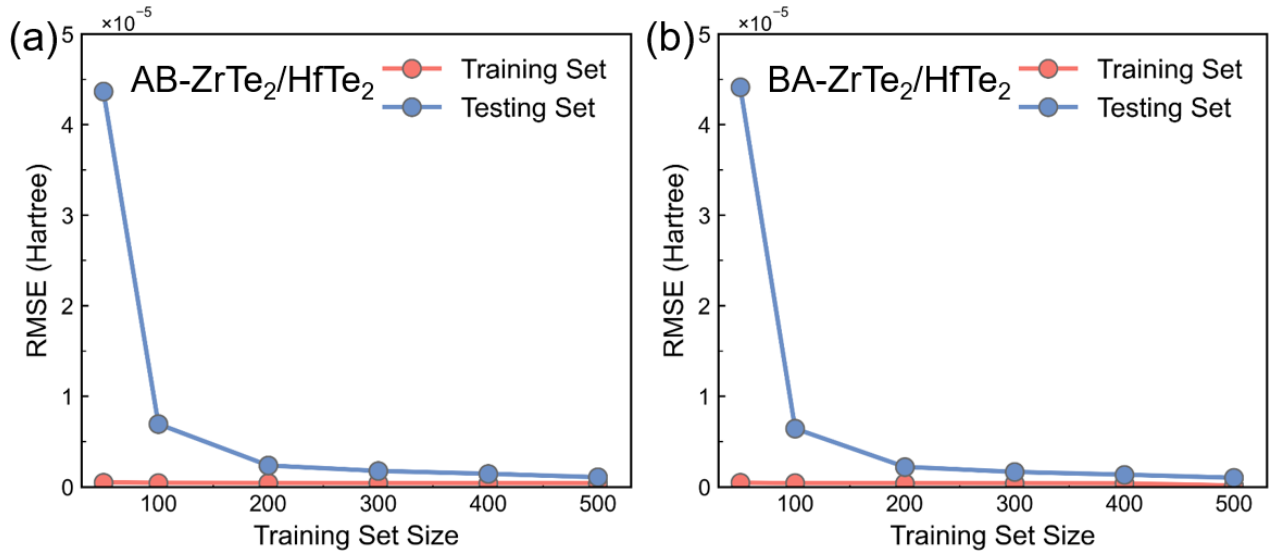

Figure S9. **Training set size dependence of the prediction errors of the machine learned Kohn-Sham Hamiltonians.** Root mean square errors (RMSEs) of the machine learning (ML)-predicted Kohn-Sham (KS) Hamiltonian matrix elements are shown as a function of training set size for **a**) AB-$ZrTe_2$/$HfTe_2$ and **b**) BA-$ZrTe_2$/$HfTe_2$. Coral red circles and lines represent the RMSEs for the training set, whereas cornflower blue circles and lines represent those for the testing set. The horizontal axis denotes the number of structures used for model training, and the vertical axis gives the corresponding RMSE values with the indicated scaling factor of $\times 10^{-5}$.

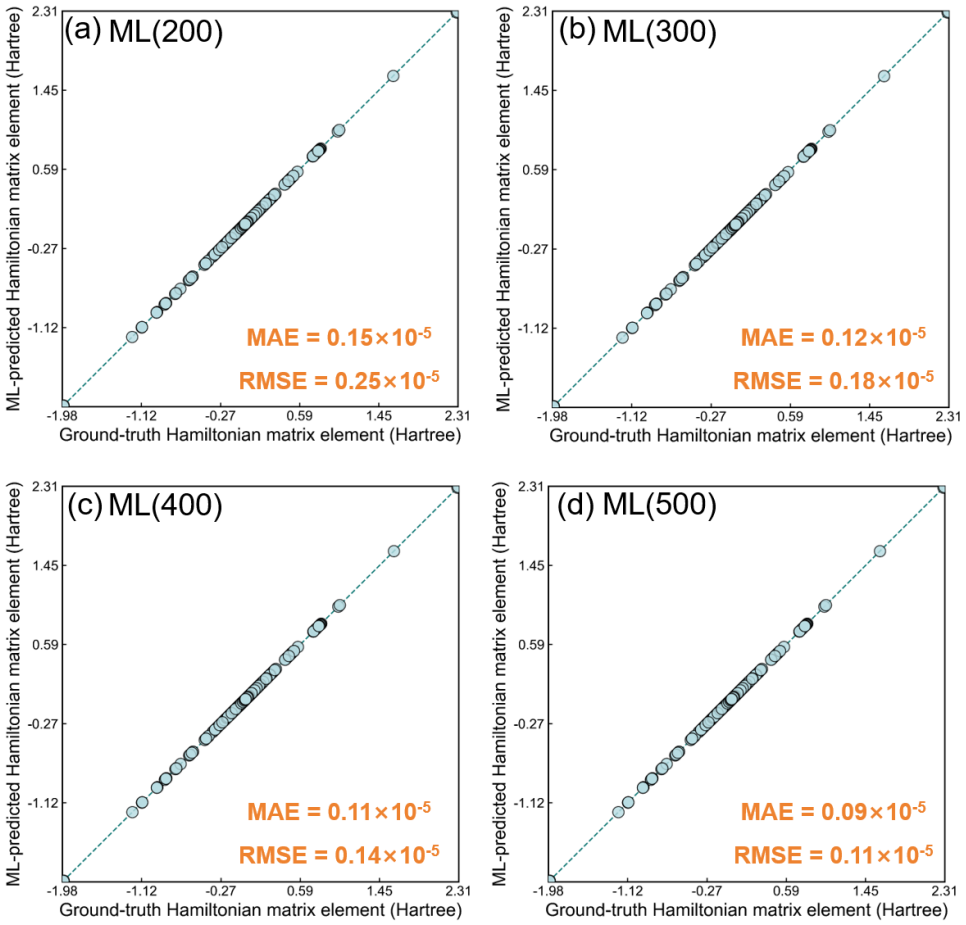

Figure S10. **Agreement between machine-learning-predicted and HSE06 reference Kohn-Sham Hamiltonian matrix elements for AB-$ZrTe_2$/$HfTe_2$.** Parity plots compare the machine learning (ML)-predicted and HSE06 reference Kohn-Sham (KS) Hamiltonian matrix elements obtained using kernel ridge regression (KRR) models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures. Circular markers represent individual Hamiltonian matrix elements, with the HSE06 reference values shown on the horizontal axis and the corresponding ML-predicted values on the vertical axis; both axes are given in Hartree. The cyan dashed diagonal line represents perfect agreement between prediction and reference values, corresponding to $y = x$. The mean absolute error (MAE) and root mean square error (RMSE) are reported in each panel.

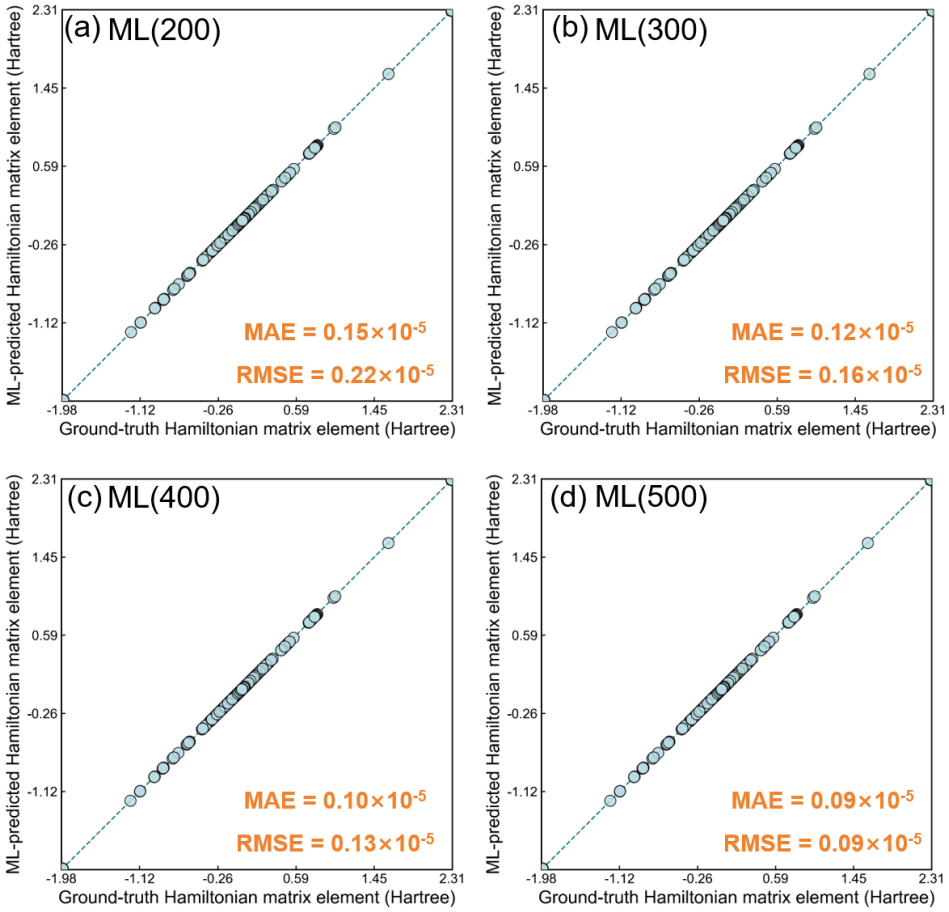

Figure S11. **Agreement between machine-learning-predicted and HSE06 reference Kohn-Sham Hamiltonian matrix elements for BA-$ZrTe_2$/$HfTe_2$.** Parity plots compare the machine learning (ML)-predicted and HSE06 reference Kohn-Sham (KS) Hamiltonian matrix elements obtained using kernel ridge regression (KRR) models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures. Circular markers represent individual Hamiltonian matrix elements, with the HSE06 reference values shown on the horizontal axis and the corresponding ML-predicted values on the vertical axis; both axes are given in Hartree. The cyan dashed diagonal line represents perfect agreement between the predicted and reference values, corresponding to $y = x$. The mean absolute error (MAE) and root mean square error (RMSE) are reported in orange text in each panel.

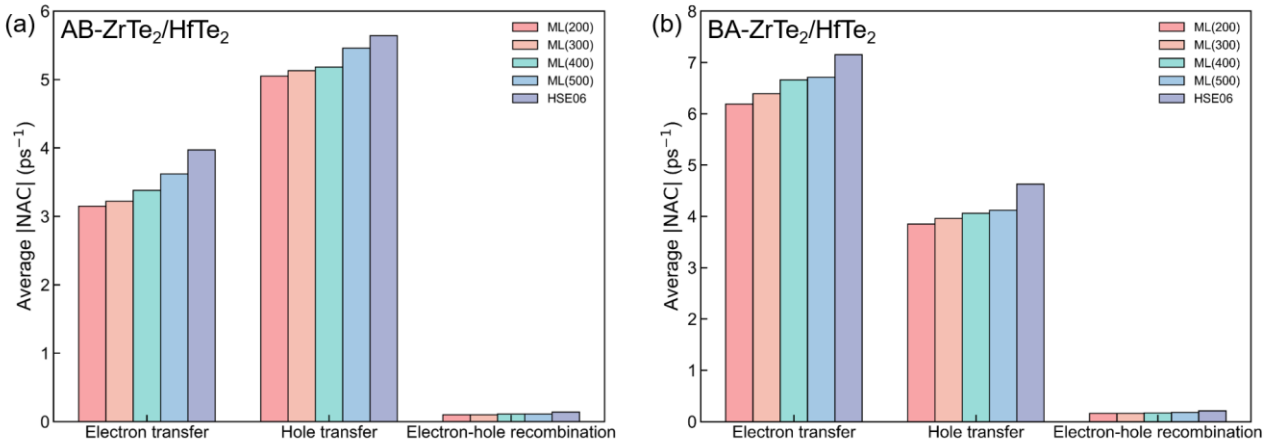

Figure S12. **Benchmark of machine learning predicted nonadiabatic couplings against explicit HSE06 calculations.** Average absolute nonadiabatic couplings (NACs) evaluated over 100 consecutive time steps are compared for **a**) AB-$ZrTe_2$/$HfTe_2$ and **b**) BA-$ZrTe_2$/$HfTe_2$. For each stacking configuration, the NACs associated with electron transfer, hole transfer, and electron-hole recombination are shown. Pink, light-orange, cyan, light-blue, and lavender bars represent the results obtained from kernel ridge regression (KRR) models trained with 200, 300, 400, and 500 structures, denoted as ML(200), ML(300), ML(400), and ML(500), and from explicit HSE06 calculations, respectively. ML denotes machine learning. The average absolute NAC values are reported in ps^{-1}.

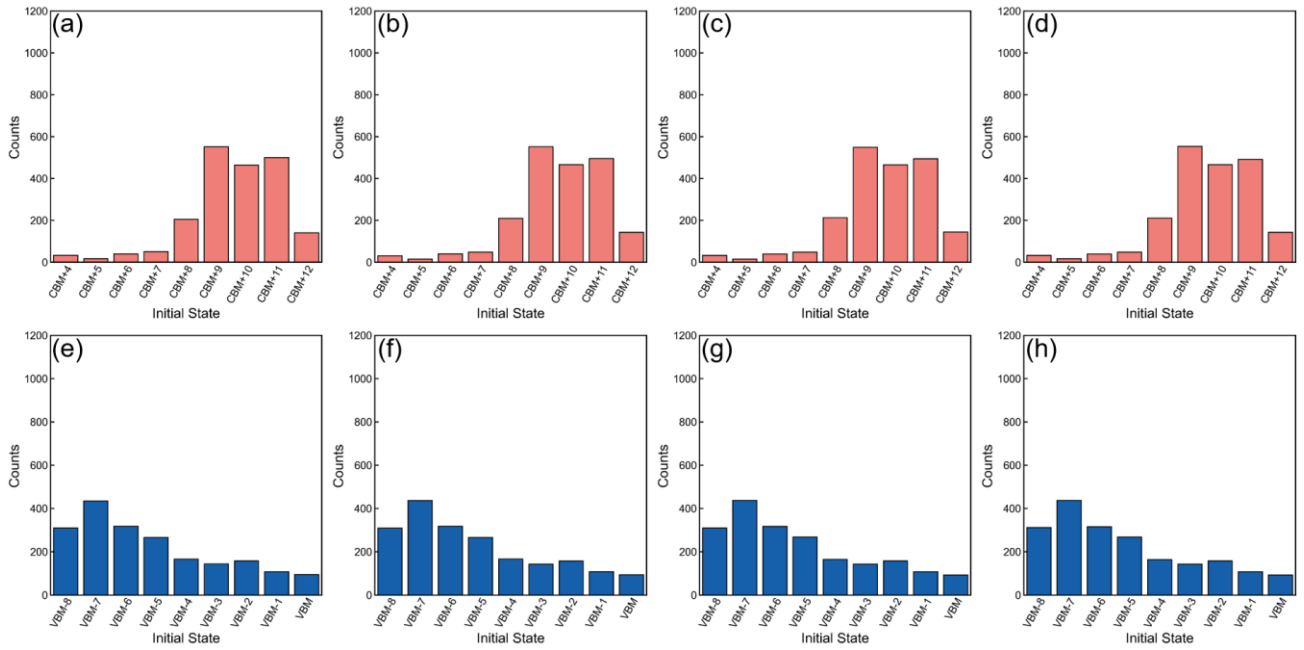

Figure S13. **Training set size dependence of the initial state distributions for carrier transfer in AB-$ZrTe_2$/$HfTe_2$.** Distributions of the electronic states selected as the initial states for electron and hole transfer are shown using models trained with different numbers of structures. Panels **a-d**) show the initial state distributions for electron transfer using training-set sizes of 200, 300, 400, and 500 structures, respectively, whereas panels **e-h**) show the corresponding distributions for hole transfer using the same training set sizes. Red bars represent the occurrence counts of the selected initial states for electron transfer, whereas blue bars represent those for hole transfer. The horizontal axis denotes the selected initial electronic state, and the vertical axis gives its occurrence count.

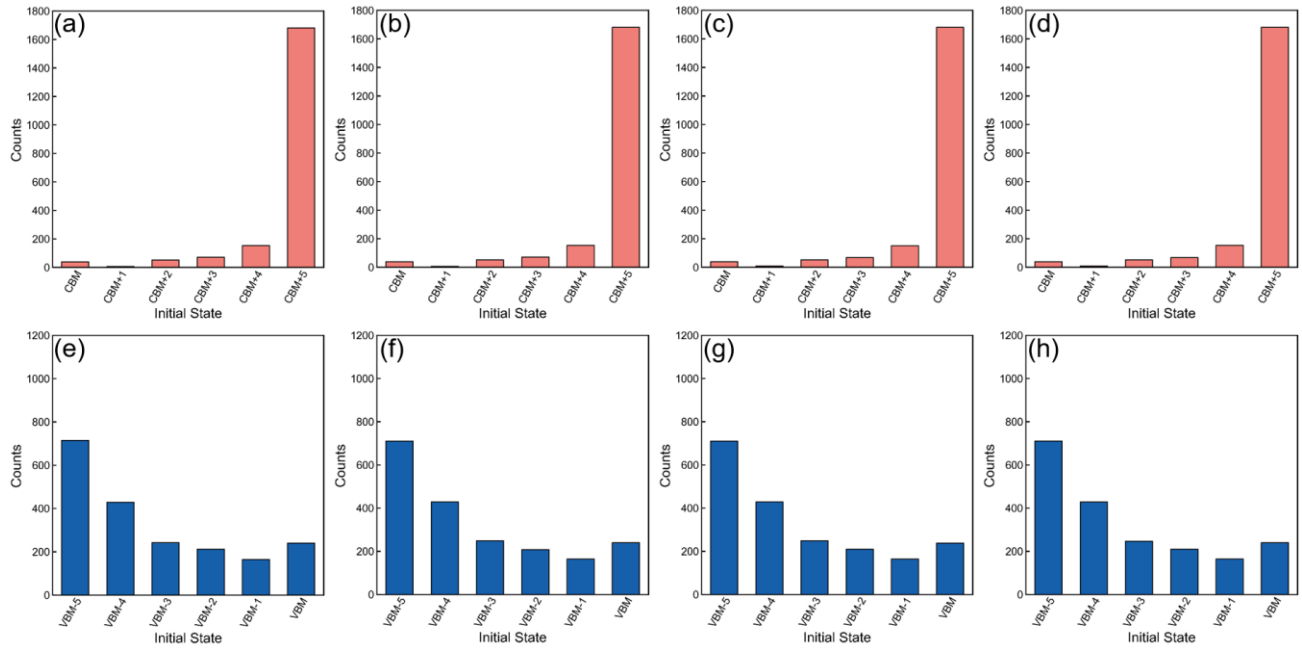

Figure S14. **Training set size dependence of the initial state distributions for carrier transfer in BA-$ZrTe_2$/$HfTe_2$.** Distributions of the electronic states selected as the initial states for electron and hole transfer are shown using models trained with different numbers of structures. Panels **a-d**) show the initial state distributions for electron transfer using training-set sizes of 200, 300, 400, and 500 structures, respectively, whereas panels **e-h**) show the corresponding distributions for hole transfer using the same training set sizes. Red bars represent the occurrence counts of the selected initial states for electron transfer, whereas blue bars represent those for hole transfer. The horizontal axis denotes the selected initial electronic state, and the vertical axis gives its occurrence count.

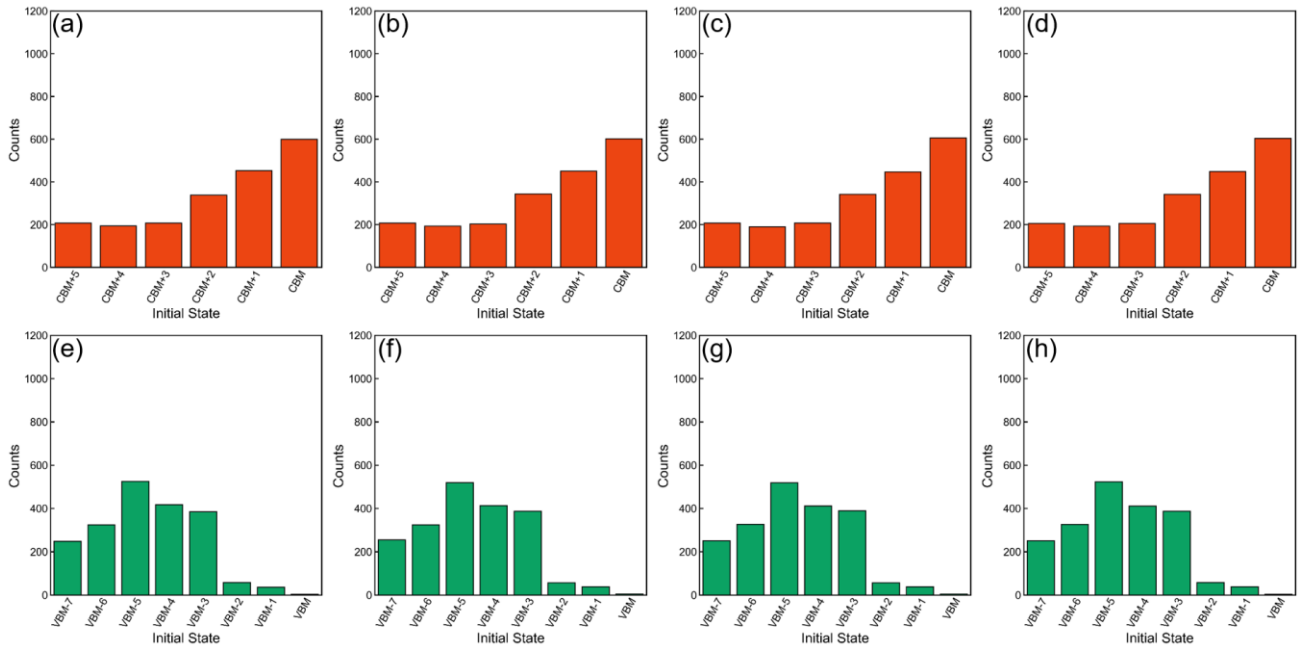

Figure S15. **Training set size dependence of the initial state distributions for carrier transfer in AB-$TiSe_2$/$HfSe_2$.** Distributions of the electronic states selected as the initial states for electron and hole transfer are shown using models trained with different numbers of structures. Panels **a-d**) show the initial state distributions for electron transfer using training set sizes of 200, 300, 400, and 500 structures, respectively, whereas panels **e-h**) show the corresponding distributions for hole transfer using the same training set sizes. Orange red bars represent the occurrence counts of the selected initial states for electron transfer, whereas emerald green bars represent those for hole transfer. The horizontal axis denotes the selected initial electronic state, and the vertical axis gives its occurrence count.

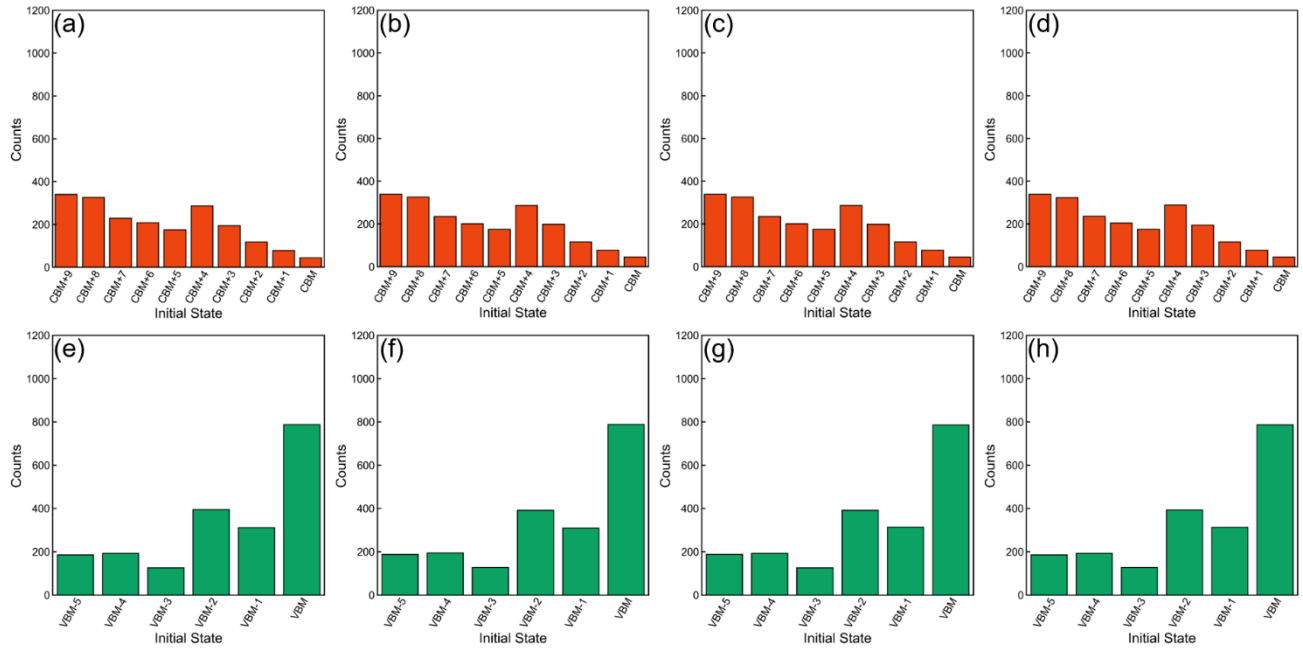

Figure S16. **Training set size dependence of the initial state distributions for carrier transfer in BA-$TiSe_2$/$HfSe_2$.** Distributions of the electronic states selected as the initial states for electron and hole transfer are shown using models trained with different numbers of structures. Panels **a-d**) show the initial state distributions for electron transfer using training set sizes of 200, 300, 400, and 500 structures, respectively, whereas panels **e-h**) show the corresponding distributions for hole transfer using the same training set sizes. Orange red bars represent the occurrence counts of the selected initial states for electron transfer, whereas emerald green bars represent those for hole transfer. The horizontal axis denotes the selected initial electronic state, and the vertical axis gives its occurrence count.

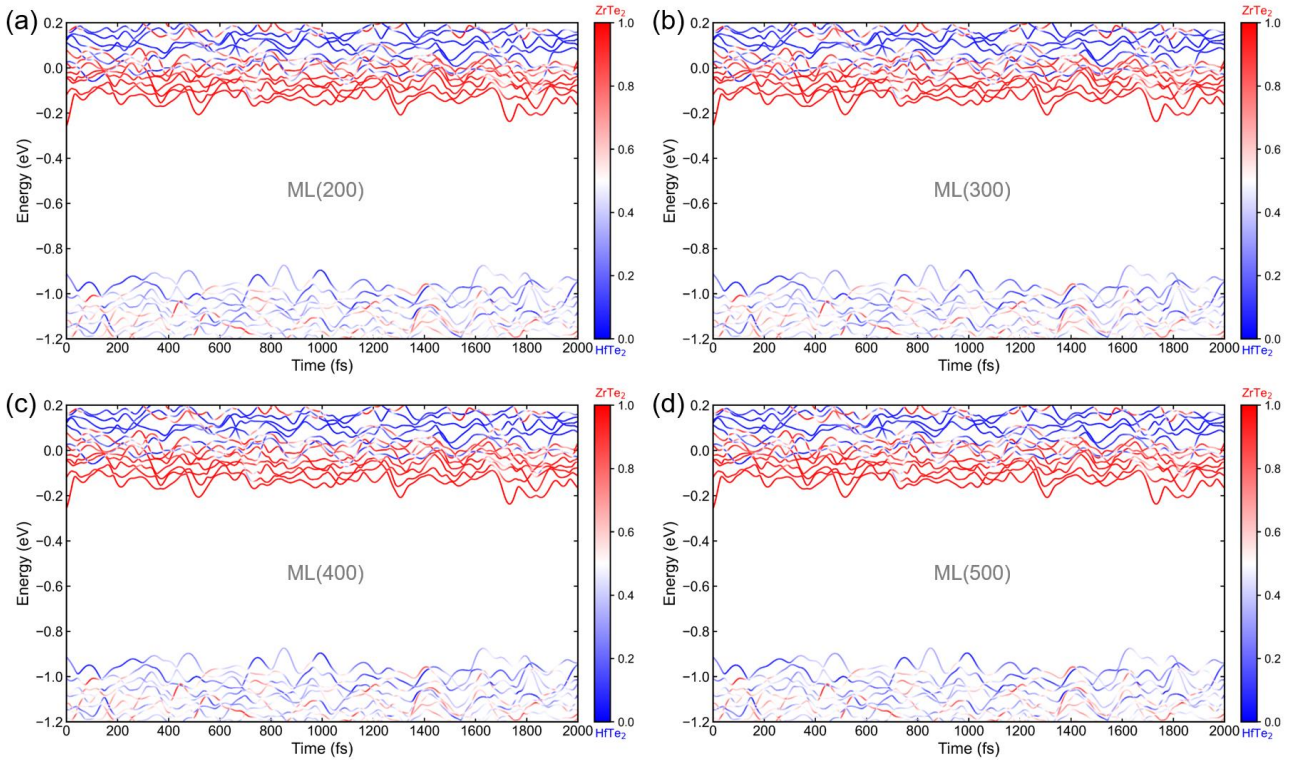

Figure S17. **Time evolution of Kohn-Sham orbital energies along the NVE trajectories of AB-$ZrTe_2$/$HfTe_2$.** Time-dependent Kohn-Sham (KS) orbital energies within the energy window from −1.2 to 0.2 eV are shown for trajectories at 300 K using machine learning models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures, denoted as ML(200), ML(300), ML(400), and ML(500), respectively. The trajectories are propagated in the microcanonical (NVE) ensemble. Red curves indicate electronic states predominantly localized on the $ZrTe_2$ layer, blue curves indicate states predominantly localized on the $HfTe_2$ layer, and pale or white regions indicate states with mixed contributions from both layers. The horizontal axis gives the simulation time in fs, and the vertical axis gives the KS orbital energy in eV.

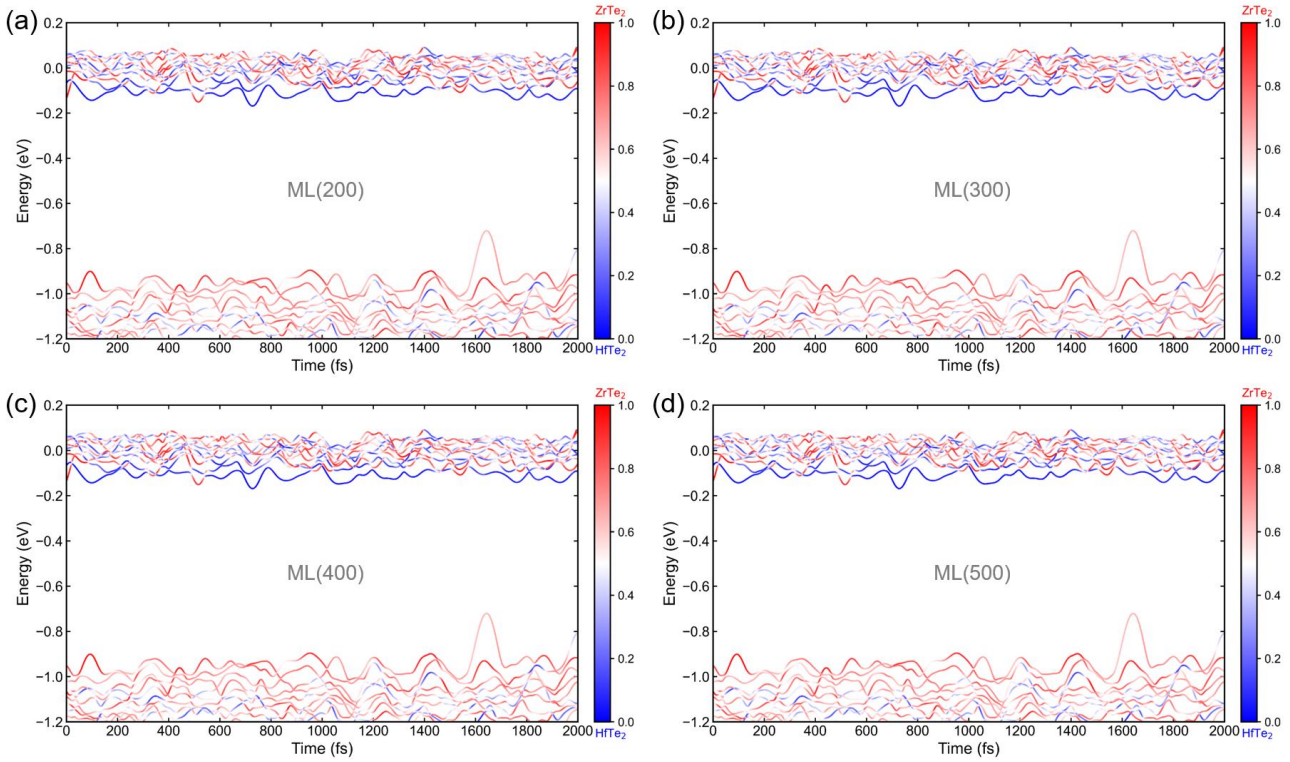

Figure S18. **Time evolution of Kohn-Sham orbital energies along the NVE trajectories of BA-$ZrTe_2$/$HfTe_2$.** Time-dependent Kohn-Sham (KS) orbital energies within the energy window from −1.2 to 0.2 eV are shown for trajectories at 300 K using machine learning models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures, denoted as ML(200), ML(300), ML(400), and ML(500), respectively. The trajectories are propagated in the microcanonical (NVE) ensemble. Red curves indicate electronic states predominantly localized on the $ZrTe_2$ layer, blue curves indicate states predominantly localized on the $HfTe_2$ layer, and pale or white curves indicate states with mixed contributions from both layers. The horizontal axis gives the simulation time in fs, and the vertical axis gives the KS orbital energy in eV.

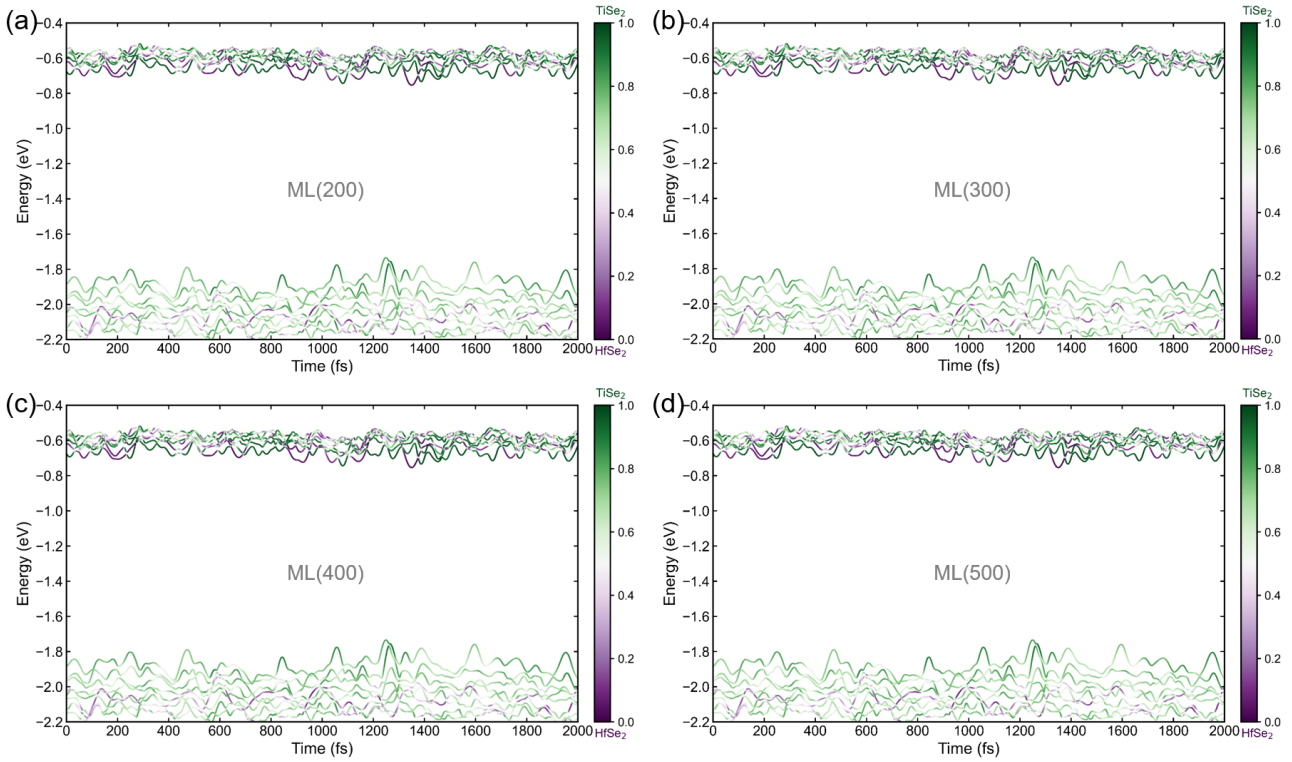

Figure S19. **Time evolution of Kohn-Sham orbital energies along the NVE trajectories of AB-$TiSe_2$/$HfSe_2$.** Time-dependent Kohn-Sham (KS) orbital energies within the energy window from −2.2 to −0.4 eV are shown for trajectories at 300 K using machine learning models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures, denoted as ML(200), ML(300), ML(400), and ML(500), respectively. The trajectories are propagated in the microcanonical (NVE) ensemble. Green curves indicate electronic states predominantly localized on the $TiSe_2$ layer, purple curves indicate states predominantly localized on the $HfSe_2$ layer, and pale or white curves indicate states with mixed contributions from both layers. The horizontal axis gives the simulation time in fs, and the vertical axis gives the KS orbital energy in eV.

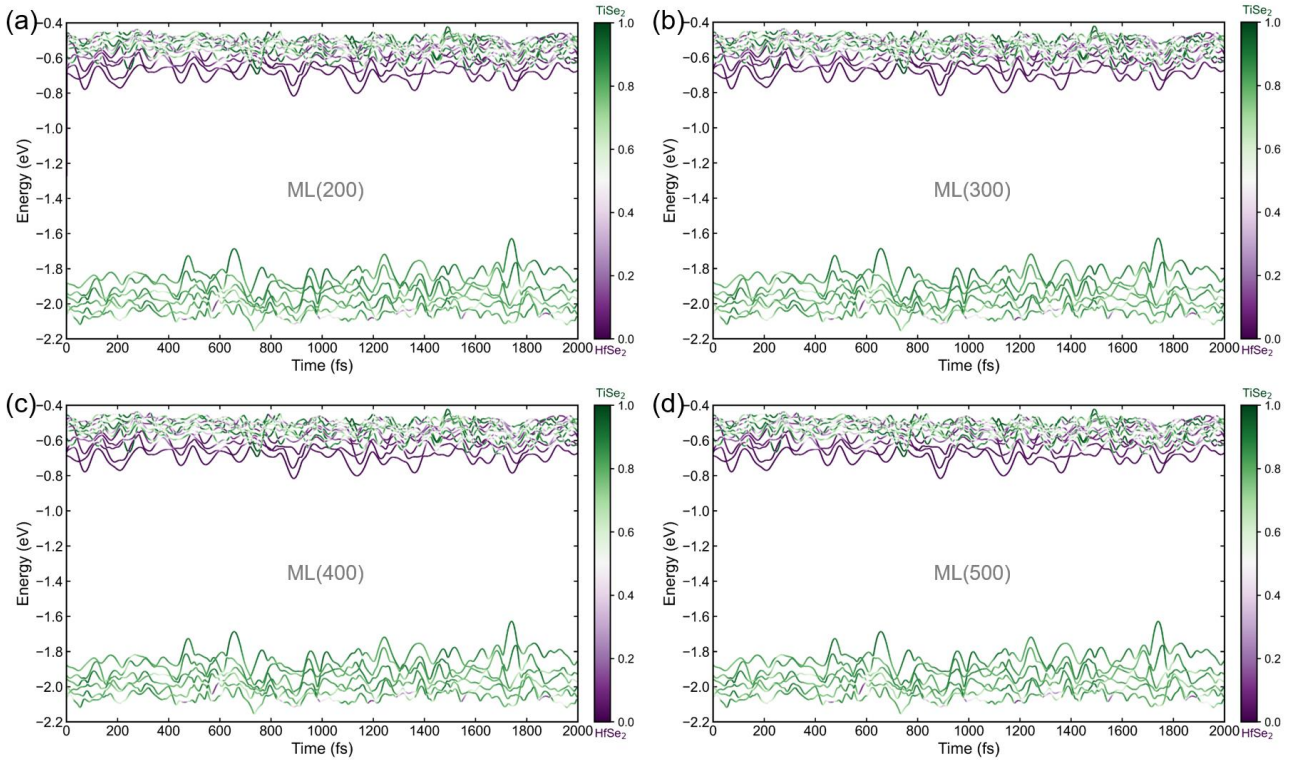

Figure S20. **Time evolution of Kohn-Sham orbital energies along the NVE trajectories of BA-$TiSe_2$/$HfSe_2$.** Time-dependent Kohn-Sham (KS) orbital energies within the energy window from −2.2 to −0.4 eV are shown for trajectories at 300 K using machine learning models trained with **a**) 200, **b**) 300, **c**) 400, and **d**) 500 structures, denoted as ML(200), ML(300), ML(400), and ML(500), respectively. The trajectories are propagated in the microcanonical (NVE) ensemble. Green curves indicate electronic states predominantly localized on the $TiSe_2$ layer, purple curves indicate states predominantly localized on the $HfSe_2$ layer, and pale or white curves indicate states with mixed contributions from both layers. The horizontal axis gives the simulation time in fs, and the vertical axis gives the KS orbital energy in eV.

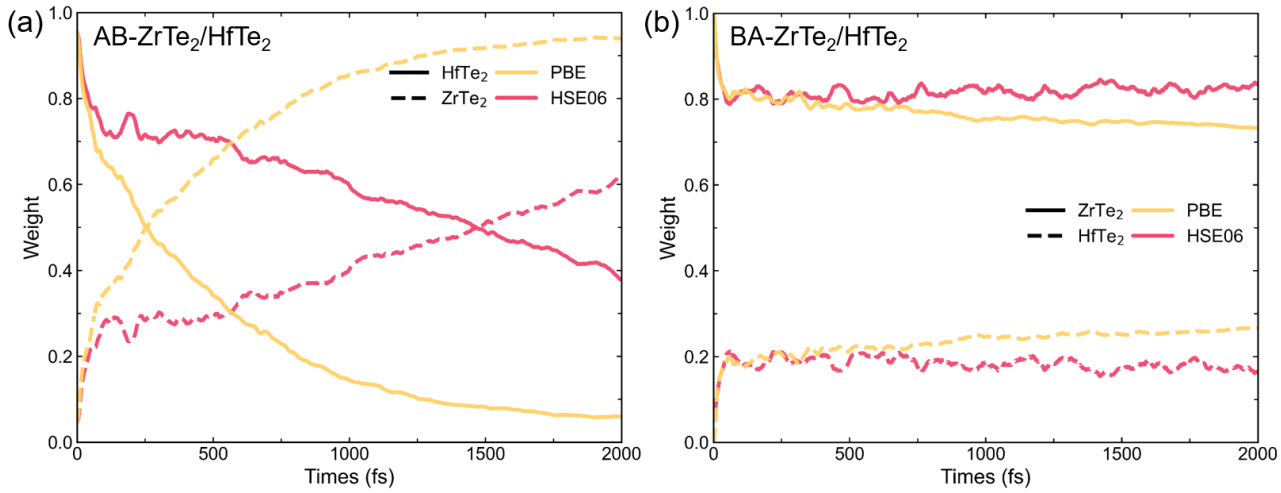

Figure S21. **Functional dependence of interlayer electron transfer dynamics in $ZrTe_2/HfTe_2$.** Time-dependent layer-resolved electron weights are compared at the Perdew-Burke-Ernzerhof (PBE) and Heyd-Scuseria-Ernzerhof hybrid (HSE06) levels for **a**) AB-$ZrTe_2/HfTe_2$ and **b**) BA-$ZrTe_2/HfTe_2$. Yellow curves represent the PBE results, whereas red curves represent the HSE06 results. In panel **a**), solid and dashed lines denote the electron weights on the $HfTe_2$ and $ZrTe_2$ layers, respectively; in panel **b**), solid and dashed lines denote the electron weights on the $ZrTe_2$ and $HfTe_2$ layers, respectively. The horizontal axis gives time in fs, and the vertical axis gives the corresponding layer-resolved electron weight.

.

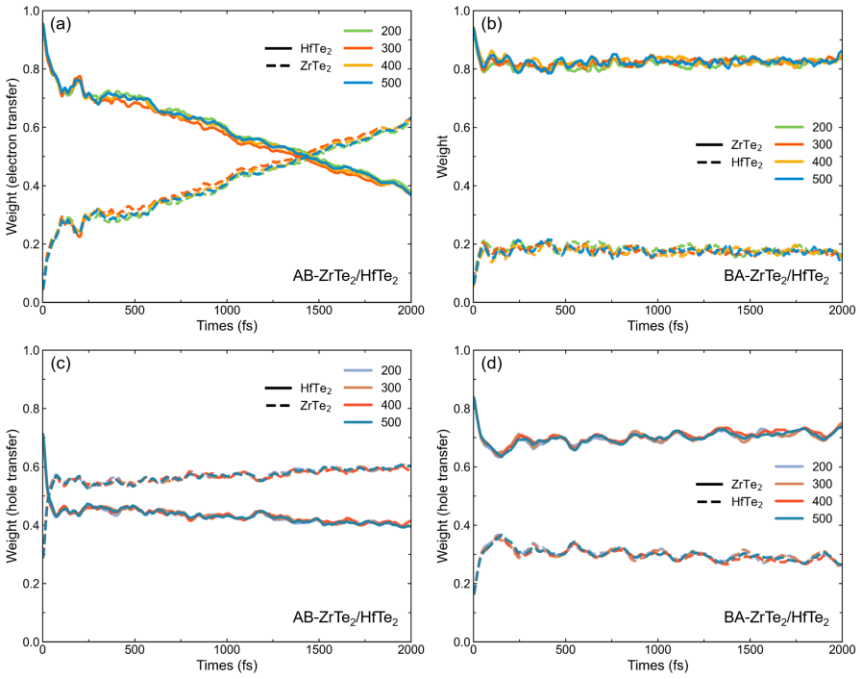

Figure S22. **Convergence of interlayer carrier transfer dynamics in $ZrTe_2/HfTe_2$ with respect to the KRR training set size.** Time-dependent layer-resolved carrier weights obtained using kernel ridge regression (KRR) models with different training set sizes are shown for **a**) electron transfer in AB-$ZrTe_2/HfTe_2$, **b**) electron transfer in BA-$ZrTe_2/HfTe_2$, **c**) hole transfer in AB-$ZrTe_2/HfTe_2$, and **d**) hole transfer in BA-$ZrTe_2/HfTe_2$. In panels **a**) and **b**), the differently colored curves correspond to KRR models trained with 200, 300, 400, and 500 structures, as indicated in the legends; the same correspondence is shown in panels **c**) and **d**) for hole transfer. In panels **a**) and **c**), solid and dashed lines represent the carrier weights on the $HfTe_2$ and $ZrTe_2$ layers, respectively, whereas in panels **b**) and **d**), solid and dashed lines represent the carrier weights on the $ZrTe_2$ and $HfTe_2$ layers, respectively. The horizontal axis gives time in fs, and the vertical axis gives the corresponding electron or hole weight. The trajectories obtained with different training-set sizes nearly overlap, indicating that a training set of 200 structures is sufficient to yield converged carrier transfer dynamics.

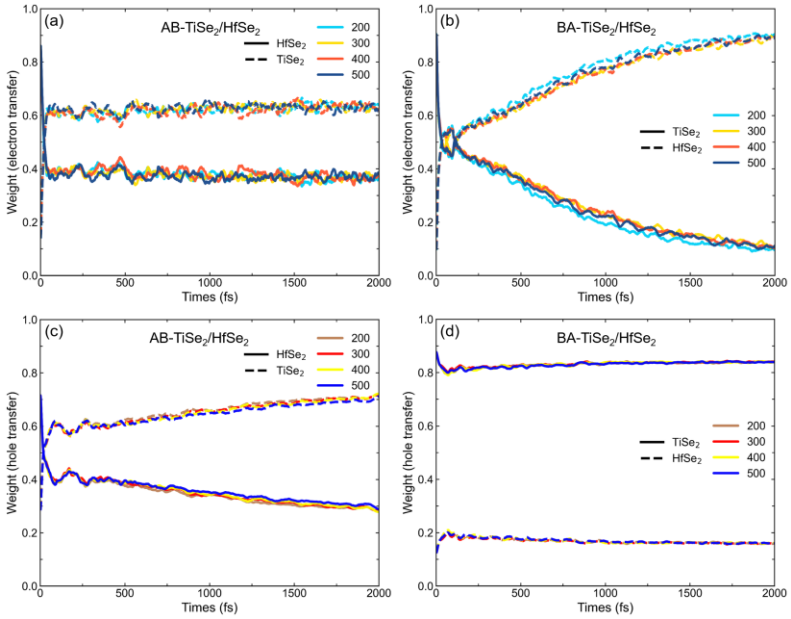

Figure S23. **Convergence of interlayer carrier transfer dynamics in $TiSe_2/HfSe_2$ with respect to the KRR training set size.** Time-dependent layer-resolved carrier weights obtained using kernel ridge regression (KRR) models with different training set sizes are shown for **a**) electron transfer in AB-$TiSe_2/HfSe_2$, **b**) electron transfer in BA-$TiSe_2/HfSe_2$, **c**) hole transfer in AB-$TiSe_2/HfSe_2$, and **d**) hole transfer in BA-$TiSe_2/HfSe_2$. In panels **a**) and **b**), cyan, yellow, orange-red, and blue curves correspond to KRR models trained with 200, 300, 400, and 500 structures, respectively. In panels **c**) and **d**), brown, red, yellow, and blue curves correspond to training-set sizes of 200, 300, 400, and 500 structures, respectively. In panels **a**) and **c**), solid and dashed lines represent the carrier weights on the $HfSe_2$ and $TiSe_2$ layers, respectively, whereas in panels **b**) and **d**), solid and dashed lines represent the carrier weights on the $TiSe_2$ and $HfSe_2$ layers, respectively. The horizontal axis gives time in fs, and the vertical axis gives the corresponding electron or hole weight. The trajectories obtained with different training set sizes nearly overlap, indicating that a training set of 200 structures is sufficient to yield converged carrier transfer dynamics.

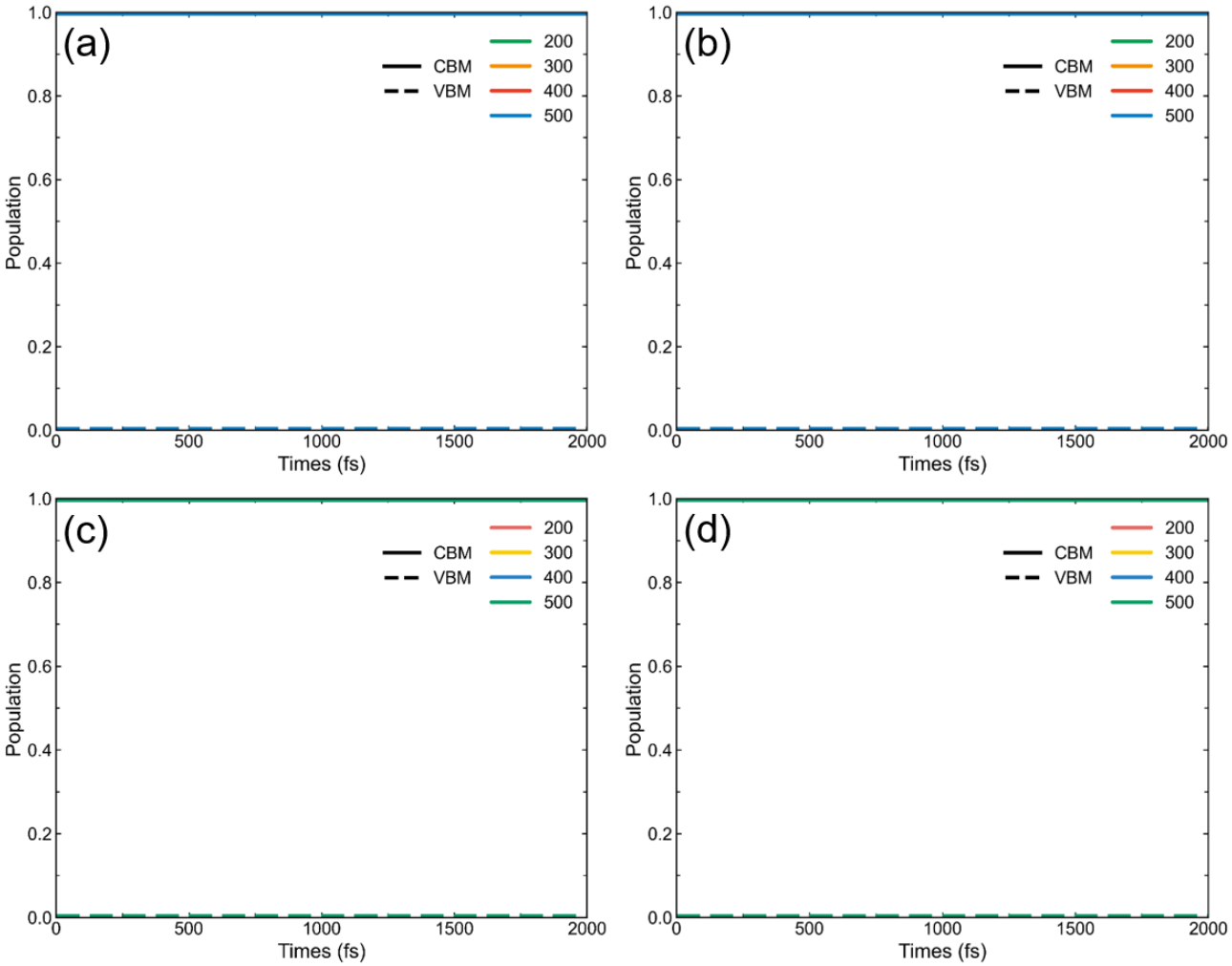

Figure S24. **Training set size dependence of electron hole recombination dynamics in the four heterostructure configurations.** Time-dependent electronic state populations are shown for **a**) AB-$ZrTe_2$/$HfTe_2$, **b**) BA-$ZrTe_2$/$HfTe_2$, **c**) AB-$TiSe_2$/$HfSe_2$, and **d**) BA-$TiSe_2$/$HfSe_2$ using models trained with 200, 300, 400, and 500 structures. Different colors correspond to the different training set sizes, as indicated in the legends. Solid lines represent the population of the conduction-band minimum (CBM), whereas dashed lines represent the population of the valence-band maximum (VBM). The horizontal axis gives time in fs, and the vertical axis gives the corresponding electronic state population.

Table S1. Mean absolute errors (MAEs) obtained using KRR models for the AB-$ZrTe_2$/$HfTe_2$ structure at different training set sizes, including Hamiltonians, band gaps, eigenvalues, eigenvectors, and the overlap between MO_{ML} and MO_{ref}. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures. "---" indicates values omitted, identical to those in the preceding table.

Train Set Size	Test Set size	Train Set MAE (Ha)	Test Set MAE (Ha)	Band Gap MAE (meV)	Eigenvalue MAE (meV)	Eigenvectors MAE (Ha)	MO-MO Overlaps
50 (2.5%)	200 (10.0%)	$1.61e^{-6}$	$4.58e^{-6}$	$2.19e^{-2}$	$3.84e^{-3}$	$2.08e^{-3}$	$5.55e^{-2}$
100 (5.0%)	---	$1.57e^{-6}$	$2.22e^{-6}$	$6.53e^{-4}$	$1.05e^{-3}$	$6.61e^{-4}$	$9.74e^{-3}$
200 (10.0%)	---	$1.24e^{-6}$	$1.57e^{-6}$	$4.00e^{-4}$	$6.18e^{-4}$	$4.34e^{-4}$	$5.31e^{-3}$
300 (15.0%)	---	$1.07e^{-6}$	$1.23e^{-6}$	$2.96e^{-4}$	$4.38e^{-4}$	$3.13e^{-4}$	$3.07e^{-3}$
400 (20.0%)	---	$9.71e^{-7}$	$1.08e^{-6}$	$2.61e^{-4}$	$3.77e^{-4}$	$2.70e^{-4}$	$2.32e^{-3}$
500 (25.0%)	---	$9.01e^{-7}$	$9.77e^{-7}$	$2.21e^{-4}$	$3.27e^{-4}$	$2.36e^{-4}$	$2.01e^{-3}$

Table S2. Mean absolute errors (MAEs) obtained using KRR models for the BA-$ZrTe_2$/$HfTe_2$ structure at different training set sizes, including Hamiltonians, band gaps, eigenvalues, eigenvectors, and the overlap between MO_{ML} and MO_{ref}. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures. "---" indicates values omitted, identical to those in the preceding table.

Train Set Size	Test Set size	Train Set MAE (Ha)	Test Set MAE (Ha)	Band Gap MAE (meV)	Eigenvalue MAE (meV)	Eigenvectors MAE (Ha)	MO-MO Overlaps
50 (2.5%)	200 (10.0%)	$1.60e^{-6}$	$4.38e^{-6}$	$7.47e^{-3}$	$3.69e^{-3}$	$2.44e^{-3}$	$9.93e^{-2}$
100 (5.0%)	---	$1.56e^{-6}$	$2.13e^{-6}$	$3.51e^{-3}$	$9.77e^{-4}$	$6.62e^{-4}$	$1.15e^{-2}$
200 (10.0%)	---	$1.23e^{-6}$	$1.51e^{-6}$	$3.16e^{-4}$	$5.94e^{-4}$	$4.27e^{-4}$	$4.71e^{-3}$
300 (15.0%)	---	$1.06e^{-6}$	$1.18e^{-6}$	$2.58e^{-4}$	$4.34e^{-4}$	$3.16e^{-4}$	$3.18e^{-3}$
400 (20.0%)	---	$9.53e^{-7}$	$1.04e^{-6}$	$2.24e^{-4}$	$3.72e^{-4}$	$2.73e^{-4}$	$2.39e^{-3}$
500 (25.0%)	---	$8.79e^{-7}$	$9.37e^{-7}$	$1.98e^{-4}$	$3.25e^{-4}$	$2.39e^{-4}$	$2.11e^{-3}$

Table S3. Mean absolute errors (MAEs) obtained using KRR models for the AB-$TiSe_2$/$HfSe_2$ structure at different training set sizes, including Hamiltonians, band gaps, eigenvalues, eigenvectors, and the overlap between MO_{ML} and MO_{ref}. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures. "---" indicates values omitted, identical to those in the preceding table.

Train Set Size	Test Set size	Train Set MAE (Ha)	Test Set MAE (Ha)	Band Gap MAE (meV)	Eigenvalue MAE (meV)	Eigenvectors MAE (Ha)	MO-MO Overlaps
50 (2.5%)	200 (10.0%)	$1.17e^{-6}$	$1.34e^{-5}$	$8.64e^{-1}$	$1.87e^{-2}$	$1.12e^{-2}$	$5.41e^{-1}$
100 (5.0%)	---	$2.12e^{-6}$	$3.51e^{-6}$	$1.30e^{-2}$	$3.00e^{-3}$	$9.77e^{-4}$	$1.86e^{-2}$
200 (10.0%)	---	$1.69e^{-6}$	$2.39e^{-6}$	$6.77e^{-4}$	$1.55e^{-3}$	$5.49e^{-4}$	$5.63e^{-3}$
300 (15.0%)	---	$1.46e^{-6}$	$1.77e^{-6}$	$4.66e^{-4}$	$9.82e^{-4}$	$3.73e^{-4}$	$3.63e^{-3}$
400 (20.0%)	---	$1.31e^{-6}$	$1.54e^{-6}$	$3.78e^{-4}$	$8.08e^{-4}$	$3.16e^{-4}$	$2.60e^{-3}$
500 (25.0%)	---	$1.20e^{-6}$	$1.35e^{-6}$	$3.25e^{-4}$	$6.83e^{-4}$	$2.66e^{-4}$	$2.30e^{-3}$

Table S4. Mean absolute errors (MAEs) obtained using KRR models for the BA-$TiSe_2$/$HfSe_2$ structure at different training set sizes, including Hamiltonians, band gaps, eigenvalues, eigenvectors, and the overlap between MO_{ML} and MO_{ref}. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures. "---" indicates values omitted, identical to those in the preceding table.

Train Set Size	Test Set size	Train Set MAE (Ha)	Test Set MAE (Ha)	Band Gap MAE (meV)	Eigenvalue MAE (meV)	Eigenvectors MAE (Ha)	MO-MO Overlaps
50 (2.5%)	200 (10.0%)	$1.12e^{-6}$	$1.47e^{-5}$	$8.43e^{-1}$	$2.30e^{-2}$	$1.43e^{-2}$	$6.37e^{-1}$
100 (5.0%)	---	$2.13e^{-6}$	$3.78e^{-6}$	$7.41e^{-3}$	$3.68e^{-3}$	$1.12e^{-3}$	$2.08e^{-2}$
200 (10.0%)	---	$1.74e^{-6}$	$2.55e^{-6}$	$7.76e^{-4}$	$1.85e^{-3}$	$6.37e^{-4}$	$7.23e^{-3}$
300 (15.0%)	---	$1.50e^{-6}$	$1.87e^{-6}$	$5.52e^{-4}$	$1.11e^{-3}$	$4.20e^{-4}$	$4.20e^{-3}$
400 (20.0%)	---	$1.35e^{-6}$	$1.63e^{-6}$	$4.59e^{-4}$	$9.13e^{-4}$	$3.57e^{-4}$	$3.29e^{-3}$
500 (25.0%)	---	$1.24e^{-6}$	$1.42e^{-6}$	$4.06e^{-4}$	$7.57e^{-4}$	$2.98e^{-4}$	$2.60e^{-3}$

Table S5. Training time on a single CPU core for ML models mapping from PBE (atomic guess) to HSE06 (converged) for the AB-$ZrTe_2$/$HfTe_2$ structure. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures, and the number of submodels was consistently maintained at 1,065. "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Test Set size	Sub-model Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	200 (10.0%)	1065	16.83 s
---	100 (5.0%)	---	---	25.35 s
---	200 (10.0%)	---	---	45.94 s
---	300 (15.0%)	---	---	63.45 s
---	400 (20.0%)	---	---	73.13 s
---	500 (25.0%)	---	---	92.36 s

Table S6. For the AB-$ZrTe_2$/$HfTe_2$ heterostructure at different training set sizes, the mean time on a single CPU core to map the Hamiltonian from PBE (atomic guess) to HSE06 for a single structure using KRR models is reported. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	5.87 s
---	100 (5.0%)	5.80 s
---	200 (10.0%)	6.46 s
---	300 (15.0%)	6.92 s
---	400 (20.0%)	7.67 s
---	500 (25.0%)	8.14 s

Table S7. Training time on a single CPU core for ML models mapping from PBE (atomic guess) to HSE06 (converged) for the BA-$ZrTe_2$/$HfTe_2$ structure. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures, and the number of submodels was consistently maintained at 1,065. "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Test Set size	Sub-model Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	200 (10.0%)	1065	18.68 s
---	100 (5.0%)	---	---	26.64 s
---	200 (10.0%)	---	---	43.25 s
---	300 (15.0%)	---	---	65.13 s
---	400 (20.0%)	---	---	73.62 s
---	500 (25.0%)	---	---	90.83 s

Table S8. For the BA-$ZrTe_2$/$HfTe_2$ heterostructure at different training set sizes, the mean time on a single CPU core to map the Hamiltonian from PBE (atomic guess) to HSE06 for a single structure using KRR models is reported. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	6.11 s
---	100 (5.0%)	6.16 s
---	200 (10.0%)	6.24 s
---	300 (15.0%)	7.56 s
---	400 (20.0%)	7.27 s
---	500 (25.0%)	7.56 s

Table S9. Training time on a single CPU core for ML models mapping from PBE (atomic guess) to HSE06 (converged) for the AB-$TiSe_2$/$HfSe_2$ structure. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures, and the number of submodels was consistently maintained at 1,065. "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Test Set size	Sub-model Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	200 (10.0%)	1065	19.50 s
---	100 (5.0%)	---	---	25.84 s
---	200 (10.0%)	---	---	41.12 s
---	300 (15.0%)	---	---	63.31 s
---	400 (20.0%)	---	---	70.84 s
---	500 (25.0%)	---	---	87.80 s

Table S10. For the AB-$TiSe_2$/$HfSe_2$ heterostructure at different training set sizes, the mean time on a single CPU core to map the Hamiltonian from PBE (atomic guess) to HSE06 for a single structure using KRR models is reported. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	6.18 s
---	100 (5.0%)	6.49 s
---	200 (10.0%)	6.76 s
---	300 (15.0%)	7.52 s
---	400 (20.0%)	8.29 s
---	500 (25.0%)	9.14 s

Table S11. Training time on a single CPU core for ML models mapping from PBE (atomic guess) to HSE06 (converged) for the BA-$TiSe_2$/$HfSe_2$ structure. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). The test set was fixed at 200 structures, and the number of submodels was consistently maintained at 1,065. "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Test Set size	Sub-model Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	200 (10.0%)	1065	18.76 s
---	100 (5.0%)	---	---	27.04 s
---	200 (10.0%)	---	---	45.69 s
---	300 (15.0%)	---	---	58.62 s
---	400 (20.0%)	---	---	74.40 s
---	500 (25.0%)	---	---	94.29 s

Table S12. For the BA-$TiSe_2$/$HfSe_2$ heterostructure at different training set sizes, the mean time on a single CPU core to map the Hamiltonian from PBE (atomic guess) to HSE06 for a single structure using KRR models is reported. The training-set size was set to 50, 100, 200, 300, 400, and 500 snapshots (corresponding to 2.5%, 5.0%, 10.0%, 15.0%, 20.0%, and 25.0% of the 2 ps *NVE* trajectories simulation at 300 K). "---" indicates values omitted, identical to those in the preceding table.

CPU Model	Train Set Size	Elapsed Time
Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**1 cores**)	50 (2.5%)	6.40 s
---	100 (5.0%)	6.12 s
---	200 (10.0%)	6.83 s
---	300 (15.0%)	7.86 s
---	400 (20.0%)	8.52 s
---	500 (25.0%)	9.09 s

Table S13. Comparison of initial electron states in NAMD with dominant TDDFPT transitions. The bold numbers indicate the virtual states involved in the dominant TDDFPT transitions that exactly correspond to the CBM states selected as the initial electron states in the NAMD simulations.

	NAMD initial electron state	TDDFPT transition orbitals	TDDFPT oscillator strength (f)
AB-$ZrTe_2$/$HfTe_2$	$HfTe_2$ CBM = **440**	415→**440**; 410→**440**	0.1129; 0.4616
BA-$ZrTe_2$/$HfTe_2$	$ZrTe_2$ CBM = **438**	417→**438**; 416→**438**	1.6309; 0.1396
AB-$TiSe_2$/$HfSe_2$	$HfSe_2$ CBM = **434**	408→**434**; 384→**434**	1.1418; 0.4059
BA-$TiSe_2$/$HfSe_2$	$TiSe_2$ CBM = **440**	436→**440**; 434→**440**	0.1165; 0.6615

Table S14. Comparison of initial hole states in NAMD with dominant TDDFPT transitions. The bold numbers indicate the occupied states involved in the dominant TDDFPT transitions and their correspondence to the VBM states selected as the initial hole states in the NAMD simulations. For most systems, these occupied states either exactly match the selected VBM initial hole states or differ by only one adjacent valence state.

	NAMD initial hole state	TDDFPT occupied orbitals	TDDFPT oscillator strength (f)
AB-$ZrTe_2$/$HfTe_2$	$HfTe_2$ VBM = **429**	**428**→457; **428**→458	0.0227; 0.0123
BA-$ZrTe_2$/$HfTe_2$	$ZrTe_2$ VBM = **427**	**428**→447; **428**→457	0.0148; 0.0112
AB-$TiSe_2$/$HfSe_2$	$HfSe_2$ VBM = **427**	**428**→460; **428**→434	0.0156; 0.0137
BA-$TiSe_2$/$HfSe_2$	$TiSe_2$ VBM = **429**	**429**→433; **429**→458	0.0211; 0.0198

Table S15. RMSD values between the relaxed charge-transfer structures obtained from constrained density functional theory (CDFT) calculations and the corresponding ground-state structures for different stacking configurations of $ZrTe_2/HfTe_2$ and $TiSe_2/HfSe_2$ heterostructures. The "Electron Transfer (-1)" column corresponds to the constrained transfer of one electron from one layer to the other, while the "Hole Transfer (+1)" column corresponds to the constrained transfer of one hole.

		RMSD from Ground-State Structure (Å)
AB-$ZrTe_2/HfTe_2$	Electron Transfer (-1)	0.0069
	Hole Transfer (+1)	0.0370
BA-$ZrTe_2/HfTe_2$	Electron Transfer (-1)	0.0053
	Hole Transfer (+1)	0.0046
AB-$TiSe_2/HfSe_2$	Electron Transfer (-1)	0.0053
	Hole Transfer (+1)	0.0600
BA-$TiSe_2/HfSe_2$	Electron Transfer (-1)	0.0718
	Hole Transfer (+1)	0.0556

Table S16. Calculated reorganization energies (λ, meV) for electron and hole transfer in the four heterostructure configurations.

	electron transfer λ	hole transfer λ
AB-$TiSe_2$/$HfSe_2$	1.16	1.57
BA-$TiSe_2$/$HfSe_2$	1.51	1.20
AB-$ZrTe_2$/$HfTe_2$	1.05	1.18
BA-$ZrTe_2$/$HfTe_2$	1.36	0.74

Table S17. Based on 200 snapshots extracted from a 2 ps *NVE* trajectories simulation at 300 K of the AB-$ZrTe_2$/$HfTe_2$ structure, the mean computational time per structure was obtained for Hamiltonian evaluations by dividing the total wall-clock time by the number of snapshots. The timings were measured on 36 CPU cores at three levels: PBE (atomic guess), PBE (converged), and HSE06 (converged).

	PBE Atomic Guess	**PBE Converged**	**HSE06 Converged**
CPU Model	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)
Mean Computation Time	20.52 s	3 min 6.78 s (186.78 s)	21 min 12.50 s (1272.50 s)

Table S18. Based on 200 snapshots extracted from a 2 ps *NVE* trajectories simulation at 300 K of the BA-$ZrTe_2$/$HfTe_2$ structure, the mean computational time per structure was obtained for Hamiltonian evaluations by dividing the total wall-clock time by the number of snapshots. The timings were measured on 36 CPU cores at three levels: PBE (atomic guess), PBE (converged), and HSE06 (converged).

	PBE Atomic Guess	**PBE Converged**	**HSE06 Converged**
CPU Model	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)
Mean Computation Time	29.43 s	6 min 15.77 s (375.77 s)	23 min 54.78 s (1434.78 s)

Table S19. Based on 200 snapshots extracted from a 2 ps *NVE* trajectories simulation at 300 K of the AB-$TiSe_2$/$HfSe_2$ structure, the mean computational time per structure was obtained for Hamiltonian evaluations by dividing the total wall-clock time by the number of snapshots. The timings were measured on 36 CPU cores at three levels: PBE (atomic guess), PBE (converged), and HSE06 (converged).

	PBE Atomic Guess	**PBE Converged**	**HSE06 Converged**
CPU Model	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)
Mean Computation Time	18.92 s	2 min 26.75 s (145.75 s)	19 min 52.78 s (1192.78 s)

Table S20. Based on 200 snapshots extracted from a 2 ps *NVE* trajectories simulation at 300 K of the BA-$TiSe_2$/$HfSe_2$ structure, the mean computational time per structure was obtained for Hamiltonian evaluations by dividing the total wall-clock time by the number of snapshots. The timings were measured on 36 CPU cores at three levels: PBE (atomic guess), PBE (converged), and HSE06 (converged).

	PBE Atomic Guess	**PBE Converged**	**HSE06 Converged**
CPU Model	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz (**36 cores**)
Mean Computation Time	18.94 s	2 min 36.56 s (156.56 s)	20 min 34.94 s (1234.94 s)

Table S21. Average wall-clock time required for a single structure calculation at the PBE and HSE06 levels for four representative stacking configurations. The last column shows the ratio of HSE06 to PBE computation times.

	PBE Converged	HSE06 Converged	HSE06/PBE
AB-$ZrTe_2$/$HfTe_2$	186.78 s	1272.50 s	6.81
BA-$ZrTe_2$/$HfTe_2$	375.77 s	1434.78 s	3.82
AB-$TiSe_2$/$HfSe_2$	145.75 s	1192.78 s	8.18
BA-$TiSe_2$/$HfSe_2$	156.56 s	1234.94 s	7.89

Table S22. With a training-set size of 200, the computational time required to obtain the Hamiltonians of the full trajectories for four heterostructures (AB-$ZrTe_2$/$HfTe_2$, BA-$ZrTe_2$/$HfTe_2$, AB-$TiSe_2$/$HfSe_2$, and BA-$TiSe_2$/$HfSe_2$) was evaluated using both DFT calculations and ML predictions. For the DFT calculations, each snapshot was computed at the PBE and HSE06 levels. The PBE-level calculations provided the wavefunction files, which were essential for accelerating the subsequent HSE06 computations. For the ML predictions, the workflow included (i) calculating the Hamiltonians of the entire 2000-snapshot trajectory at the PBE (atomic guess) level, (ii) computing HSE06 Hamiltonians for a selected subset of 200 snapshots to train the model, (iii) training the ML model, and (iv) mapping the PBE (atomic guess) Hamiltonians to the corresponding HSE06 Hamiltonians for the full trajectory using the trained model.

DFT Calculation				
	AB-$ZrTe_2$/$HfTe_2$	BA-$ZrTe_2$/$HfTe_2$	AB-$TiSe_2$/$HfSe_2$	BA-$TiSe_2$/$HfSe_2$
PBE-level Calculation (36 cores)	186.78 s * 2000 = 103.77 h	375.77 s * 2000 = 208.76 h	145.75 s * 2000 = 80.97 h	156.56 s * 2000 = 86.98 h
HSE06- level Calculation (36 cores)	1272.5 s * 2000 = 706.94 h	1434.78 s * 2000 = 797.1 h	1192.78 s * 2000 = 662.66 h	1234.94 s * 2000 = 686.08 h
Total Time	810.71 h	1005.86 h	743.63 h	773.05 h

ML Predicted				
	AB-$ZrTe_2$/$HfTe_2$	BA-$ZrTe_2$/$HfTe_2$	AB-$TiSe_2$/$HfSe_2$	BA-$TiSe_2$/$HfSe_2$
PBE Atomic Guess-level Calculation (36 cores)	20.52 s * 2000 = 11.40 h	29.43 s * 2000 = 16.35 h	18.92 s * 2000 = 10.51 h	18.94 s * 2000 = 10.52 h
HSE06-level Calculation for Training set (36 cores)	1272.5 s * 200 = 70.69 h	1434.78 s * 200 = 79.71 h	1192.78 s * 200 = 66.27 h	1234.94 s * 200 = 68.61 h
ML-Train Time (1 cores)	0.013 h	0.012 h	0.011 h	0.012 h
ML-Predicted Time (1 cores)	6.46 s * 2000 = 3.59 h	6.24 s * 2000 = 3.47 h	6.76 s * 2000 = 3.76 h	6.83 s * 2000 = 3.79 h
Total Time	85.69 h	99.54 h	80.55 h	82.94 h

Table S23. Structural information of the primitive cells used to construct the $ZrTe_2/HfTe_2$ and $TiSe_2/HfSe_2$ heterostructures. The listed parameters include the in-plane lattice constants, gamma angle, number of atoms in the primitive cell, space group, and point group.

	a=b (Å)	gamma angle (°)	number of atoms in the primitive cell	space group	point group
$ZrTe_2$	3.97	120.00	3	$P6_3/mmc$	D_6h
$HfTe_2$	3.95	120.00	3	$P6_3/mmc$	D_6h
$TiSe_2$	3.51	120.00	3	$P6_3/mmc$	D_6h
$HfSe_2$	3.69	120.00	3	$P6_3/mmc$	D_6h

Table S24. Structural information of the AB- and BA-stacked $ZrTe_2/HfTe_2$ and $TiSe_2/HfSe_2$ heterostructures. The listed parameters include the in-plane lattice constants, gamma angle, number of atoms in the primitive cell, and number of atoms in the 3 × 3 supercell.

	a (Å)	b (Å)	gamma angle (°)	number of atoms in the primitive cell	number of atoms in the supercell
AB-$ZrTe_2/HfTe_2$	11.87	20.56	90.00	6	108
BA-$ZrTe_2/HfTe_2$	11.87	20.56	90.00	6	108
AB-$TiSe_2/HfSe_2$	10.80	18.71	90.00	6	108
BA-$TiSe_2/HfSe_2$	10.80	18.71	90.00	6	108