Supporting Information of

Oxidation Mechanisms and Oxygen Migration Dynamics on Octa-Penta Graphene

Chenyi Zhou,[1] Rubin Huo,[1] Boyi Situ,[1] Zihan Yan,[2] Yusong Tu,[1] Zhe Zhang,[1,*]

[1]*College of Physics Science and Technology, Yangzhou University, Jiangsu 225009, China*

[2]*School of Engineering, Westlake University, Hangzhou, Zhejiang 300024, China*

[*]Corresponding Authors: zzhang@yzu.edu.cn

1. The detailed explanation of the machine-learning-based molecular dynamics (MLMD) procedure.

(1) The MLMD Algorithm: The MLMD algorithm trains a machine learning model on quantum mechanical data to learn relationships between atomic configurations, potential energy, and atomic forces. The trained model replaces DFT in molecular dynamics (MD) simulations, predicting forces/energies for new configurations and enabling faster simulations.

(2) Force Field Structure: The MLMD force field is derived from a machine learning model trained on DFT data. It utilizes the learned potential energy surface to compute forces/energies for untrained atomic configurations, accurately modeling complex interactions (e.g., oxygen migration on OPG monolayers). Physical properties like bond lengths and angles are optimized to match the system’s behavior.

(3) Machine Learning Model Algorithm: The model learns the potential energy surface by fitting DFT-derived energy/force data. Inputs include atomic positions and types, while outputs are total energy and forces. Trained on diverse OPG and oxygen configurations, the model generalizes to predict energies/forces for unseen systems. Model selection and training align with VASP-recommended practices.

(4) Input and Output Information: Inputs include atomic types, initial configurations, potential energy surfaces, and atomic positions (iteratively updated during MD). Outputs comprise atomic forces, total energy, and time-resolved atomic trajectories. These enable analysis of dynamic processes (e.g., oxygen migration, bond formation) and oxidation mechanisms in OPG materials.

2. The approach to estimate the force in MLMD .

(1) Generation of training data

DFT-based force labels: The training dataset was constructed using first-principles DFT calculations performed with VASP. The atomic forces computed from these calculations serve as the ground-truth labels for training the machine learning force field (MLFF).

Dataset composition: To ensure robust force predictions, we generated 3543

atomic configurations for OPG-L and 2663 configurations for OPG-Z, each containing 1–3 oxygen atoms randomly adsorbed on the monolayer. These configurations were selected to capture diverse oxygen adsorption environments and migration pathways.

(2) Training the machine learning model

Model selection: We employed a neural network potential trained on the DFT dataset to learn the underlying potential energy surface and force interactions.

Feature engineering: The model was trained using local atomic descriptors, such as interatomic distances, bond angles, and coordination numbers, ensuring that the representation effectively captures the physical interactions governing force predictions.

Validation: The trained model was evaluated on an independent test set to ensure high accuracy in force predictions, with root mean square errors (RMSEs) of 0.142 eV/Å for OPG-L and 0.119 eV/Å for OPG-Z, demonstrating reliable force estimation.

(3) Force Prediction During MLMD Simulations

Real-time force calculation: During MLMD simulations, the trained model predicts atomic forces based on new configurations encountered in the trajectory. These predicted forces are then used to update atomic positions and velocities.

Self-consistency check: The predicted forces are validated against a subset of DFT-calculated forces to confirm their accuracy throughout the simulation.

(4) Integration with molecular dynamics simulations

Time-stepping algorithm: The predicted forces are incorporated into the MLMD framework using the Verlet integration scheme, which efficiently propagates atomic motion over time.

Trajectory evolution: The MLMD simulations are performed over 50 ps, allowing for a detailed analysis of oxygen migration and reaction dynamics on the OPG monolayers.

3. Workflow example: input and output process of a representative on-the-fly MLFF training case

Below is a simplified and clarified example illustrating the input and output data workflow for one representative training instance in our study:

(1) Prepare a molecular dynamics (MD) Simulation

The initial step involves setting up an ab-initio MD simulation through the following inputs: the initial atomic structure defined in POSCAR, and systematically tuned computational parameters in the INCAR, KPOINTS, and POTCAR files.

(2) Start on-the-fly training from scratch

The MLFF implementation is activated using specific tags in the INCAR file, grouped under the ML_ prefix. The fundamental requirement is enabling the master switch:

ML_LMLFF = .TRUE.

Without this Boolean parameter, all MLFF-related functionalities remain disabled and

VASP defaults to standard ab-initio calculations. To initiate active learning, the meta-tag must be explicitly configured:

ML_ISTART = 0

In this mode, VASP dynamically alternates between performing ab-initio reference calculations and MLFF predictions during MD simulation. Alongside standard MD trajectory outputs (XDATCAR, OUTCAR), VASP also generates three critical MLFF-specific files:

ML_LOGFILE: Records real-time training metrics, including error statistics (identified by ERR-prefixed entries) and convergence diagnostics.

ML_ABN: Archives atomic configurations selected for training and maintains reference datasets.

ML_FFN: Stores the iteratively updated ML force field in a binary format.

These MLFF-specific files are continuous updated during the simulation. Upon completion of the specified number of NSW steps, ML_ABN consolidates the final training dataset, while ML_FFN contains the finalized force field.

(3) Continue on-the-fly training from an existing training database

This step enables iterative refinement of the MLFF by extending training to new configurations (e.g., varied temperatures, compositions, or phases). To continue training:

(i) Initialize a new MD simulation with updated POSCAR (potentially reused from a previous simulation's CONTCAR).

(ii) Transfer the previous training data by copying: **cp ML_ABN ML_AB**

(iii) Set the training mode and restart VASP with:

ML_LMLFF = .TRUE.

ML_ISTART = 1 .

The hybrid training procedure continues seamlessly, now integrating prior and newly generated structural data. After multiple training iterations, the final combined dataset and optimized force field parameters are stored in the updated ML_ABN and ML_FFN files, respectively. To reset the training history, remove ML_AB.

(4) Applying the finalized MLFF in production simulations

The finalized MLFF obtained from Step 3 is deployed in prediction mode by:

(i) Transferring the trained force field to a new simulation environment:

cp ML_FFN ML_FF

(ii) Configuring VASP exclusively to use MLFF predictions by setting:

ML_LMLFF = .TRUE.

ML_ISTART = 2

In this state, VASP exclusively utilizes the MLFF to predict forces and energies, bypassing computationally expensive ab-initio calculations. This drastically reduces the computational cost per ionic step, often by orders of magnitude, compared to standard hybrid/DFT simulations.

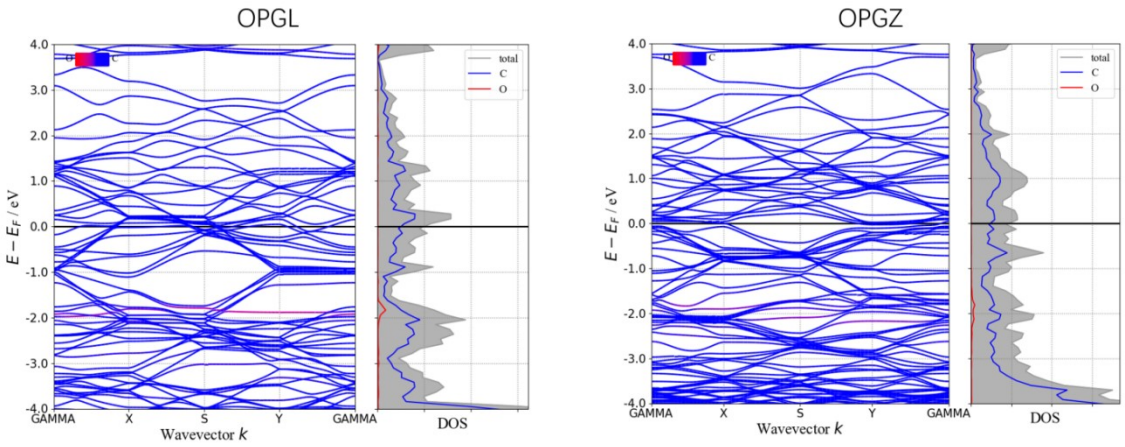


Fig. S1 Electronic band structures and density of states (DOS) for pristine OPG-*L* and OPG-*Z*, respectively.

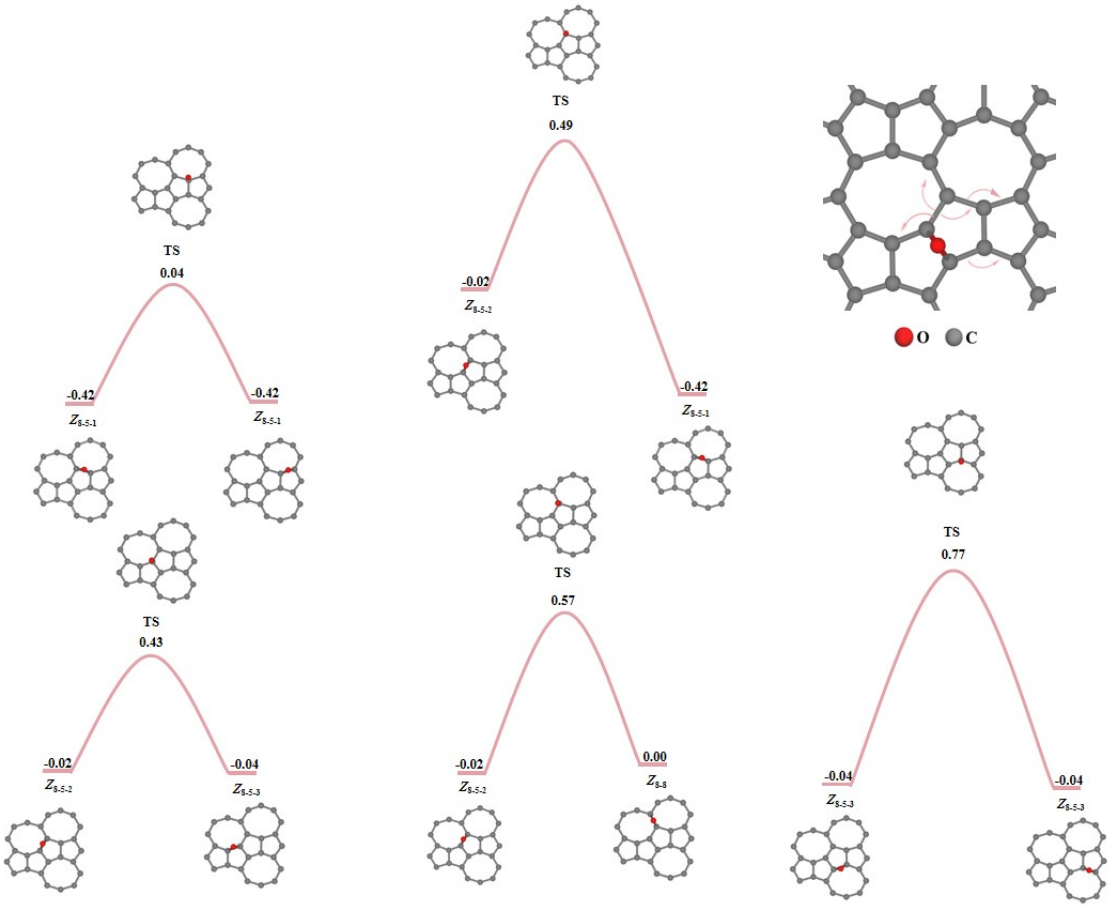
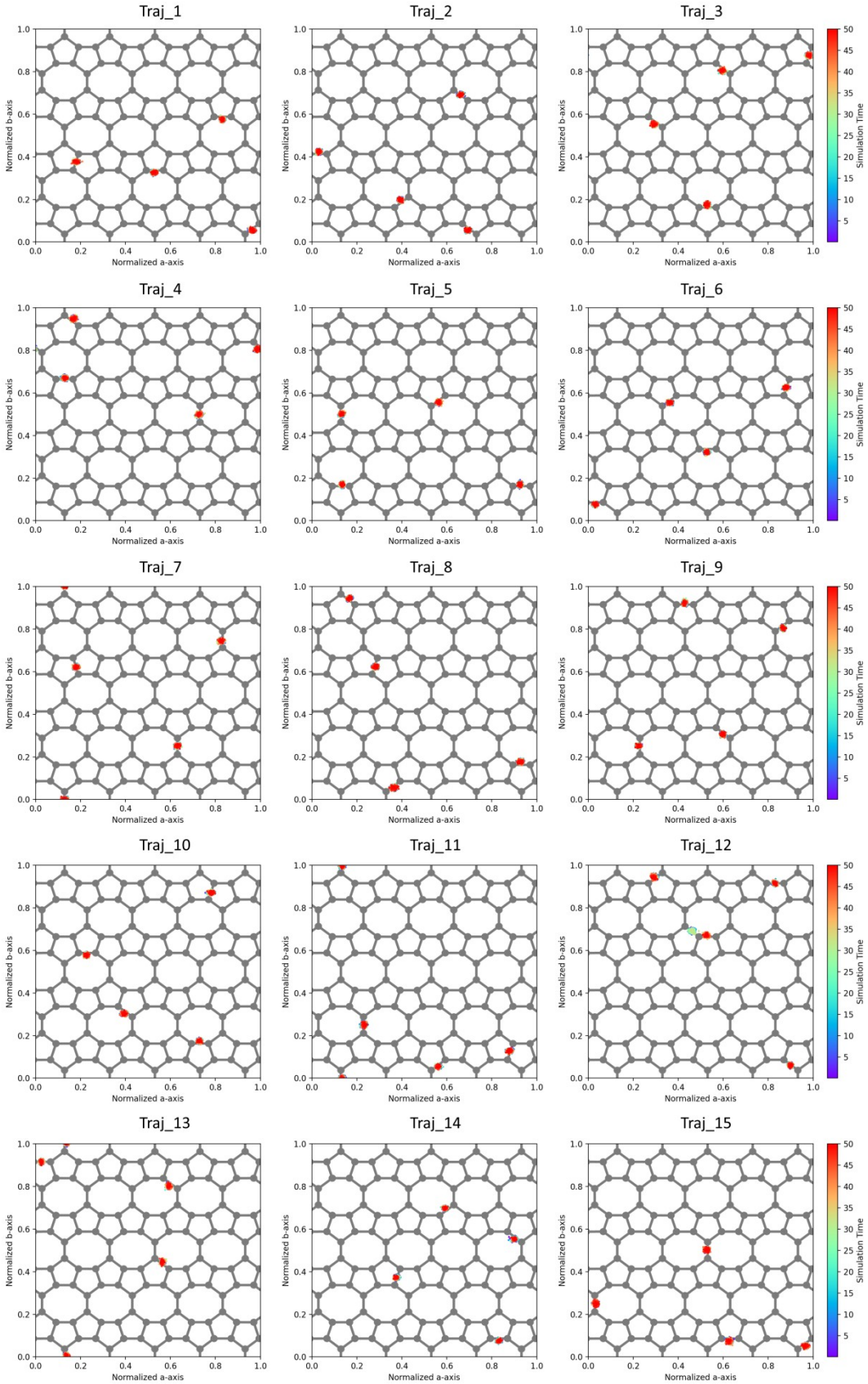
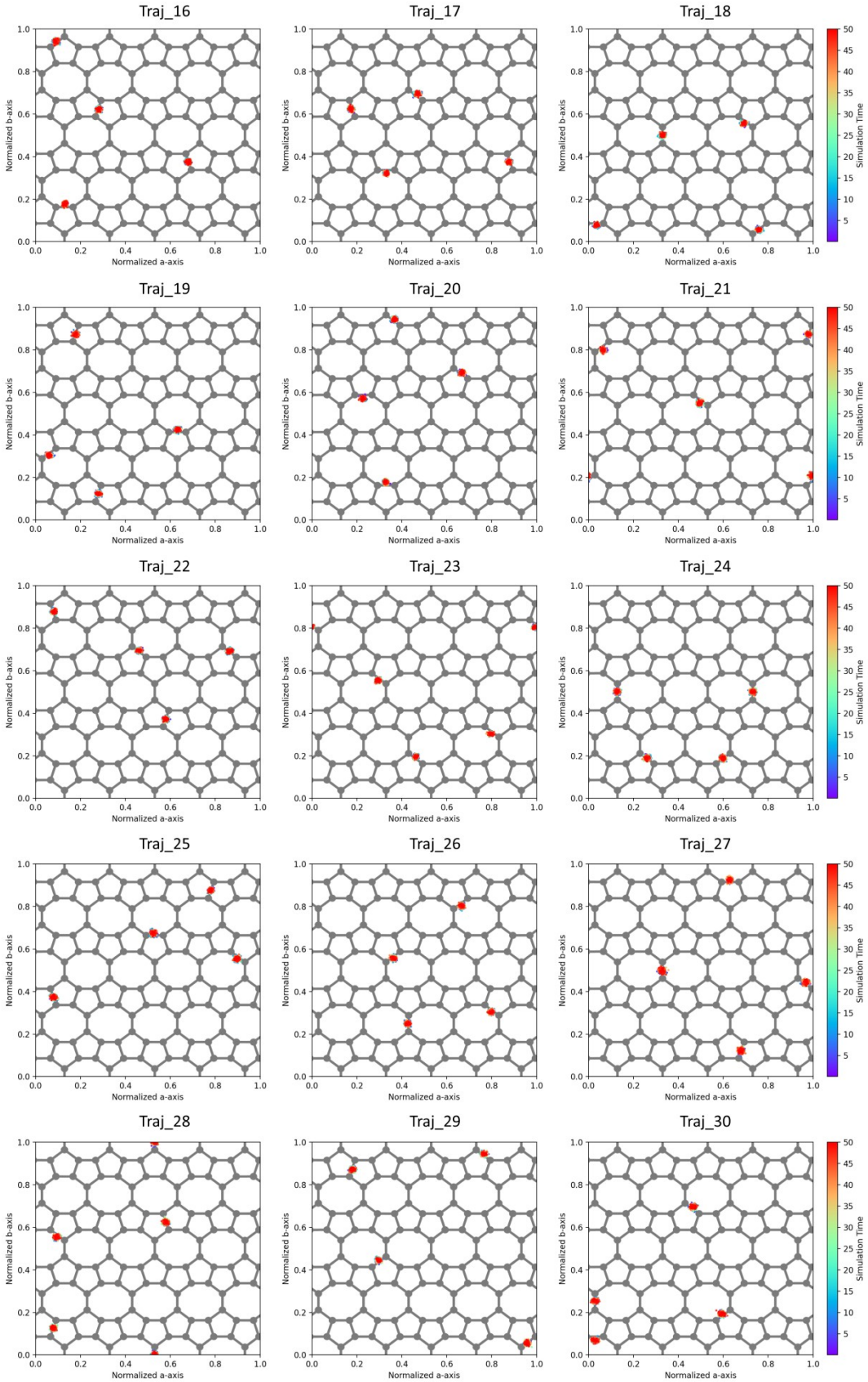
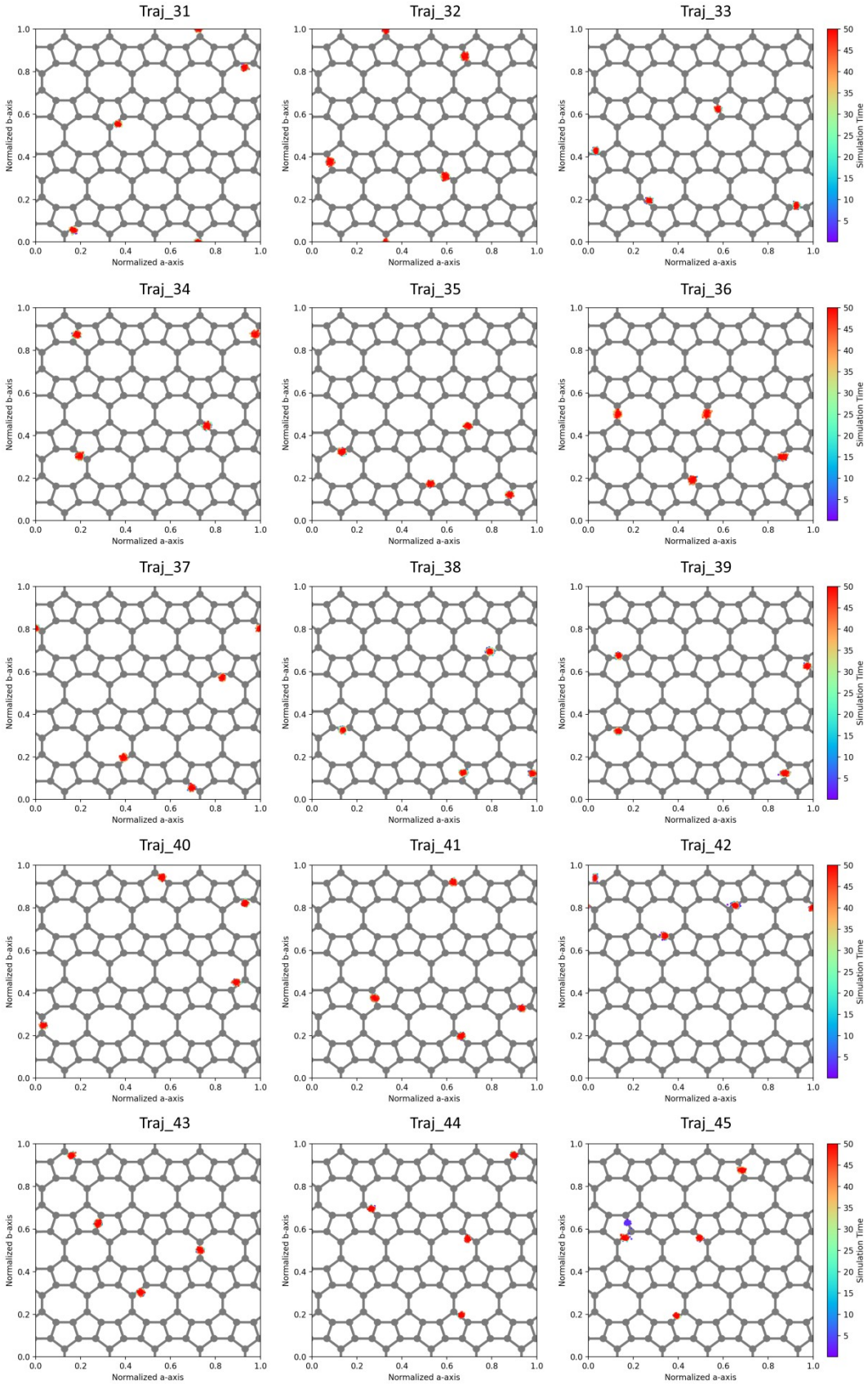
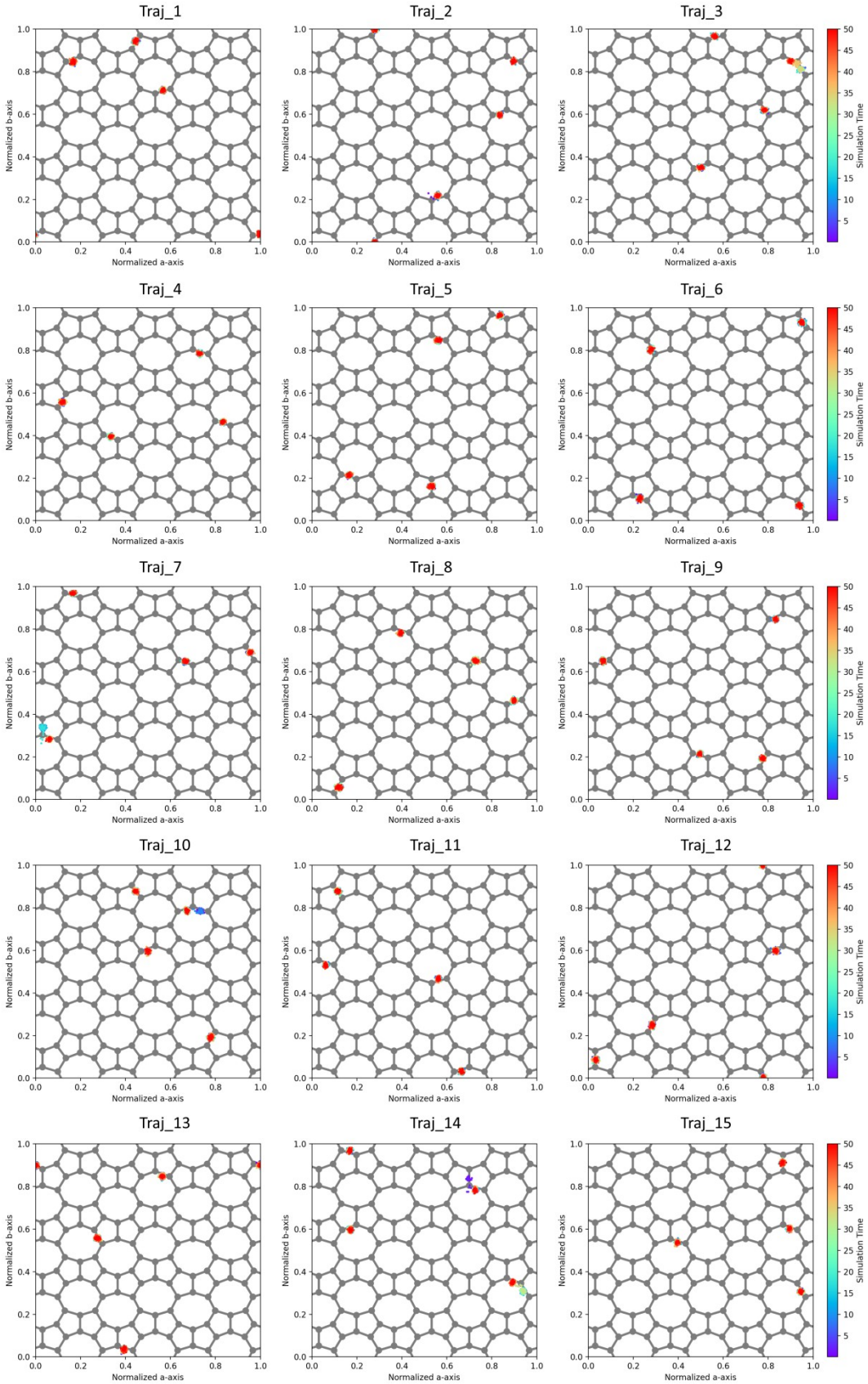


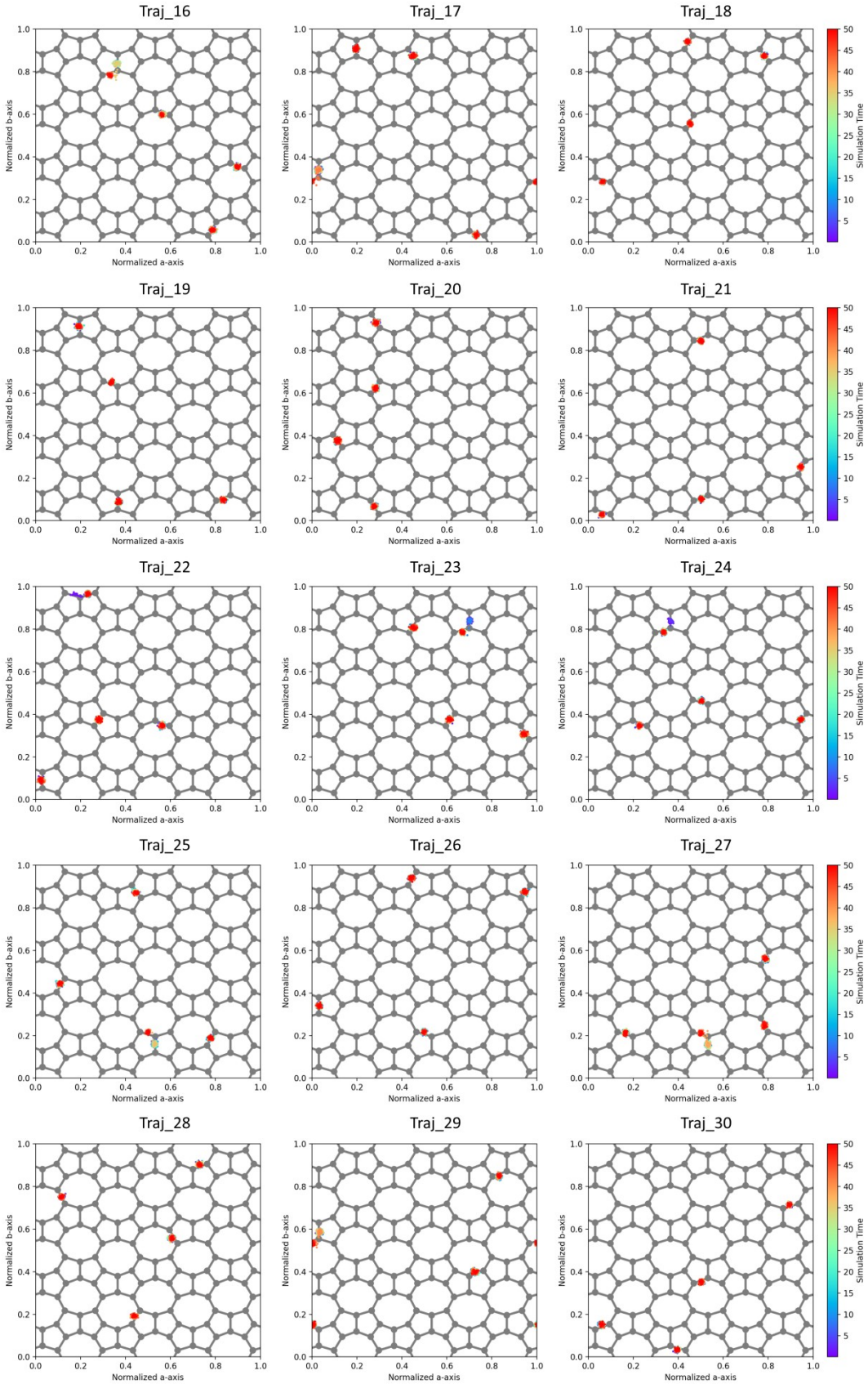
Fig. S2 Extra oxygen migration pathways on the OPG-*Z*. The energy value of $Z_{8\text{-}8}$ is shifted to 0 eV. The non-equivalent carbon sites are labeled by the same notations used in Fig. 2. The direction of the arrow indicates the possible pathways of the oxygen migration.


Traj_1
Traj_2
Traj_3
Traj_4
Traj_5
Traj_6
Traj_7
Traj_8
Traj_9
Traj_10
Traj_11
Traj_12
Traj_13
Traj_14
Traj_15
Normalized a-axis
Normalized b-axis
Simulation Time


Traj_16
Traj_17
Traj_18
Traj_19
Traj_20
Traj_21
Traj_22
Traj_23
Traj_24
Traj_25
Traj_26
Traj_27
Traj_28
Traj_29
Traj_30
Normalized a-axis
Normalized b-axis
Simulation Time


Traj_31
Traj_32
Traj_33
Traj_34
Traj_35
Traj_36
Traj_37
Traj_38
Traj_39
Traj_40
Traj_41
Traj_42
Traj_43
Traj_44
Traj_45
Normalized a-axis
Normalized b-axis
Simulation Time


Traj_1
Traj_2
Traj_3
Traj_4
Traj_5
Traj_6
Traj_7
Traj_8
Traj_9
Traj_10
Traj_11
Traj_12
Traj_13
Traj_14
Traj_15
Normalized a-axis
Normalized b-axis
Simulation Time


Traj_16
Traj_17
Traj_18
Traj_19
Traj_20
Traj_21
Traj_22
Traj_23
Traj_24
Traj_25
Traj_26
Traj_27
Traj_28
Traj_29
Traj_30
Normalized a-axis
Normalized b-axis
Simulation Time

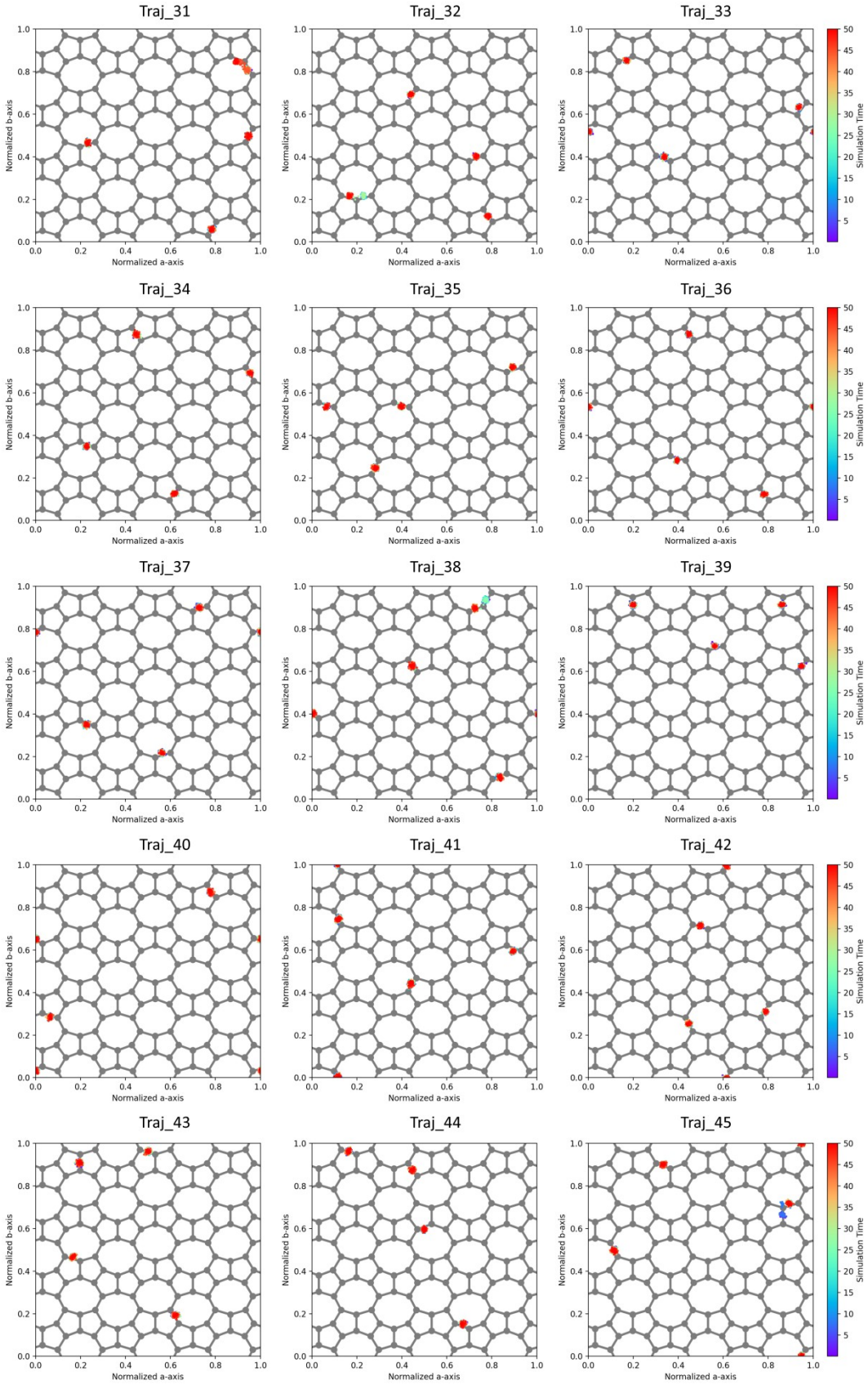


Fig. S3 MLMD simulations trajectories of oxygen migration on OPG-*L* and OPG-*Z,* respectively. The colored dots denote the positions of 4 oxygen atoms, and the colors denote the simulation time. The static OPG monolayer is used as the background for clarity.