Supplementary Information

Supplementary Information

1 Training results of the GAP-CN potential

Numerical validation The quality of the developed GAP-CN potential is crucial for the accuracy of subsequent modelling of nitrogen-doped diamond samples and is validated against DFT-based energies per atom and forces. This comparison is shown in figure SI 1**a.** and SI 1**b.**. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are taken as the measure of the difference between the result from GAP-CN and the result from DFT. It is observed in the scatter plot that energy data demonstrated a good fit as the value of the MAE is less than 20 meV/atom and RMSE is in the region of 51 meV/atom, whereas, Force data showed a very strong fit as the MAE is less than 0.15 eV/Å and RMSE is in the region of less than 0.3 eV/Å.

Further validation of the force and energy for each defective carbon-nitrogen-related structure, along with the distribution of each configuration, is illustrated in Figure SI 1**c.**, **d.** and **e.**. From the graph, it is shown that both the RMSE and MAE for force errors are below 0.3 eV/Å, while for energy errors, they are all below 0.01 eV/atom. It is notable that the defective structure containing one interstitial nitrogen and one substitutional nitrogen, marked as 'Ni_Ns', occupies the largest number of configurations to maintain consistent force and energy accuracy with other types of defective structures.

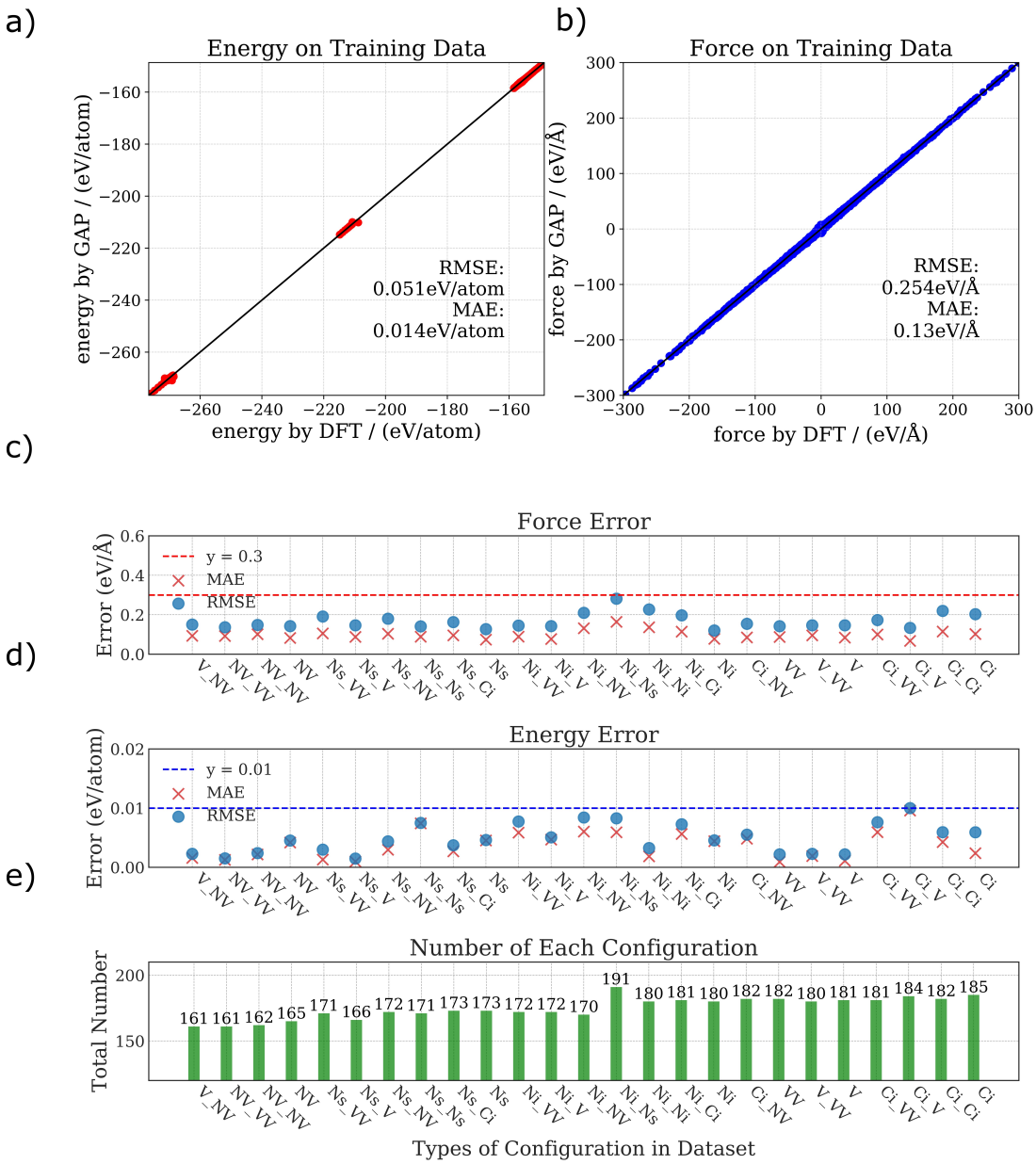

Figure SI 1: **a.** A scatter plot comparing the energy results obtained from DFT and GAP. **b.** A scatter plot comparing the force results between DFT and GAP. An illustration of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for various defective structures concerning **c.** force and **d.** energy. Plot **e.** demonstrates the distribution of each defective structure within the data samples.

Supplementary Figures

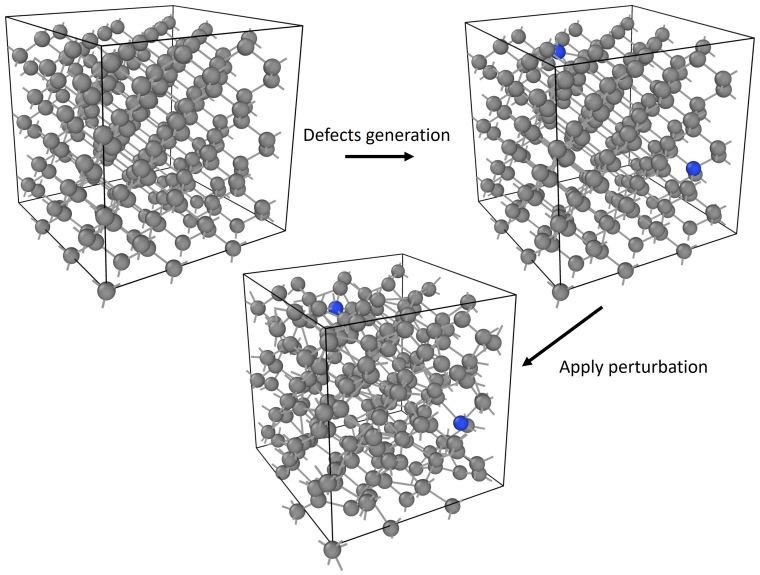

Figure SI 2: An illustration of the construction of one sample containing a substitutional nitrogen (N_s) and an interstitial nitrogen (N_i) in a $3 \times 3 \times 3$ supercell.

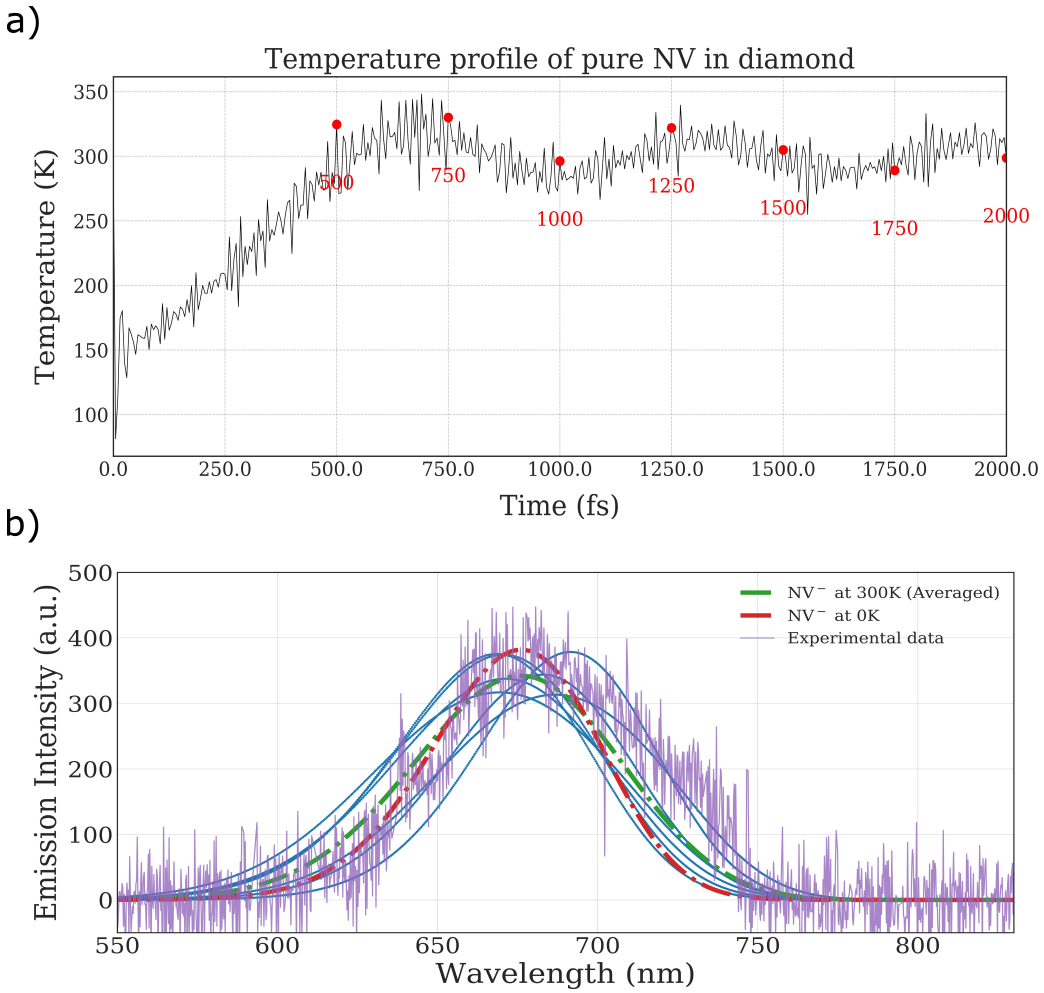

Figure SI 3: **a.** Temperature profile of the molecular dynamics simulation of a pure NV centre in diamond at 300 K. The red points indicate sampled structures used for TDDFT simulations, with emission spectrum results presented in **b.**

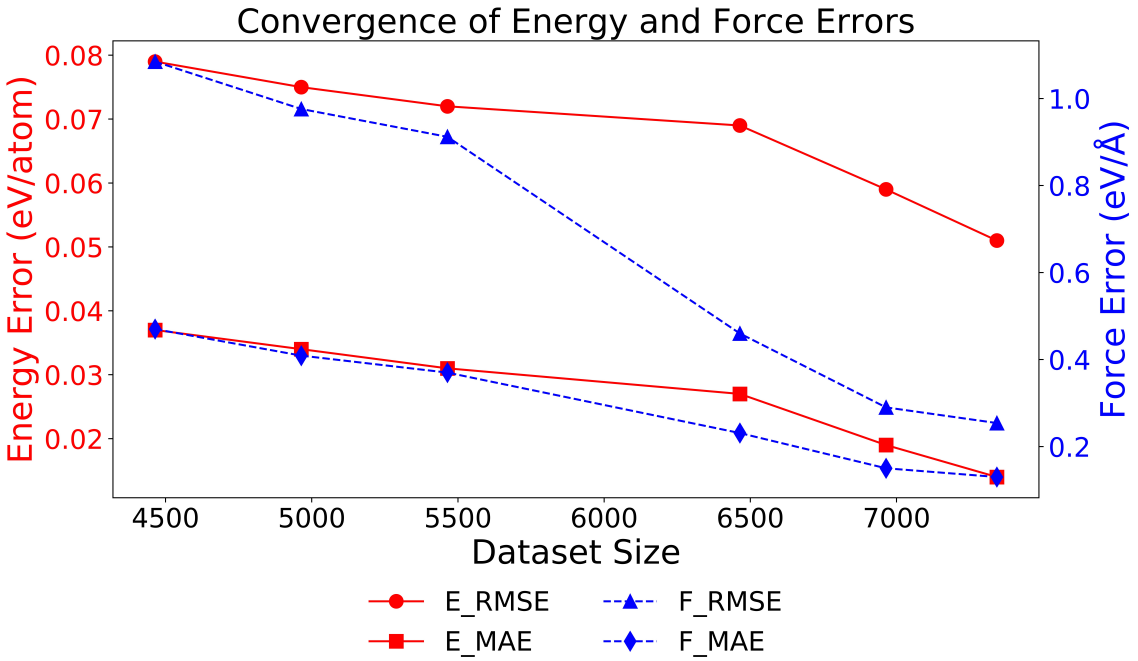

Figure SI 4: Graphs showing the convergence of the GAP-CN potential training with respect to force and energy. The number of training configurations for sets 1 to 6 are 4464, 4964, 5464, 6454, 6964, and 7343, respectively.

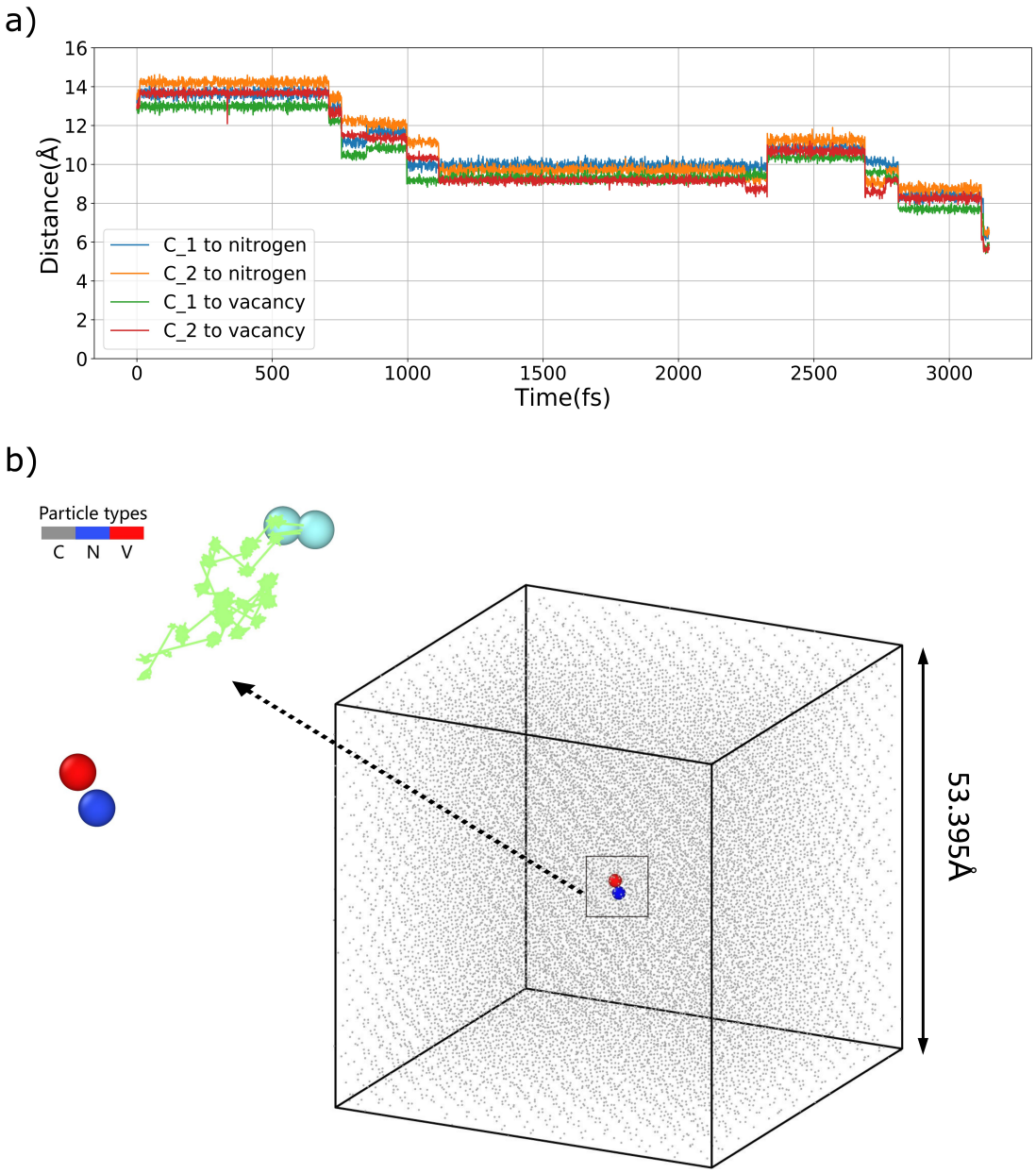

Figure SI 5: Examples of GAP-CN-assisted molecular dynamics simulations in a $15 \times 15 \times 15$ conventional diamond supercell (27,000 atoms) containing one NV center and one C_i. **a.** shows the distance traces between two interstitial carbons and the nitrogen and vacancy sites, while **b.** illustrates the trajectory paths of the interstitial carbons within the diamond lattice.

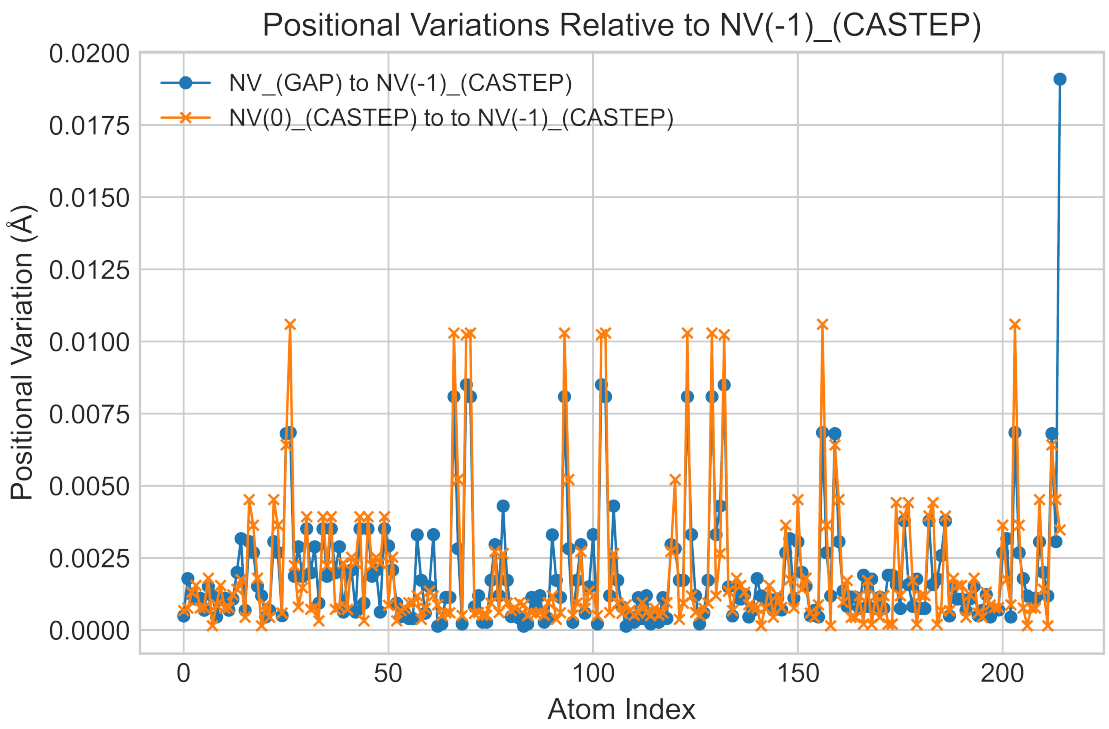

Figure SI 6: Impact of distance variations on the optimized NV results across different methods and charge states