Supplemental Information: Hydrogen in Disordered Titania: Connecting Local Chemistry, Structure, and Stoichiometry through Accelerated Exploration

James Chapman [*1,4], Kyoung E. Kweon [†1], Yakun Zhu[‡1], Kyle Bushick[1,3], Leonardus Bimo Bayu Aji[1], Christopher A. Colla[1], Harris Mason[1], Nir Goldman[1,2], Nathan Keilbart[1], S. Roger Qiu[1], Tae Wook Heo[1], Jennifer Rodriguez[1], and Brandon C. Wood[§1]

[1]*Materials Science Division, Lawrence Livermore National Laboratory, Livermore, CA, USA*
[2]*Department of Chemical Engineering, University of California, Davis, California 95616, United States*
[3]*Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States*
[4]*Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215, United States*

December 21, 2022

S1 Atomic force neural network

S1.1 Atomic-level structural descriptors

An atom's local geometry is decomposed into numerical descriptors which are then mapped to a specific atomic force component. These descriptors aim to capture unique pairwise aspects of an atom's local atomic environment by observing the changes in the atomic probability densities at various intervals around the atom. The functional form of the atomic-level descriptors are defined as [1]:

$$\mathbf{v}_{i,\alpha;k_{[i,j]}}^{[\psi_i,\psi_j]} = \frac{1}{a_{k_{[i,j]}}^3 (2\pi)^{3/2}} \sum_{j\neq i} \frac{r_{ij}^\alpha}{r_{ij}} exp\left(-\frac{(r_{ij}-a_{k_{[i,j]}})^2}{w^2}\right) f_{cut}(r_{ij}) \tag{1}$$

Here, r_i and r_j are the Cartesian coordinates of atoms i and j, $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$, r_{ij}^α is the projection of $\mathbf{r}_j - \mathbf{r}_i$ onto any arbitrary direction α. $[\psi_i, \psi_j]$ represents the chemical identities of the interaction. $k_{[i,j]}$ represents a given atomic feature for the specific $[i,j]$ interaction. The summation runs over the neighbor list set $\{j\}$ of atom i, while $\frac{1}{k_{[i,j]}^3 (2\pi)^{3/2}}$ is a normalization constant. The damping function $f_{cut}(r_{ij}) = \frac{1}{2}[cos(\frac{\pi r_{ij}2}{R_{cut}^2}) + 1]$, which accounts for a smooth degradation in an atom's contribution to the atomic force exerted on atom i, has a cut-off radius R_{cut} chosen to be 7 Å. The gaussian functions are placed at various distances, $a_{k_{[i,j]}}$ away from atom i, with widths controlled by w. The $a_{k_{[i,j]}}$ values are controlled manually by matching their peaks with those observed in the radial distribution function of the material. Once the peaks have been matched, the remaining distances are then filled in by placing equidistantly spaced features. The gaussian functions are placed from 0.1 Å to 7 Å, to ensure that the underlying potential energy surface is appropriately sampled. In this work, the final atomic feature vector, for a given atom i, along direction α will be given as:

$$V_{i,\alpha} = \{v_{i,\alpha;k_{[i,i]}}^{[\psi_i,\psi_i]}, v_{i,\alpha;k_{[i,j]}}^{[\psi_i,\psi_j]}\} \tag{2}$$

*Corresponding Author, jc112358@bu.edu
†Corresponding Author, kweon1@llnl.gov
‡Corresponding Author, zhu15@llnl.gov
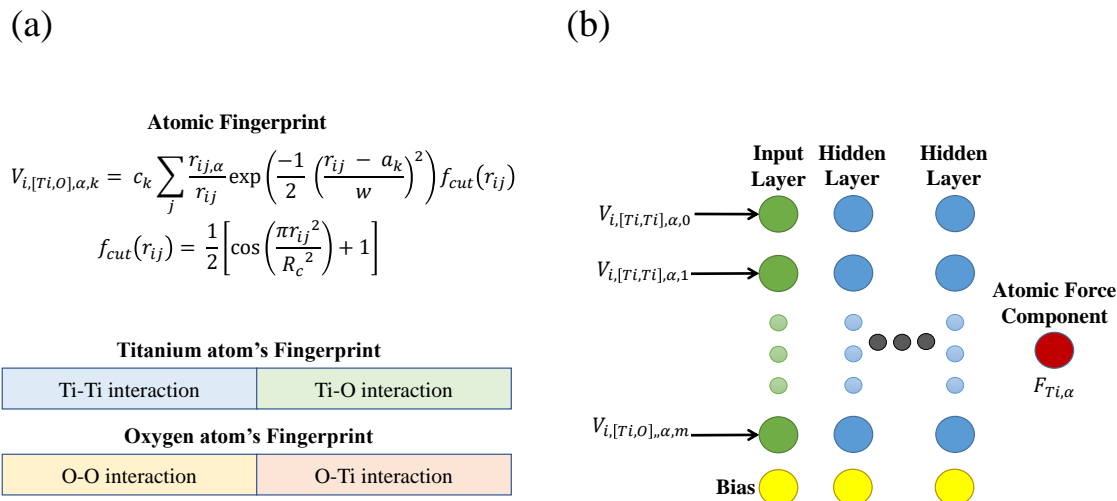§Corresponding Author, wood37@llnl.gov

(a)   (b)

Atomic Fingerprint

$$V_{i,[Ti,O],\alpha,k} = c_k \sum_j \frac{r_{ij,\alpha}}{r_{ij}} \exp\left(\frac{-1}{2}\left(\frac{r_{ij}-a_k}{w}\right)^2\right) f_{cut}(r_{ij})$$

$$f_{cut}(r_{ij}) = \frac{1}{2}\left[\cos\left(\frac{\pi r_{ij}^2}{R_c^2}\right) + 1\right]$$

Titanium atom's Fingerprint

Ti-Ti interaction	Ti-O interaction

Oxygen atom's Fingerprint

O-O interaction	O-Ti interaction

Input Layer Hidden Layer Hidden Layer

$V_{i,[Ti,Ti],\alpha,0}$

$V_{i,[Ti,Ti],\alpha,1}$

Atomic Force Component

$F_{Ti,\alpha}$

$V_{i,[Ti,O],,\alpha,m}$

Bias

Figure S1: (a) Workflow of the atomic feature construction process for the titanium AFNN. The atomic feature functional form and corresponding eometric damping function (top) for the case of a titanium atom are shown. A visual depiction of the concatenated final feature vector for titanium and oxygen (bottom) are provided. The current specie's self interaction features are always included first, while the corresponding hetergenous interactions are included after the self-tineraction terms. (b) Visual description of the AFNN architecure is shown. The entire concatenated feature vector is fed into the AFNN, along with a bias term. This information is passed through several hidden layers, culminating in the prediction of a particular atomic force component. The example shown here is for the case of titaium.

S1.2   Model Construction

The Atomic force neural network (AFNN) employs a feature set containing self-interaction and hetergenous interaction terms [1]. Each term includes 48 gaussian functions, with means spaced $0.14375\,\mathring{A}$ apart as they move away from an atom's center. The first gaussian is palced $0.1\,\mathring{A}$ away from an atom's center. The width of each gaussian was set uniformly at $0.2\,\mathring{A}$. The same feature set was used for both chemical species studied in this work. The final feature vector contained 96 features, as it is a concatenation of interaction terms. Figure S1 provides a visual depiction of the functional form used to calculate the feature vectors, as well as the interaction term concatenation process.

Table S1 provides technical information regarding the neural network artchitectures used for the Ti and O AFNNs. Figure S1 (b) also provides a visual description of the AFNN setup. For both Ti and O AFNNs, 96 features were fed into the input layer of the neural network, along witha single bias term. The input layer was connected to a hidden layer containing 192 neurons and a bias node. A second, identical hidden layer was connected to the second layer, followed by an final expanded layer containing 768 neurons. These 768 neurons were summed together linearly to predict the atomic force component. The artchitecture employed here was identical between the chemical specie's respective models.

Table S1: Neural network architecture for the oxygen and titanium AFNNs. The number of layers provided does not include the input layer.

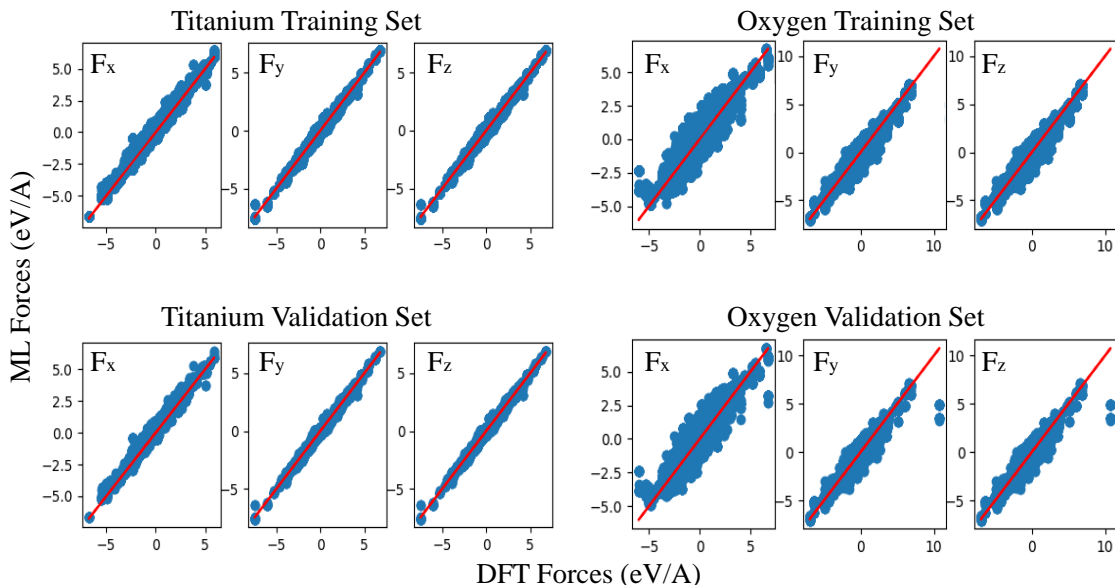| Chemical Identity	# Layers	# Neurons
Ti	3	192, 192, 768
O	3	192, 192, 768

Figure S2: (left) Atomic forces for the training (top) and validation (bottom) sets for titanium force prediction, shown against the corresponding DFT predictions. (right) Atomic forces for the training (top) and validation (bottom) sets for oxygen force prediction, shown against the corresponding DFT predictions.

The final mathematical form of the NN is expressed in matrix form as:

$$\vec{u}^{(1)}(i) = \vec{V}_i * \hat{w}^{(0,1)} \tag{3}$$

$$\vec{u}^{(n)}(i) = \vec{f}(\vec{u}^{(n-1)}(i)) * \hat{w}^{(n-1,n)} + b^{(n-1,n)} \ni n > 1 \tag{4}$$

Here, $\vec{u}^{(n)}$ represents a given set of neurons for a particular layer n, with $n = 1$ representing the input layer. $\vec{V}_i$ represents the atomic feature vector used for a particular atom i. $w^{(n-1,n)}$ is the weight matrix of size $(M_{n-1} \times M_n)$, where M_n is the number of neurons is a given layer n. As mentioned previously, $\vec{f} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. $b^{(n-1,n)}$ represents the bias term associated with a given neuron, outside of the input layer. The final layer used to predict a given atomic force component, is given as $f_i = b^{last} + \sum_m^{M_{last}} u_m^{last}(i)$.

S1.3 Prediction of atomic forces

Figure S2 shows parity plots for the AFNN and density functional theory (DFT) predicted atomic forces for the DFT-generated reference data. The top plots show the force predictions on the training sets of Ti and O, while the bottom plots provide forces for the validation sets. Good overall agreement is shown for both species, though a larger spread exists for the case of oxygen. Detailed statistics used to validate the AFNN can be found in Table S2

Table S2: Statistics for the atomic force preditions, suing the AFNN, compared to DFT. RMSE indicates the root mean square error, STD is the standard deviation, and KS represents the the p-value from a Kolmogorov-Smirnov [2] test.

Element	Data Set	RMSE ($\frac{eV}{\mathring{A}^2}$)	STD ($\frac{eV}{\mathring{A}^2}$)	KS
Ti				
	Training	0.15	0.16	0.97
	Validation	0.17	0.17	0.95
O				
	Training	0.34	0.32	0.91
	Validation	0.38	0.38	0.9

S1.4 Dynamic properties of TiO$_2$

Mean square displacements (MSD) were calculated using classical molecular dynamics (MD) and ab intio MD (AIMD). Classical MD simulations were performed in LAMMPS [3] using the AFNN. Each chmical species was tracked over the course of the simulation. Time was broken into windows (shifted initial temporal starting points) to allow for a smoother resulting MSD. A fickian diffusion model [4] was used to extract the self-diffusion constants from the resulting MSDs. Figure **??** shows the MSDs and corresponding self-diffusion constants for both Ti and O, showing good agreement between DFT and the AFNN.

S1.5 Structural properties of non-stoichiometric TiO$_{1.88}$

Time-averaged radial distribution function (RDF) for the AFNN, and static RDF for DFT, were calculated to show the structural agreement between the AFNN and DFT for the case of TiO$_{1.88}$. Results can be found in Fig. S4. The AFNN predicts a slight shift to the right in the first peak of the RDF, corresponding to Ti-O interactions. This shift is primarily due to a small shift to the left of the O-O peak, indicating that oxygen atoms are slightly closer together in the AFNN systems. However, it should be noted that the AFNN simulations cover a much larger region of the TiO$_{1.88}$ phase space than DFT, which only contain a single snapshot. Figure S4 shows the RDF predictions for the AFNN and DFT,decomposed into O-O, Ti-O, and Ti-Ti interaction curves.

S1.6 Graph characterization of amorphous phase space

The GCN is described by an adjacency matrix, with matrix elements defined as:

$$G^{i-j}_{k_i,k_j} = \frac{1}{d_{k_i,k_j}} \ni d_{k_i,k_j} \leq R_c \tag{5}$$

Here, i and j represent the chemical species of the atoms contained in the GCN. k_i and k_j are the atomic indices of a particular atom, from chemical specie i and j respectively. d_{k_i,k_j} is defined as the l^2-norm between two atoms. R_c is the cutoff radius specified when constructing the GCN. Each matrix element, $\frac{1}{d_{k_i,k_j}}$, represents the weight of a given edge for a specific pair of adjacent nodes in the graph. The degree of each node is defined as the sum of the elements in a node's edge set. The degree sets are then fed into the scalar graph order parameter (SGOP) [5] scheme for the final characterization of the atomic environments. In this work Ti-O GCNs are used to classify the amorphous titania phase space. The SGOP functional form is defined as:

$$\theta_{i-j,R_c} = \sum_s^S \left(\sum_m^{D_s} P(d_m) \log_b P(d_m) + d_m P(d_m) \right)^3 \tag{6}$$

Here, i and j represent the chemical identities of the atoms contained in the GCN. R_c is the cutoff radius specified when constructing the GCN. We make the assumption that a particular GCN is disconnected, and
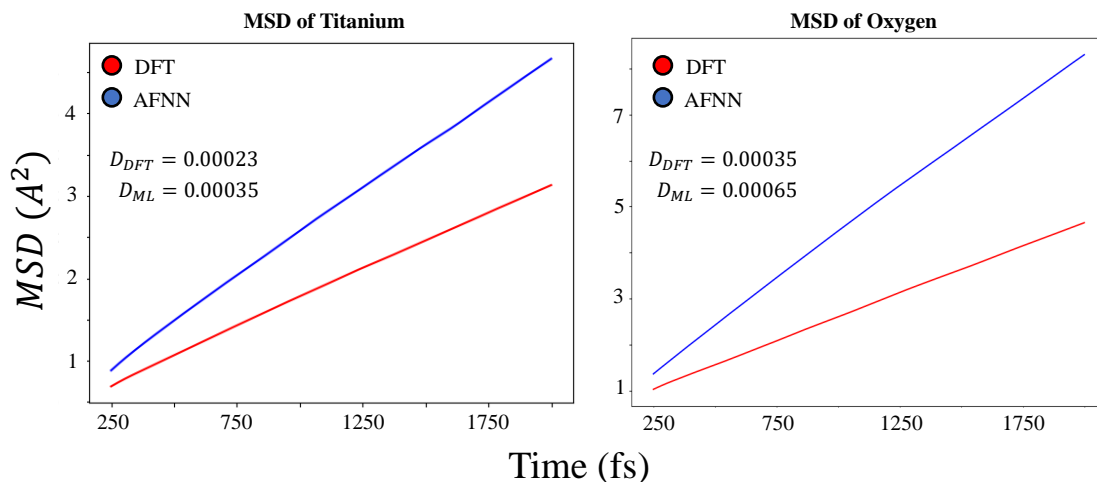
Figure S3: (left) Mean square displacement of titanium atoms at 2550K, plotted against time, shown in femtoseconds, for AFNN MD (blue) and AIMD (red) simulations. (right) Mean square displacement of oxygen atoms at 2550K, plotted against time, shown in femtoseconds, for AFNN MD (blue) and AIMD (red) simulations. Inserted values indicate the corresponding diffusion constant, calculated from the slope of the MSD.
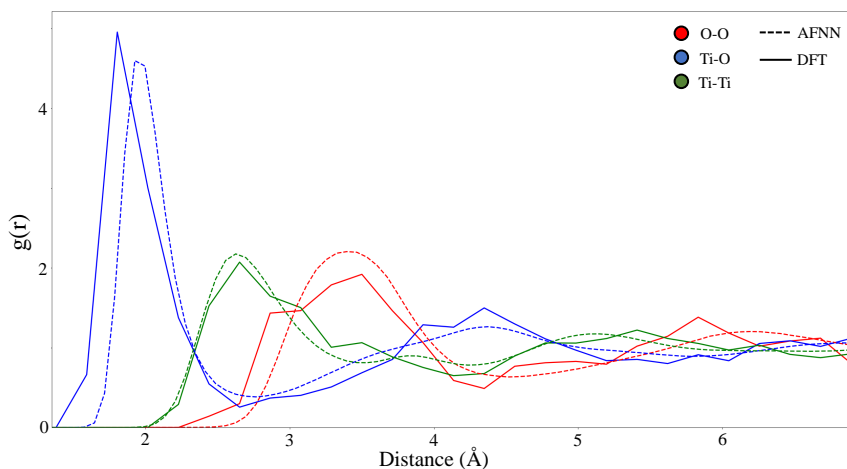

Figure S4: Molecular dynamics derived time-averaged pair-correlation function of amorphous $TiO_{1.88}$ at T=2250K, for both DFT and the AFNN. The pair-correlation function is decomposed based on chemical species interactions (colors). The AFNN is shown as the solid line, while DFT is given as a dashed line.
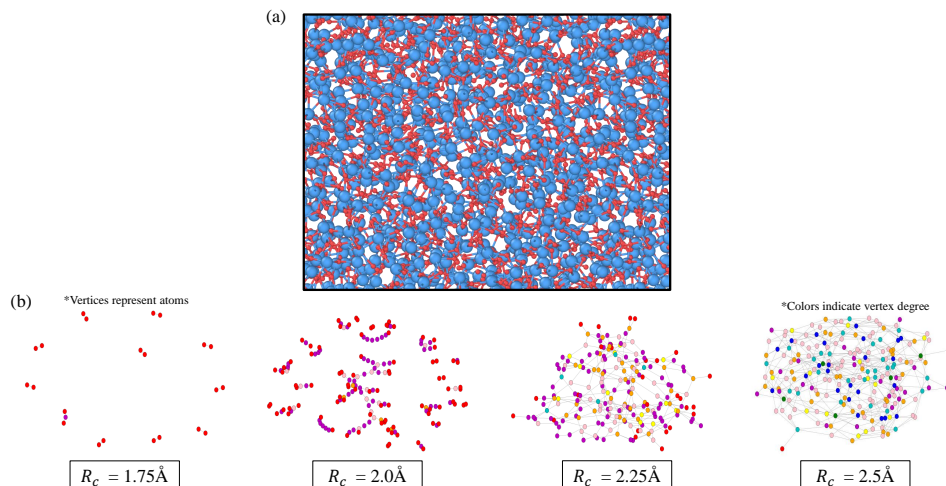
Figure S5: (a) Atomic-level structure of amorphous TiO_2 with red atoms representing oxygen and blue atoms representing titanium. (b) Graph coordination networks of (a) at various cutoff distances. Colors are relative to the smallest vertex degree within the graph, with purple representing the smallest degree and blue representing the largest degree. One can see that as the graph cutoff is increased, the GCN becomes increasingly interconnected and chaotic, providing a good indication of the local environments at various cutoff radii.

that the underlying network exists as a set of subgraphs, S, with s indexing a particular subgraph. Note that in the event a GCN is fully connected the outer sum disappears and no further changes are required to the formalism. D_s is the set of unique node degrees in a subgraph, with P_{d_m} being the probability of a given degree, d_m, occurring in the subgraph. A Vector Graph Order Parameter (VGOP) [5] is then assembled from a list of SGOP values, calculated using a cutoff radius set of $1.75\text{Å}, 2.0\text{Å}, 2.25\text{Å}, 2.5\text{Å}, 2.75\text{Å}, 3.0\text{Å}$, which were chosen based on the profile of the first Ti-O peak in the radial distribution function. Principal component analysis [6] (PCA) is used to reduce the number of features in the VGOP and allow for the visual inspection of the underlying data. Z-score normalization [7] is used to normalize the VGOPs to aid in the PCA decomposition. In this work the first two principal components comprised at least 95% of the underlying variance, and therefore the remaining components were discarded.

As discussed in the main text, the 1944-atom trajectories were used to explore the configuration space of the amorphous phase. Four stoichiometries were considered for the case of TiO_x ($x = 2.0$, 1.95, 1.9, 1.85), to gauge how the structural space changes as a function of oxygen concentration. Figure S6 shows the VGOP PCA decomposition for all cases considered in this work. The red points in Figure S6 correspond to the centroid of each cluster, which is calculated as the average x and y coordinate over all points in a given stoichiometry. The location of the red points corresponds to the black points shown in Figure 5 (b) in the main text. A subset of the total dataset was used in Figure S6 to give the reader a better idea of how the majority of each phase's points radially extend away from the cluster's centroid.

S2 DFT Binding Energies

S2.1 96-atom unit cell

Binding energy distributions are shown for the non-stoichiometric case of $TiO_{1.88}$ in Fig S7. Binding energy distributions are nearly identical between $TiO_{1.88}$ and TiO_2, indicating that the concentration of oxygen in the system plays little-to-no role in the hydrogen insertion energetics, at least for the cases examined in this work.
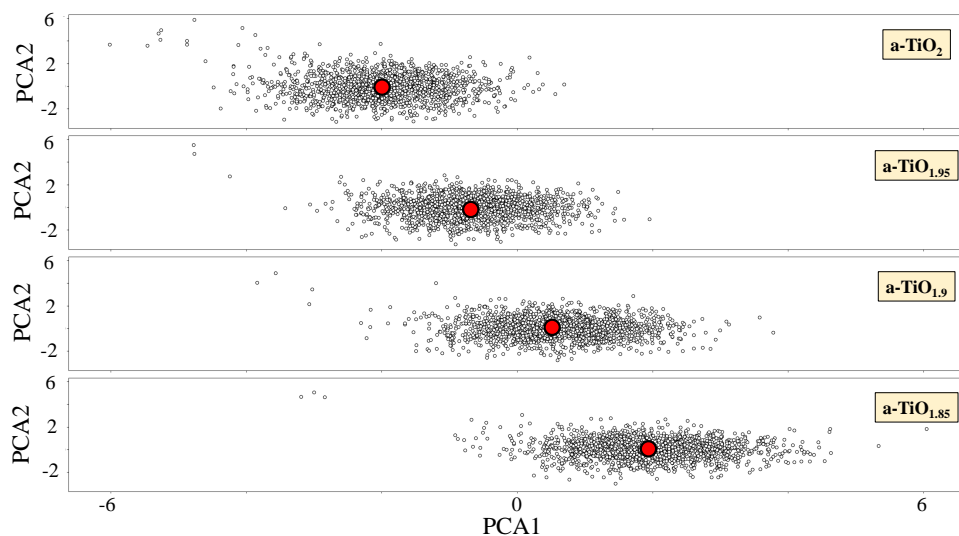
Figure S6: Decomposed PCA reduction for the amorphous phase space, as characterized using VGOP. Each row corresponds to a unique stoichiometry. Red points indicate the location each cluster's centroid. White points shown here represent approximately 25% of the total amount of data, which is reduced here for visualization purposes.
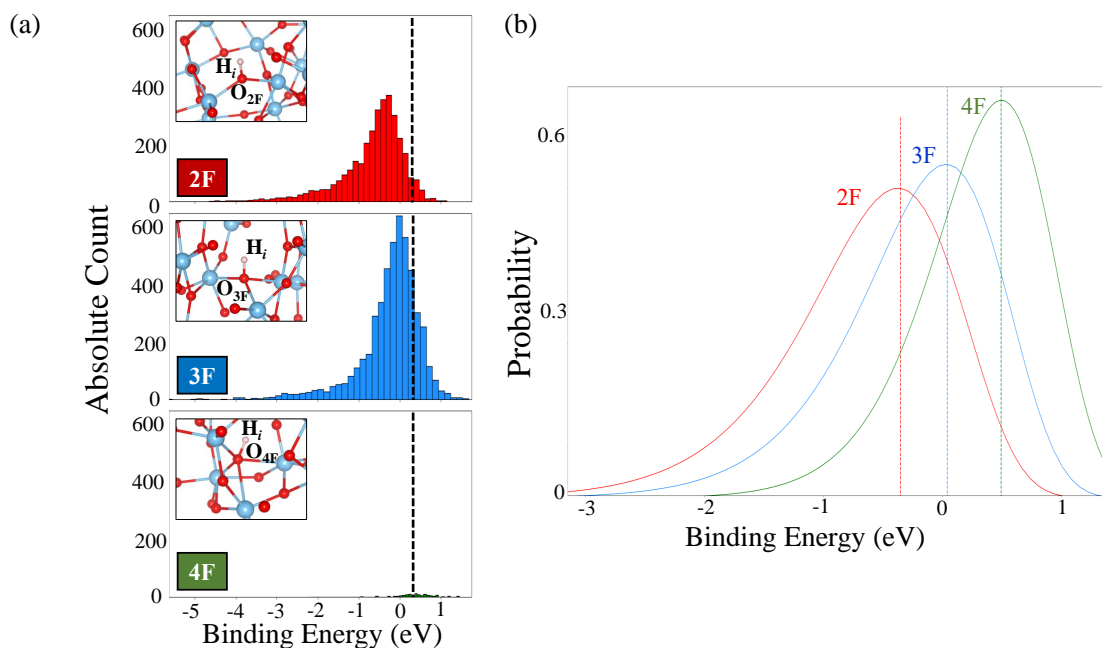

Figure S7: (a) Histograms of the hydrogen binding energies for the various CN environments. Values shown here are the absolute number of samples, signifying the significant reduction in data set size for 4F environments. Values are color coded based on the CN environment. The dashed vertical lines indicate the mean for the distributions. (b) Fitted hydrogen binding energies for $TiO_{1.88}$. The distributions are colored according to the coordination number of the particular oxygen site. Here, the x-axis represents the binding energy, while the y-axis represents the probability of that binding energy occurring with respect to the number of environments for that CN type. The dashed black line represents the hydrogen binding energy in crystalline rutile. Inserted images in (a) show the oxygen environments encountered by hydrogen in the amorphous $TiO_{1.88}$ phase space.
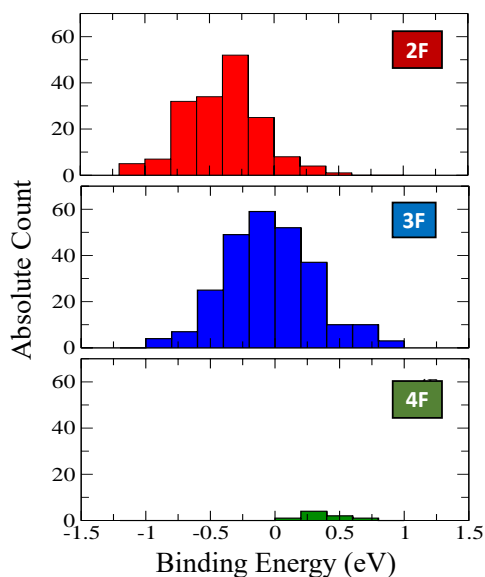
Figure S8: Histograms of the hydrogen binding energies for the various CN environments. Values shown here are the absolute number of samples, signifying the significant reduction in data set size for 4F environments. Values are color coded based on the CN environment. Here, the x-axis represents the binding energy.

S2.2 216-atom unit cell

Binding energy distributions are shown for the 216-atom unit cell systems of TiO$_2$ in Fig S8. 216-atom systems are taken from the DFT training data. While the number of binding energies represented here is significantly less than that of the 96-atom case, due to the cost of performing DFT on 216-atom unit cells, a trend does emerge for the 216-atom case in which binding energies less than -2 eV are not seen, while there exist many such configurations for 96-atom structures. While this may be due to the limited number of samples in the 216-atom case, it may be possible that such an effect is due to the hydrogen concentration, however we do not examine this in detail in this work.

S3 Experimental Characterization
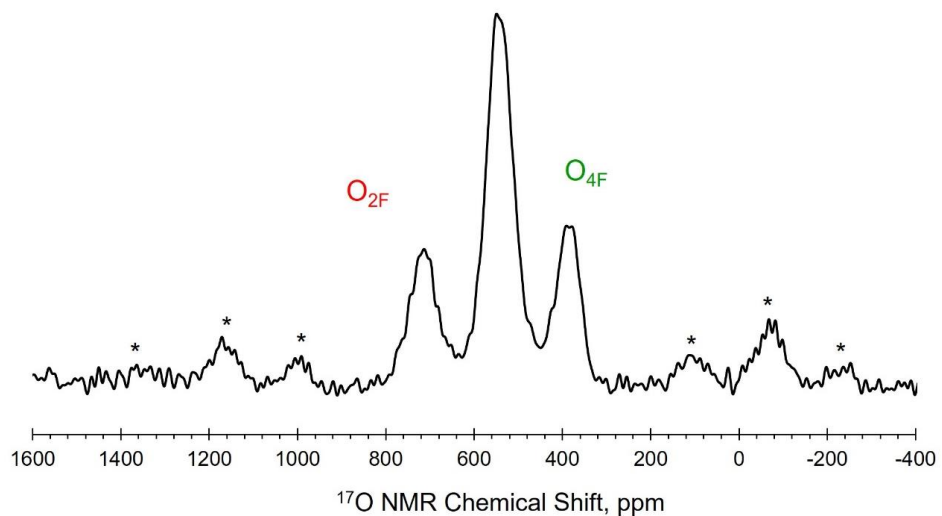
Figure S9: Experimental ^{17}O solid-state NMR of synthesized amorphous TiO$_2$ powder. The spectrum was referenced to tap water. The asterisks denote the sidebands. The synthesis procedures, phase, and stoichiometry characterizations for the powdered TiO$_2$ are given in supplementary information.
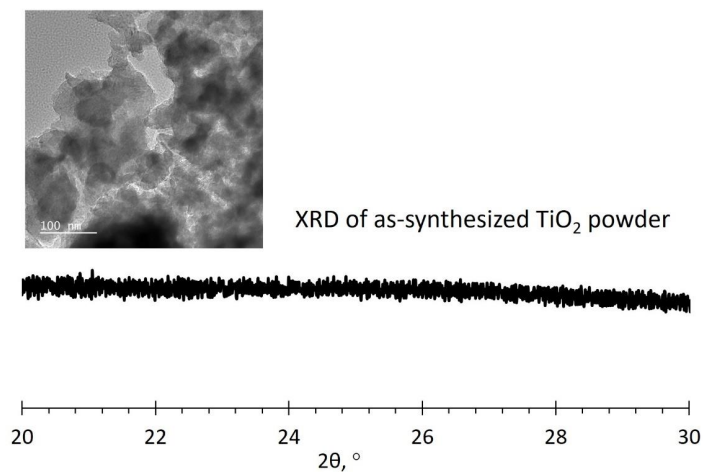

Figure S10: TEM imaging and XRD spectra of synthesized TiO$_2$ powder. The synthesized TiO$_2$ powders aggregates into nanometer scale clusters and are amorphous phase.
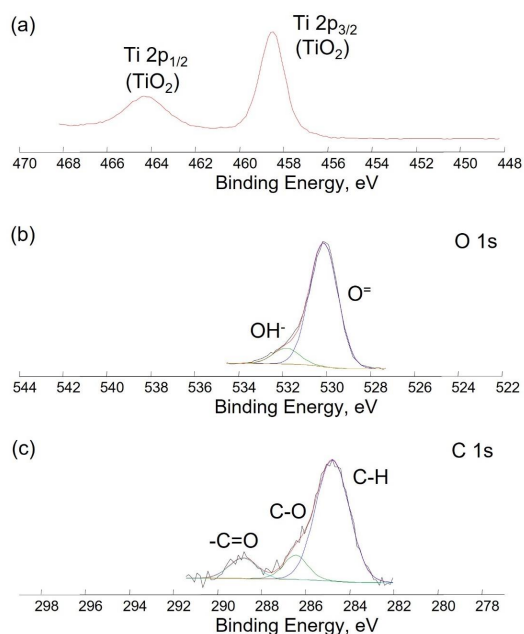
Figure S11: XPS spectra for the examination of chemical species and stoichiometry: (a) Ti $2p_{3/2}$ and Ti $2P_{1/2}$ peaks that corresponding to TiO_2. The stoichiometry of Ti oxide is 2 and is in align with the RBS results. (b) O 1s peak corresponding to OH- and O= groups. (c) C 1s peak corresponding to -C=O, C-O and C-H groups where were sourced from the precursor for TiO_2 synthesis. The atomic ratio of O/Ti is about 2.6, which is higher than 2 in TiO_2. The extra oxygen is contributed by the OH-, C-O, and -C=O groups.
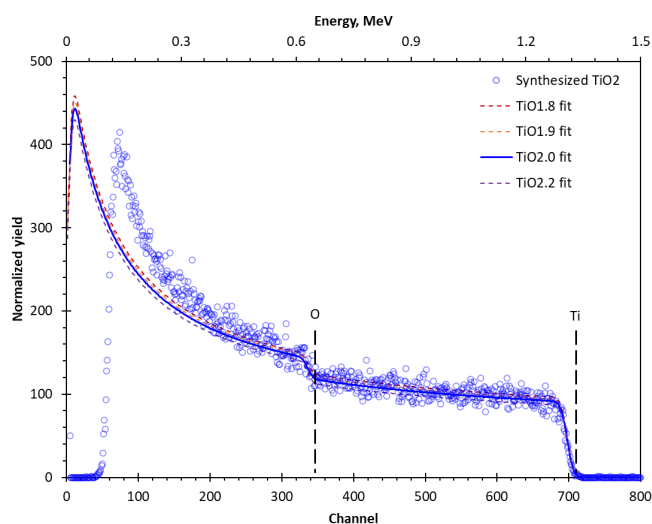
Figure S12: Rutherford backscattering spectra from TiO_x film. Symbols are experimental points, while solid lines are results of RUMP-code simulations. For clarity, only every 15th experimental point is depicted. The position of ^{16}O, and ^{48}Ti are peaks are marked by arrows. The best fit to the experimental data is obtained with O/Ti = 2.

Acknowledgements

References

1. Batra, R. *et al.* General Atomic Neighborhood Fingerprint for Machine Learning-Based Methods. *The Journal of Physical Chemistry C* **123,** 15859–15866. eprint: `https://doi.org/10.1021/acs.jpcc.9b03925`. `https://doi.org/10.1021/acs.jpcc.9b03925` (2019).

2. Jr., F. J. M. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46,** 68–78. eprint: `https://www.tandfonline.com/doi/pdf/10.1080/01621459.1951.10500769`. `https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769` (1951).

3. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Compu. Phys.* **117,** 1–19. ISSN: 0021-9991. `http://lammps.sandia.gov/` (1995).

4. Kärger, J. & Ruthven, D. M. Diffusion in nanoporous materials: fundamental principles, insights and challenges. *New J. Chem.* **40,** 4027–4048. `http://dx.doi.org/10.1039/C5NJ02836A` (5 2016).

5. Chapman, J., Goldman, N. & Wood, B. C. Efficient and universal characterization of atomic structures through a topological graph order parameter. *npj Computational Materials* **8,** 37. ISSN: 2057-3960. `https://doi.org/10.1038/s41524-022-00717-7` (Mar. 2022).

6. Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M. & Hooman, A. An Overview of Principal Component Analysis. *Journal of Signal and Information Processing* **04,** 173–175. ISSN: 2159-4465, 2159-4481. `http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jsip.2013.43B031` (2021) (2013).

7. Friedman, L. & Komogortsev, O. V. Assessment of the Effectiveness of Seven Biometric Feature Normalization Techniques. *IEEE Transactions on Information Forensics and Security* **14,** 2528–2536 (2019).