Tuning the Electronic Structure of Anatase Through Fluorination

Dario Corradini†, Damien Dambournet, and Mathieu Salanne†

Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8234, PHENIX, Paris, France

A highly fluorinated anatase lattice has been recently reported, providing a new class of materials whose general chemical formula is $\text{Ti}_1_{-x} \square_{x}X_{4-x}O_{2-4x}$ ($X^- = \text{F}^-$ or $\text{OH}^-$). To characterise the complex structural features of the material and the different $F$ environments, we here apply a computational screening procedure. After deriving a polarisable force-field from DFT simulations, we screen in a step-wise fashion a large number of possible configurations differing in the positioning of the titanium vacancies ($\square$) and of the fluorine atoms. At each step only 10% of the configurations are retained. At the end of the screening procedure, a configuration is selected and simulated using DFT-based molecular dynamics. This allows us to analyse the atomic structure of the material, which is strongly disordered, leading to a strong decrease (by 0.8 eV) of the band gap compared to conventional anatase.

Titanium dioxide, $\text{TiO}_2$, is a widely studied material. $\text{TiO}_2$ has in fact several promising applications, for example in the fields of photocatalysis, green chemistry and energy storage [1-8]. Naturally occurring polymorphs of $\text{TiO}_2$ include rutile, anatase and brookite. Recently, interest in the polymorphs of $\text{TiO}_2$ has been sparked in particular by their possible application as anodic materials in Li-ion batteries [9-14]. Fluorinated $\text{TiO}_2$ has also been investigated [15-20] since the presence of F in the compound might improve the sought characteristics of the material [17] or stabilise the highly reactive $\{001\}$ facets of the anatase crystal [19,21]. The nature of the fluorinated compound depends strongly on the fluorination technique employed [18,22]. So far, the stabilisation of fluorine within the anatase lattice of $\text{TiO}_2$ has been poorly understood, probably because of the structural complexity of the fluorinated material.

Pure anatase is a tetragonal crystal, with $c \approx 2.5a$, and its lattice is characterised by $\text{TiO}_6$ octahedral units. Recently, a novel synthesis technique conducted in our laboratory [23] has led to the preparation of a highly fluorinated anatase material in which fluoride or hydroxide anions replace the oxides in their lattice sites and the resulting charge deficiency is compensated by the formation of a cationic vacancy ($\square$) every four substitutions. The material obtained has thus the general formula $\text{Ti}_{1-x} \square_xX_{4-x}O_{2-4x}$, where $X^- = \text{F}^-$ or $\text{OH}^-$ (the amount of $F^-$ may vary depending on the synthesis conditions). Elemental analysis and synchrotron diffraction have revealed the existence of more than 20% cation vacancies. In fact the stoichiometric formula $\text{Ti}_{0.78} \square_{0.22} \text{X}_{0.88} \text{O}_{1.12}$ has been assigned to the most fluorinated composition of the material. By using $^{19}$F NMR spectroscopy, it has also been possible to discern three different coordination modes for the F atoms: $\text{F} - \text{Ti}_1 \square_2$, $\text{F} - \text{Ti}_2 \square_1$ and $\text{F} - \text{Ti}_3$, highlighting the complex structural arrangement present in the material.

Here we report the results of a computational study of the fully-fluorinated, hydroxide-free material (i.e. $\text{Ti}_{0.78} \square_{0.22} \text{F}_{0.88} \text{O}_{1.12}$) performed in order to better characterise its structural features and the effect of fluorination on the electronic structure. The enormous number of possible structural arrangements of the vacancies and of the F atoms in the anatase structure renders the problem untreatable directly by ab initio simulations. Therefore we apply a screening procedure on the possible configurations of the material, in the spirit of the emerging high-throughput techniques [24,25], by using classical Molecular Dynamics (MD). Several force-fields have been previously proposed for pure $\text{TiO}_2$ [26,33]. In this work, we use a polarisable force-field valid for the pure phase [31] as well as for the fluorinated material. We have extracted its parameters from Density Functional Theory (DFT) simulations, via a well-established force and dipole fitting procedure [35,36]. We have chosen to derive a new force-field instead of using an already available one for $\text{TiO}_2$. This is motivated by the fact that we want the force-field to be compatible with O to F substitutions, as well as with other oxide species, e.g. $\text{SiO}_2$, for future studies [34]. The details on the force-field employed are discussed in Supplementary Section S1, while an additional validation of the parameters involving fluorine atoms is presented in Supplementary Section S2.

In order to generate fluorinated samples starting from the pure $\text{TiO}_2$ anatase, we apply a screening procedure, similar in spirit to what done by Wilmer et al. for metal-organic frameworks [37] or by Coudert for zeolites [38]. At the fixed target composition $\text{Ti}_{0.78} \square_{0.22} \text{F}_{0.88} \text{O}_{1.12}$, we consider samples containing F in all possible environments $\text{F} - \text{Ti}_1 \square_2$, $\text{F} - \text{Ti}_2 \square_1$, and $\text{F} - \text{Ti}_3$, as suggested by NMR [23]. We leave the ratio of $\text{F}$ in the different environments free to vary at random. The starting fluorinated structures are generated from the $4 \times 4 \times 2$ pure anatase $\text{TiO}_2$ structure [39] ($\text{Ti}_{12}\text{O}_{25}$) leading to a system thus composed: $\text{Ti}_{100} \square_{28} \text{F}_{112} \text{O}_{144}$. We generate these configurations by erasing 28 Ti ions at random with no constraints on the creation of adjacent vacancies and...
we randomly substitute 112 O with 112 F. We impose that all F and O must be attached to at least one Ti. The screening procedure is then initiated. The protocol is as follows:

1) we perform single-point energy calculations on $\sim 1.5 \cdot 10^5$ configurations; we then retain the $\sim 1.5 \cdot 10^4$ configurations with the lowest energy for the following step.

2) we perform 0 K geometry optimisations of the atomic positions, keeping the length of the cell vectors fixed; we retain at maximum the $1.5 \cdot 10^3$ configurations with the lowest energy for the following step.

3) we perform 0 K cell optimisations of both the atomic positions and the lengths of the cell vectors, while keeping the box angles fixed at $\alpha = \beta = \gamma = \pi/2$; we retain at maximum the $1.5 \cdot 10^2$ configurations with the lowest energy for the following step.

4) we temper the configurations performing 10 ps $NVT$ runs at finite temperatures from $T_1 = 25$ K to $T_{12} = 300$ K, every $\Delta T = 25$ K. The 15 configurations with the lowest energy at $T_{12} = 300$ K are retained for the following step.

5) for the remaining samples, we perform a series of longer MD simulations at 300 K, first in the $NVT$ and then in the $NPT$ ensemble.

6) we then simulate the configuration with the lowest potential energy for 10 ps using using DFT–based molecular dynamics. We extract structural (bond length, fluorine environments) and electronic (density of states) characteristics of the material from this simulation.

Testing all the starting configurations in a generic, entirely ab initio based high-throughput procedure would be impossible. Generally, such studies involve static calculations only since performing ab initio MD simulations is computationally too expensive. Nevertheless, it is interesting to test whether our selected configurations, i.e., the 10 configurations remaining at the end of step 5) of the screening procedure would have also been selected if ab initio static calculations had been performed. To test this, we take their initial structures and perform a full DFT relaxation. Then we take the same number of random configurations from the starting pool of configurations. We find that the configurations given by the classical screening all have a lower final DFT energy than the ones taken at random. The results of this validation are shown in Supplementary Fig. S7.

Next, we analyse how the initial structural arrangements correlate with the energy of the configurations. The results are shown in Fig. 1. We see that the lowest energies (at 0 K) correlate with a higher fraction of F – Ti$_2$□$_1$. This is consistent with previous static DFT calculations performed on a system with only one vacancy and four O/F substitutions [29], which showed that having the F closest to the vacancy stabilises the structure. In Fig. 1 we also report the initial energies of the best configurations given at the end by the screening procedure. We observe that these final screening configurations are found closer to the average F speciation values rather than at the highest F – Ti$_2$□$_1$ relative compositions. Some of them have strongly been stabilised during the procedure, showing the importance of taking into account relaxation and thermal effects (see also Supplementary Fig. S8).

In order to compare the structural properties of the material as found in the experiments with our simulations, we plot together the experimental x-ray structure factor $S(k)$ and the one that we calculate from our trajectories using the Ashcroft–Langreth partial structure factors according to the formula:

$$S_{tot}(k) = \frac{\sum_{\alpha,\beta} \sqrt{x_\alpha x_\beta} f_\alpha(k) f_\beta(k) S_{\alpha\beta}(k)}{\sum_{\alpha} x_\alpha f_\alpha^2(k)}$$

(1)

where $\alpha, \beta = Ti, O, F$. $x_\alpha$ are the relative atomic concentrations of atoms of type $\alpha$, $S_{\alpha\beta}(k)$ are the partial structure factors calculated from the simulation trajectories using

$$S_{\alpha\beta}(k) = \langle \hat{\rho}_\alpha(k) \hat{\rho}_\beta(k) \rangle$$

(2)

where the dynamic variable $\hat{\rho}_\alpha(k)$ represents the Fourier
component of the atomic density of type $\alpha$ atoms at wave vector $\mathbf{k}$:

$$\hat{\rho}_\alpha(\mathbf{k}) = N_\alpha^{-1/2} \sum_{i=1}^{N_\alpha} \exp(\mathbf{i} \mathbf{k} \cdot \mathbf{r}_i)$$  \hspace{1cm} (3)

with $\mathbf{r}_i$ the position of atom $i$, and $N_\alpha$ the number of atoms of type $\alpha$ in the system. The angular brackets denote a thermal average, which was in practice evaluated as the time average over the whole simulation. Finally, $f_\alpha(k)$ are the $k$-dependent atomic x-ray scattering factors. They are calculated using the analytic approximation:

$$f_\alpha(k) = c_\alpha + \sum_{i=1}^4 a_{\alpha,i} \exp\left[ -b_{\alpha,i} \left( \frac{k}{4\pi} \right)^2 \right]$$  \hspace{1cm} (4)

where the coefficients $a_{\alpha,i}$, $b_{\alpha,i}$ and $c_\alpha$ are taken from Ref. 40 for O$^{2-}$ and from Ref. 11 for Ti$^{4+}$ and F$^-$. The structure factor calculated from the DFT-based molecular dynamics simulation performed on the final configuration is compared to the experimental signal in Fig. 2. The agreement between the two sets of data is good, taking into account that the experiments have been performed on nanoparticles, which leads to a strong broadening of the peaks, and that part of the fluoride ions are replaced by hydroxide groups. This may also affect the comparison, notwithstanding that F and OH are isoelectronic and thus their contribution to x-ray diffraction should not differ much if they occupy similar sites.

We also calculate the speciation of fluoride during the DFT-based molecular dynamics simulation. Contrarily to the initial configurations analysed in Fig. 1, the local relaxation of the F atoms (especially around the vacancies) leads to a wide distribution of Ti–F distances. It is therefore necessary to introduce a cutoff distance for assigning an environment to the F atoms. In Fig. 3 we show the time evolutions of the concentrations of $\mathrm{F} - \mathrm{Ti}_1\square_2$, $\mathrm{F} - \mathrm{Ti}_2\square_1$ and $\mathrm{F} - \mathrm{Ti}_3$ for a cutoff of 2.7 Å which corresponds to the first minimum of the Ti–F radial distribution function. We observe that after 2 ps of simulation, the concentrations equilibrate around average values of 13/66/21% for $\mathrm{F} - \mathrm{Ti}_1\square_2$, $\mathrm{F} - \mathrm{Ti}_2\square_1$ and $\mathrm{F} - \mathrm{Ti}_3$, respectively. This compares very well with the percentages measured by NMR in the experimental sample, i.e., 13/70/17% [23]. We can therefore conclude that the structure yielded by our screening procedure is realistic. This allows us to analyse it further in order to predict the material properties. We note that the $\mathrm{F} - \mathrm{Ti}_2\square_1$ average concentration is larger than the corresponding fraction in the initial pool of configurations as shown in Fig. 1, because the fluoride atoms positions relax around the titanium vacancies during the first 2 ps of the simulation. However, no strong lattice rearrangements are observed, as can be seen from Supplementary Fig. S9.

The electronic structure is of particular interest for many applications, since TiO$_2$-based materials are widely used in photocatalysis. We have therefore calculated the electronic density of states of fluorinated anatase on a series of snapshots extracted from our DFT-based molecular dynamics simulation, and compared it with the case of pure TiO$_2$ anatase. We have used the hybrid functional HSE06 [22, 23] for these calculations. In agreement with previous works [41], we see in Fig. 4 that the valence band edge of pure TiO$_2$ anatase is dominated by O 2p, and the conduction band edge is formed by Ti 3d. The band gap is much narrower in $\mathrm{F} - \mathrm{Ti}_1\square_2$, $\mathrm{F} - \mathrm{Ti}_2\square_1$, by 0.8 eV. Unlike the case of conventional doping with heteroatoms [45], the additional 2p states associated with

![Figure 2. Structure Factor](image)

**Figure 2. Structure Factor.** Comparison of the structure factor $S(k)$ at ambient conditions measured in experiments (red line) and calculated from a DFT-based molecular dynamics simulation performed on the configuration selected by the screening procedure (black line).

![Figure 3. Fluorine coordination number](image)

**Figure 3. Fluorine coordination number.** Evolution of the proportion of each coordination mode for the fluorine during the DFT–based MD simulation. The cutoff distance for defining Ti–F neighbour atoms is set to 2.7 Å.
fluoride ions do not locate at the top of the valence band, but rather at its bottom. The strong narrowing of the band gap is therefore due to the different structure of the material. In a previous study on TiO$_2$ nanocrystals, Chen et al. have shown that, due to the presence of structural disorder, their materials exhibit a band gap substantially smaller than the one of pure bulk materials [40]. It is very likely that similar effects are at play here. In conclusion, in order to characterise fluorinated anatase Ti$_{0.78}$O$_{1.12}$F$_{0.88}$, we have developed a screening procedure employing a polarisable force-field. It has allowed us (i) to select the best configurations starting from a very large pool (hundreds of thousands) of possible configurations; (ii) to reproduce the experimental structure, (iii) to study details of the partial atomic and electronic structure using DFT-based molecular dynamics. Our results show that fluorinated anatase has a highly disordered structure, which results in a lower band gap, by 0.8 eV, compared to conventional anatase. Therefore we conclude that fluorination appears as a very promising route for tuning material properties. This may be exploited for several applications, for example photocatalysis.

**METHODS**

We have performed the classical simulations using the software CP2K [17], using the Quickstep algorithm. We have used the GGA PBE [48] exchange-correlation functional and we have employed the DZVP-MOLOPT-SR-GTH basis set [49]. Moreover, we have used the Goedecker-Teter-Hutter [50] pseudopotentials; for Ti atoms, the electronic orbitals explicitly represented are 3$s^2$3$p^6$3$d^2$4$s^2$, for O atoms 2$s^2$2$p^4$ and for F atoms 2$s^2$2$p^5$. We have set a plane wave cut-off of 400 Ry. We have added dispersive interactions through the use of the DFT-D3 correction [51], with a cutoff radius of 30 Å. We have accumulated the trajectory for 10 ps, with the simulations time step being 0.5 fs. We have conducted the simulation in the NVT ensemble with a
target temperature of 300 K. We have calculated the electronic density of states on a series of snapshot extracted from the trajectory, using the HSE06 functional [12, 13].

[1] Chen, X. & Mao, S. S. Titanium Dioxide Nanomaterials: Synthesis, Properties, Modification and Applications. Chem. Rev. 107, 2891–2959 (2007).
[2] Kavan, L., Grätzel, M., Gilbert, S. E., Klemenz, C. & Scheel, H. J. Electrochemical and Photoelectrochemical Investigation of Single-Crystal anatase. J. Am. Chem. Soc. 118, 6716–6723 (1996).
[3] O’Regan, B. & Grätzel, M. A Low–Cost, High–Efficiency Solar Cell Based on Dye–Sensitized Colloidal TiO₂ Films. Nature 353, 737–740 (1991).
[4] Fujishima, A. & Honda, K. Electrochemical Photolysis of Water at a Semiconductor Electrode. Nature 238, 37–38 (1972).
[5] Hoffmann, M. R., Martin, S. T., Choi, W. & Babenmann, D. W. Environmental Applications of Semiconductor Photocatalysis. Chem. Rev. 95, 69–96 (1995).
[6] Kudo, A. & Miseki, Y. Heterogeneous Photocatalyst Materials for Water Splitting. Chem. Soc. Rev. 38, 253–278 (2008).
[7] Ravelli, D., Dondi, D., Fagnoni, M. & Albini, A. Photocatalysis: A Multi–Faceted Concept for Green Chemistry. Chem. Soc. Rev. 38, 2009–2111 (2009).
[8] Kamat, P. V. TiO₂ Nanostructures: Recent Physical Chemistry Advances. J. Phys. Chem. C 116, 11849–11851 (2012).
[9] Wagemaker, M., Borghols, W. J. H. & Mulder, F. M. Large Impact of Particle Size on Insertion Reactions. A Case for Anatase Li⁺TiO₂. J. Am. Chem. Soc. 129, 4323–4327 (2007).
[10] Yang, M.-C., Lee, Y.-Y., Xu, B., Powers, K. & Meng, Y. S. TiO₂ Flakes as Anode Materials for Li–Ion–Batteries. J. Power Sources 207, 166–172 (2012).
[11] Morgan, B. J. & Watson, G. W. Role of Lithium Ordering in the Li⁺TiO₂–Anatase → Titanate Phase Transition. J. Phys. Chem. Lett. 2, 1657–1661 (2011).
[12] Yildirim, H., Greeley, J. P. & Sankaranarayanan, S. K. R. S. Concentration–Dependent Ordering of Lithiated Amorphous TiO₂. J. Phys. Chem. C 117, 3834–3845 (2013).
[13] Morgan, B. J. & Madden, P. A. Lithium Intercalation into TiO₂(B): A Comparison of LDA, GGA, and GGA+U Density Functional Calculations. Phys. Rev. B 86, 035147/1–035147/13 (2012).
[14] Morgan, B. J. & Watson, G. W. GGA+U Description of Lithium Intercalation into Anatase TiO₂. Phys. Rev. B 82, 144119/1–144119/11 (2010).
[15] Tosoni, S., Lamiel-Garcia, O., Hevia, D. F. & Illas, F. Theoretical Study of Atomic Fluorine Diffusion through Bulk TiO₂ Polymorphs. J. Phys. Chem. C 117, 5855–5860 (2013).
[16] Tosoni, S., Lamiel-Garcia, O., Hevia, D. F., Miguel, J. & Illas, F. Electronic Structure of F-doped Bulk Rutile, Anatase, and Brookite Polymorphs of TiO₂. J. Phys. Chem. C 116, 12738–12746 (2012).
[17] Samadpour, M. et al. Fluorine Treatment of TiO₂ for Enhancing Quantum Dot Sensitized Solar Cell Performance. J. Phys. Chem. C 115, 14400–14407 (2011).
[18] Czoska, A. M. et al. The Nature of Defects in Fluorine–Doped TiO₂. J. Phys. Chem. C 112, 8951–8956 (2008).
[19] Wang, Y. et al. A Selective Etching Phenomenon on {001} Faceted Anatase Titanium Dioxide Single Crystal Surfaces by Hydrofluoric Acid. Chem. Commun. 47, 2829–2831 (2001).
[20] Wang, Y. et al. Nature of Visible–Light Responsive Fluorinated Titanium Dioxides. J. Mater. Chem. A 1, 12948–12953 (2013).
[21] Yang, H. G. et al. Anatase TiO₂ Single Crystals with a Large Percentage of Reactive Facets. Nature 453, 638–641 (2008).
[22] Yang, S. & Halliburton, L. E. Fluorine Donors and Ti³⁺ Ions in TiO₂ Crystals. Phys. Rev. B 81, 035204/1–035204/7 (2010).
[23] Li, W. et al. High Substitution Rate in TiO₂ Anatase Nanoparticles with Cationic Vacancies for Fast Lithium Storage. Under review (2015).
[24] Curtarolo, S. et al. The High–Throughput Highway to Computational Materials Design. Nat. Mater. 12, 191–201 (2013).
[25] Zakutayev, A. et al. Theoretical Prediction and Experimental Realization of New Stable Inorganic Materials Using the Inverse Design Approach. J. Am. Chem. Soc. 135, 10048–10054 (2013).
[26] Rondinelli, J. M. et al. Accelerating Functional Materials Discovery. Am. Ceram. Soc. Bull. 92, 14–21 (2013).
[27] Catlow, C. R. A., Freeman, C. M. & Royle, R. L. Recent Studies Using Static Simulation Techniques. Physica B 311, 1–12 (1995).
[28] Catlow, C. R. A. & James, R. Disorder in TiO₂ Insulators Having the Rutile Structure. Phys. Rev. B 32, 6773–6786 (1985).
[29] Matsu, M. & Akao, M. Molecular Dynamics Simulation of the Structural and Physical Properties of the Four Polytypes of TiO₂. Mol. Simulat. 6, 239–244 (1991).
[30] Sawatari, H., Iguchi, E. & Tilley, R. Formation Energies of Point Defects in Rutile TiO₂. J. Phys. Chem. Solids 43, 1147–1155 (1982).
[31] Post, J. E. & Burnham, C. W. Ionic Modeling of Mineral Structures and Energies in the Electron Gas Approximation: TiO₂ Polymorphs, Quartz Forsterite, Diopside. Amer. Miner. 71, 142–150 (1986).
[32] Han, X. J. et al. Polarizable Interatomic Force Field for TiO₂ Parametrized Using Density Functional Theory. Phys. Rev. B 81, 134108/1–134108/9 (2010).
[33] Corradini, D., Ishii, Y., Ohtori, N. & Salanne, M. DFT–based polarizable force field for TiO₂ and SiO₂. Model. Simul. Mater. Sci. Eng. xx, xxxx–xxxx (2015).
[34] Salanne, M. et al. Including Many–Body Effects in Models for Ionic Liquids. Theor. Chem. Acc. 131, 1143/1–1143/16 (2012).
[35] Tazi, S. et al. A Transferable Ab Initio Based Force Field for Aqueous Ions. J. Chem. Phys. 136, 114507/1–114507/12 (2012).
[37] Wilmer, C. E. et al. Large–Scale Screening of Hypothetical Metal–Organic Frameworks. *Nat. Chem.* 4, 83–89 (2012).

[38] Coudert, F.-X. Systematic Investigation of the Mechanical Properties of Pure Silica Zeolites: Stiffness, Anisotropy, and Negative Linear Compressibility. *Phys. Chem. Chem. Phys.* 15, 16012–16018 (2013).

[39] Horn, M., Schwerdtfeger, C. F. & Meagher, E. P. Refinement of the Structure of Anatase at Several Temperatures. *Z. Kristallogr.* 136, 273–281 (1972).

[40] Hemmati, M., Wilson, M. & Madden, P. A. Structure of Liquid Al2O3 from a Computer Simulation Model. *J. Phys. Chem. B* 103, 4023–4028 (1999).

[41] Wilson, A. J. C. editor, *International Tables for X-Ray Crystallography* (Kluwer Academic Publisher, Dordrecht, 1992).

[42] Heyd, J., Scuseria, G. E & Ernzerhof, M. Hybrid Functionals Based on a Screened Coulomb Potential. *J. Chem. Phys.* 118, 8207–8215 (2003).

[43] Heyd, J., Scuseria, G. E & Ernzerhof, M. Erratum: “Hybrid Functionals Based on a Screened Coulomb Potential” [J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.* 124, 219906 (2006).

[44] Scanlon, D. O. et al. Band Alignment of Rutile and Anatase TiO2. *Nat. Mater.* 12, 798–801 (2013).

[45] Harb, M., Sautet, P. & Raybaud, P. Origin of the Enhanced Visible-Light Absorption in N-Doped Bulk Anatase TiO2 from First-Principles Calculations. *J. Phys. Chem. C* 115, 19394–19404 (2011).

[46] Chen, X., Liu, L., Yu, P. & Mao, S. S. Increasing Solar Absorption for Photocatalysis with Black Hydrogenated Titanium Dioxide Nanocrystals. *Science* 331, 746–750 (2011).

[47] VandeVondele, J. et al. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves. *Comp. Phys. Commun.* 167, 103–128 (2005).

[48] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868 (1996).

[49] VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* 127, 114105/1–114105/9 (2007).

[50] Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* 54, 1703–1710 (1996).

[51] Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* 132, 154104/1–154104/19 (2010).

[52] Burbano, M., Nadin, S., Marrocchelli, D., Salanne M. & Watson, G. W. Ceria Co–doping: Synergistic or Average Effect? *Phys. Chem. Chem. Phys.* 16, 8230–8331 (2014).

[53] Marrocchelli, D., Madden, P. A., Norberg, S. T. & Hull, S. Structural Disorder in Doped Zirconias, Part II: Vacancy Ordering Effects and the Conductivity Maximum. *Chem. Mater.* 23, 1365–1373 (2011).

**ACKNOWLEDGEMENTS**

We thank François-Xavier Coudert for introducing us to the screening techniques. We thank Paul A. Madden, Benjamin J. Morgan and Benjamin Rotenberg for discussions. The research leading to these results has received funding from the the European Union through the FP7–framework (FLUOSYNES, Contract PCI–GA–2012–321879). We also thank Karen W. Chapman for providing the experimental S(k) data. The work done at the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

**AUTHOR CONTRIBUTIONS**

D.C. and M.S. have designed research. D.C. has implemented the screening procedure and has performed most of the simulations. M.S. has conducted the band gap calculations. D.C. and M.S. have written the manuscript and prepared the figures. D.D. has provided the experimental input. All the authors have participated in the discussions and reviewed the manuscript.

**SUPPLEMENTARY INFORMATION**

In Supplementary Section S1, we describe the analytic form of the classical force–field employed to perform the classical MD simulations and we tabulate its parameters (see Supplementary Tables S1 and S2). In Supplementary Section S2, we compare the results reproduced by our classical force–field to DFT–calculated quantities. We first compare the results for the relative stability of structures containing only one vacancy and differing by the positioning of the F atoms (see Supplementary Fig. S6). Then we compute the DFT–energies (before and after relaxation) of the structures found by our screening procedure and we compare them to the DFT–energies of an equal number of structures selected at random from the pool of starting configurations (see Supplementary Fig. S7). Supplementary Section 3 deals with the energy distributions at the different screening steps. Fig. S8 shows the distribution of the energies of the configurations tested and retained at steps 1) to 4) of the screening procedure and Fig. S9 provides the positions of all the atoms along the DFT–based MD simulation. In Supplementary Section 4, the density of state for 10 different configurations extracted from this trajectory are shown in Fig. S10. Finally, an example CP2K input is provided in Supplementary Section S5.
S1. CLASSICAL MODEL

We describe the interaction potential between the ions by a classical polarisable force–field whose parameters we derive from \textit{ab initio} DFT simulations. Details on how to extract the parameters of the classical force–field from DFT simulations are reported in earlier works (see Ref. [34–36] in the main text). In particular, the detailed procedure used for the pure phases of TiO\textsubscript{2} together with its validation are reported in Ref. [34] (of the main text) and are not repeated here. The force–field parameters for the fluoride ions are obtained in an analogous fashion. We do report here in the following the analytic form of the classical force–field that we use, together with its parameters.

A. Polarisable ion model (PIM)

The repulsive and dispersive terms of the interactions are taken into account using the the Born–Mayer–Huggins (BMH) form of the interaction potential:

\[
V_{BMHF} = \sum_{i,j>i} A_{ij} e^{-B_{ij}r_{ij}} - f_s^{ij}(r_{ij}) \frac{C_{ij}^{ij}}{r_{ij}^8} - f_8^{ij}(r_{ij}) \frac{C_8^{ij}}{r_{ij}^8}. \tag{S5}
\]

The damping functions are Tang-Toennies functions of the form

\[
f_{ij}^{\alpha}(r_{ij}) = 1 - e^{-b_{ij}r_{ij}} \sum_{k=0}^{\infty} \frac{(b_{ij}r_{ij})^k}{k!}. \tag{S6}
\]

When performing molecular dynamics simulations, we add a Gaussian term in the Ti–O and Ti–F interactions that acts as a steep repulsive wall and accounts for the oxide/fluoride anion hard core:

\[
V_{Gaussian} = \sum_{i \in O, j \in Ti} B_{ij} e^{-d_{ij}r_{ij}^2}. \tag{S7}
\]

This extra term is used in cases where the ions are strongly polarised to avoid instability problems at very small anion–cation separations.

For the Coulombic part of the interaction potential,

\[
V_{\text{Coulomb}} = \sum_{i,j>i} \frac{q_iq_j}{r_{ij}}, \tag{S8}
\]

the formal charges for the ionic species are used, \(-2e\) for O ions, \(-e\) for F ions and \(+4e\) for Ti ions. The many-body electrostatic interactions are described by the induced dipoles \(\mu_i\), obtained at each MD step minimising the polarisation energy

\[
V_{\text{pol}} = \sum_{i} \frac{1}{2\alpha_i} |\mu_i|^2 + \sum_{i,j>i} \left[ (g_{ij}^{\alpha}g_{ij}^{\beta}(r_{ij}) - q_i^\alpha q_j^\beta(r_{ij})) T_{ij}^{\alpha\beta} - \mu_i^\alpha \mu_j^\beta T_{ij}^{\alpha\beta} \right] \tag{S9}
\]

where the Einstein summation convention is assumed, \(\alpha_i\) is the atomic polarisability and \(T\) are the multipole interaction tensors. The damping function \(g_{ij}(r_{ij})\) is of the Tang-Toennies form

\[
g_{ij}(r_{ij}) = 1 - c_{ij} e^{-b_{ij}r_{ij}} \sum_{k=0}^{4} \frac{(b_{ij}r_{ij})^k}{k!}. \tag{S10}
\]

B. Parameterisation

The repulsion and polarisation parameters of the force–field have been fitted in order to reproduce the forces and dipoles extracted from DFT calculations, using a well-established procedure, see Ref. [34] in the main text. In the present case, we obtain final \(\chi^2\) values of 0.16 and 0.37 for the fits of dipoles and forces, respectively. Such values are similar to the ones obtained in our recent work on other oxide materials, i.e. rare earth doped ceria, for example [52]. The dispersion interactions are not taken into account in a proper way in the DFT calculations we have performed. The corresponding parameters have not been fitted, instead they have been taken from our previous works on fluorides and oxides, see Ref. [35] in the main text. Finally, the Gaussian term parameters have been chosen following Marrocchelli \textit{et al.} [53]. The obtained parameters are reported in Table S1 for the BMH part of the force–field and in Table S2 for the polarisation part.

S2. COMPARISON BETWEEN \textit{AB INITIO} AND CLASSICAL SIMULATIONS

We assess the behaviour of our force–field in the fluorinated samples, by comparing the results of classical and DFT calculations. For simplicity, we consider the case of one single fluorination, as in Ref. [23] in the main text. We consider here the system Ti\textsubscript{127-1}F\textsubscript{1}O\textsubscript{252} and we distribute the F atoms either at random positions in the lattice (where they substitute O atoms) or at positions neighbouring the cationic vacancy. The latter correspond to 2–coordinated F, i.e. F–Ti\textsubscript{2} environment. We consider the cases where 0, 1, 2, 3 or 4 F are neighbouring the vacancy and have there a F–Ti\textsubscript{2} environment, with the remaining F having the F–Ti\textsubscript{3} environment. We compare the energy calculated in 0 K cell optimisations, using either DFT or our force field. The comparison for the energies is shown in Fig. S6. We see that our classic force-field is able to closely reproduce the decrease in energy with the
increase in the number of $F - Ti_2\square_1$, observed by DFT calculations (see also Ref. [23] in the main text).

Figure S6. Comparison between DFT and classical potential. Difference in energy between the case where all F atoms in $Ti_{127}\square_1F_4O_{252}$ have environment $F - Ti_3$ and the cases where F is progressively added in positions neighbouring the vacancies and has thus $F - Ti_2\square_1$ environment. We have calculated this quantity by DFT (circles) and by our classic force-field (squares).

Finally in Fig. S7 we show the comparison between the DFT energies of the configurations selected by the screening procedures and the same number of configurations taken at random from the initial pool of configurations of the $Ti_{100}\square_2F_{112}O_{144}$, before (panel a) and after (panel b) relaxation of the atomic positions and of the cell dimensions.

Figure S7. DFT energies of the configurations selected by the screening. For the configurations left at the end of the screening procedure, we take their initial structures and calculate their DFT energy (black circles) before (a) and after relaxation (b). Those are compared with the DFT energies of the same number of configurations taken at random (red squares) from the initial pools of configuration. The values of the energies are plotted relative to the lowest DFT energy configuration. Panel (a) shows the results obtained before the DFT relaxation, panel (b) shows the results obtained after the DFT relaxation.
Table S1. BMH parameters. Parameters of the BMH potential extracted from DFT simulations.

| Atom Pair | \( A_{ij} \) (Ha) | \( a_{ij} \) (\( \text{Å}^{-1} \)) | \( B_{ij} \) (Ha) | \( d_{ij} \) (\( \text{Å}^{-2} \)) | \( C_{6}^{ij} \) (Ha \( \text{Å}^6 \)) | \( C_{8}^{ij} \) (Ha \( \text{Å}^8 \)) | \( b_{ij}^{D} \) (\( \text{Å}^{-1} \)) |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| O–O       | 290.4          | 4.54668        | –              | –              | 0.48309        | 2.61949        | 2.64562        |
| O–F       | 278.4          | 4.71487        | –              | –              | 0.39890        | 1.55438        | 3.11805        |
| O–Ti      | 43.0           | 2.86431        | 50,000         | 6.4279         | –              | –              | –              |
| F–F       | 282.3          | 4.61849        | –              | –              | 0.32938        | 0.922357       | 3.59048        |
| F–Ti      | 28.3           | 3.13082        | 50,000         | 6.4279         | –              | –              | –              |
| Ti–Ti     | 1.0            | 9.44863        | –              | –              | –              | –              | –              |

Table S2. Polarisation parameters. Parameters of the polarisation part of the interaction potential extracted from DFT simulations.

| Atom / Atom Pair | \( \alpha \) (\( \text{Å}^3 \)) | \( b_{ij} \) (\( \text{Å}^{-1} \)) | \( c_{ij} \) |
|-----------------|-----------------|-----------------|-------------|
| O               | 1.59150         |                 |             |
| O–O             |                 | 4.74888         | 2.227       |
| O–F             |                 | –               | –           |
| O–Ti            |                 | 3.90122         | 2.13327     |
| F               | 1.16458         |                 |             |
| F–O             |                 | –               | –           |
| F–F             |                 | –               | –           |
| F–Ti            |                 | 4.16887         | 2.90678     |
| Ti              | 0.20442         |                 |             |
| Ti–O            |                 | 3.90122         | -1.90330    |
| Ti–F            |                 | 4.16887         | -2.66057    |
| Ti–Ti           |                 | –               | –           |
S3. SCREENING PROCEDURE

Figure S8. Energy distributions at screening steps 1) to 4). The panels on the left show the relative frequency histograms of the energies calculated for all the configurations tested at each step, while the panels on the right show the sorted energies of the configurations retained after each step. The energies shown are: (a,b) at 0 K for the starting unrelaxed structures; (c,d) at 0 K after the optimization of the atomic positions; (e,f) at 0 K after the optimization of the atomic positions and cell vector lengths; (g,h) at 300 K after the tempering from 25 to 300 K. The open circles in panels (b,d,f,h) indicate the energy of the configurations left at the end of the screening procedure at each previous step. Note that the number of retained configurations can be less than the target one (see main text). This is due to “crashed” unstable configurations that are eliminated from the pool.

Figure S9. Positions of the atoms along the DFT–based MD simulation. The positions of the Ti (blue), O (red) and F (green) atoms are shown every 10 steps of the simulation. Although structural relaxation close to the vacancies is observed, there is no major lattice rearrangement.
S4. Density of States for several Ti$_{0.78}$O$_{0.22}$F$_{0.88}$ configurations. The density of states has been calculated for 10 configurations extracted from the DFT-based molecular dynamics trajectory. The results are shown as thin black lines while their average is shown using a thick red line.

S5. CP2K INPUT FILE

We here report an example of CP2K input file, prepared for running the system Ti$_{127}$O$_{252}$F$_4$ in the NVT ensemble at $T = 300$ K, using our classical force-field.
