Journal of Materials Chemistry A

PAPER

Cite this: J. Mater. Chem. A, 2024, 12, 29222

p–d coupling: prerequisite for band-like doping levels in metal oxides†

Kangyu Zhang,[ab] Li-Chang Yin, *[ab] Guoqiang Deng,[ab] Xing-Qiu Chen,[ab] Hui-Ming Cheng [ac] and Gang Liu *[ab]

Received 9th August 2024
Accepted 24th September 2024

DOI: 10.1039/d4ta05577b

rsc.li/materials-a

Identifying the prerequisite of inducing band-like doping levels in wide bandgap metal oxides is a crucial yet open question. Herein, taking boron (B) and nitrogen (N) codoped anatase TiO_2 as an example and combining density functional theory calculation with machine learning, it has been revealed that the band-like doping levels mainly originate from a strong p–d coupling between O/N-p_π and Ti-t_{2g} orbitals. Significantly, the existence of strong p–d coupling is an intrinsic characteristic of anatase TiO_2 determined entirely by crystal symmetry, which can be used as a universal criterion to predict whether band-like doping levels can be induced in other metal oxides, and the criterion has been fully verified on rutile and brookite TiO_2, $ScTaO_4$, WO_3, SnO_2 and $MgTa_2O_6$. Besides, the strong p–d coupling offers a theoretical basis for the long-held empirical understanding that uniform doping is important in achieving stronger visible light absorption for wide bandgap metal oxide photocatalysts. Overall, the uncovered strong p–d coupling provides a simple yet profound guideline for bandgap engineering of metal oxides.

Introduction

Narrowing the bandgap of wide bandgap metal oxides is highly desirable for many metal oxide related applications.[1–3] For instance, it is crucial for harvesting light over a broader spectrum and achieving suitable band edge positions for photocatalysis[4] and optoelectronic applications.[5] A widely used approach for bandgap narrowing is non-metal doping,[6,7] particularly nitrogen (N) doping,[8–10] since N doping induces doping levels above the valence band maximum (VBM) of metal oxides. Ideally, band-like doping levels with strong dispersion and significant overlap with the VBM are indispensable for band-to-band redshift of the absorption edge[8] as well as fast charge carrier transport[11] of metal oxides. However, the N doping levels are normally localized within the bandgap due to typically low N doping concentrations.[12–17] Increasing the N doping concentrations may only result in multiple localized doping levels with different energy,[18–20] meanwhile aggravating the formation of charge trapping centers such as oxygen vacancy.[13,21–23]

Nevertheless, the emergence of red TiO_2 has gone beyond this traditional understanding.[24,25] Specifically, by gradient boron (B) and nitrogen (N) codoping, the first red anatase TiO_2 with a bandgap value of about 2 eV was obtained,[24] which exhibits a band-to-band redshift of the light absorption edge compared with pristine anatase TiO_2. Advancing from gradient to uniform doping,[25] the photocatalytic oxygen evolution activity of red anatase TiO_2 has been substantially increased, implying a much higher hole mobility contributed by band-like doping levels. Considering the fact that the atomic ratio of N to Ti is only about 5% in red anatase TiO_2,[25] it is much lower than the N content in oxynitrides. However, the bandgap value of anatase TiO_2 can still be substantially narrowed from 3.2 to about 2 eV, which is similar to that of oxynitrides.[26] This suggests the existence of certain intrinsic character allowing for band-like doping levels at a low N-doping concentration in anatase TiO_2.

Elucidating this intrinsic character necessitates the identification of characteristic B/N-codoping configurations featuring band-like doping levels, which are likely to exhibit the most significant bandgap narrowing among all configurations due to the large energy level broadening of band-like doping levels. Although the bandgap can also be narrowed by localized doping levels related to direct bonding between N dopants,[27] this can be easily excluded by scrutinizing the band structures. Therefore, calculating the bandgap values for all possible B/N-codoping configurations in anatase TiO_2 is the precondition to

[a]Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, 72 Wenhua Road, Shenyang 110016, China. E-mail: lcyin@imr.ac.cn; gangliu@imr.ac.cn

[b]School of Materials Science and Engineering, University of Science and Technology of China, 72 Wenhua Road, Shenyang 110016, China

[c]Institute of Technology for Carbon Neutrality, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Blvd, Shenzhen 518055, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ta05577b

efficiently and accurately identify those configurations featuring band-like doping levels. However, to reproduce the experimental doping level (N : Ti ≈ 5 at%)[24] as closely as possible and maintain the charge balance, one needs to substitute three lattice O atoms with three N atoms and introduce one additional interstitial B atom in a 3 × 3 × 1 anatase TiO_2 supercell (N : Ti = 8.33 at%), resulting in a total of 8144 symmetrically inequivalent B/N-codoping configurations. This makes the calculation of bandgap values for all these configurations far beyond the capability of conventional density functional theory (DFT) calculations. Thus, previous studies have only been able to consider a few specific configurations based on researchers' individual intuition and/or expertise.[24,28,29] Consequently, the intrinsic character of anatase TiO_2 allowing band-like doping levels has remained elusive over the past two decades.

Different from conventional DFT calculations, the machine learning (ML)[30–35] methods offer a promising solution to accurately and efficiently predict the bandgap values of all B/N-codoping configurations in anatase TiO_2. Since applying the ML methods requires each B/N-codoping configuration to be represented by a fixed-length array, termed structure descriptor, in this work, we first proposed an accurate structure descriptor, and then trained a highly accurate ML model for predicting the bandgap values of all B/N-codoping configurations in anatase TiO_2. Based on the ML predictions, B/N-codoping configurations with band-like doping levels were identified, and thorough analysis revealed that the existence of strong p–d coupling between O/N-p_π and Ti-t_{2g} orbitals in anatase TiO_2 is essential for inducing band-like doping levels. Since the existence of strong p–d coupling is entirely determined by crystal symmetry, we further generalized it as a universal criterion to predict whether band-like doping levels can be induced in other metal oxides, and the criterion has been fully verified on rutile and brookite TiO_2, $ScTaO_4$, WO_3, SnO_2 and $MgTa_2O_6$.

Results

The application of ML methods requires each B/N-codoping configuration in anatase TiO_2 to be represented by a structure descriptor, and the accuracy of ML prediction is largely determined by whether the structure descriptors can accurately differentiate B/N-codoping configurations. Therefore, we first proposed a novel approach for structure descriptor construction based on a concept of charge transfer path.[36] As detailed in the ESI (Fig. S1),† each B/N-codoping configuration in anatase TiO_2 can be represented by a 168-dimensional structure descriptor, and in combination with DFT calculations, a highly accurate XGBoost regression model[37] has been successfully trained to predict the bandgap values of all B/N-codoping configurations in anatase TiO_2 (Fig. S2†).

Based on the ML prediction, 20 B/N-codoping configurations with the most significant bandgap narrowing have been identified, which exhibit distinctive periodic features of dopant spatial orderings. Specifically, all three N dopants are located on the same (010) atomic plane and exhibit two characteristic spatial orderings, while the interstitial B dopant can occupy any interstitial site on either the same (highlighted yellow) or the third nearest (010) atomic plane (highlighted purple), as depicted in Fig. 1a and b and S5.† Notably, the band structures of all these 20 B/N-codoping configurations are essentially the same, where band-like doping levels are obtained through level crossing between N-p_σ (p_x/p_z) and N-p_π (p_y) doping levels, as shown in Fig. S6.† Such band-like doping levels can be further confirmed by band structure calculations at the HSE06 level[38–40] for two arbitrarily selected B/N-codoping configurations (configID-809 and configID-862) among the 20 identified ones, as shown in Fig. 1c and d.

It has also been found that when the B dopant occupies an interstitial site outside the yellow/purple (010) atomic planes found in the 20 identified configurations and the N dopant spatial ordering remains unchanged, which corresponds to another 20 configurations depicted in Fig. S7,† the level crossing between N-p_σ and N-p_π doping levels disappears, resulting in multiple localized doping levels, as shown in Fig. S8.† A closer analysis revealed that when the B dopant is located on the yellow/purple (010) atomic planes, a mirror symmetry $\sigma(xz)$ exists in the 3 × 3 × 1 anatase TiO_2 supercell, while it is absent when the B dopant is located elsewhere. Consequently, the symmetries of N-p_σ (p_x/p_z) and N-p_π (p_y) orbitals are different in the first case because of the presence of $\sigma(xz)$, but are identical in the second case due to the absence of $\sigma(xz)$. According to group theory,[41] when energy levels with different or the same symmetry approach one another, level crossing or level anti-crossing can be expected. This is the reason why level crossing can be observed in the 20 identified configurations (Fig. S6†), while level anti-crossing, giving rise to multiple localized doping levels, has been observed in another 20 configurations (Fig. S8†).

It is important to note that the bandgap narrowing achieved through level crossing between N-p_σ/p_π doping levels in the 20 identified B/N-codoping configurations is more significant (~1.0 eV at the PBE level) than in previously reported N-doped anatase TiO_2 (typically less than 0.5 eV at the PBE level[18,42]). This indicates that the interstitial B dopant can remarkably affect the energies of N doping levels to achieve a more significant bandgap narrowing. To clarify this effect, we took one identified configuration (configID-809) as an example, and compared its band/atomic structures with those of the corresponding N-doping configuration without the interstitial B dopant. As shown in Fig. 2a and d, the interstitial B dopant causes noticeable downshifting/upshifting of two N-p_σ doping levels, labeled as N-$p_\sigma(1)$ and N-$p_\sigma(2)$, and the upshifting of N-$p_\sigma(2)$ results in the more significant bandgap narrowing than the case without the interstitial B dopant. By comparing the atomic structures and charge density isosurface plots shown in Fig. 2b and e, the interstitial B dopant substantially shortens the bond length between the Ti atom labeled as Ti3 and its adjacent N dopant, thereby strengthening the corresponding p_σ bond. This not only results in the downshifting of N-$p_\sigma(1)$, but also the upshifting of N-$p_\sigma(2)$ as the repulsion between N-$p_\sigma(1)$ and N-$p_\sigma(2)$ is substantially increased.

In order to uncover the reason for the Ti3–N bond shortening, we further analyzed the projected density of states (PDOS). As shown in Fig. 2c and f, the Ti3–N bond shortening

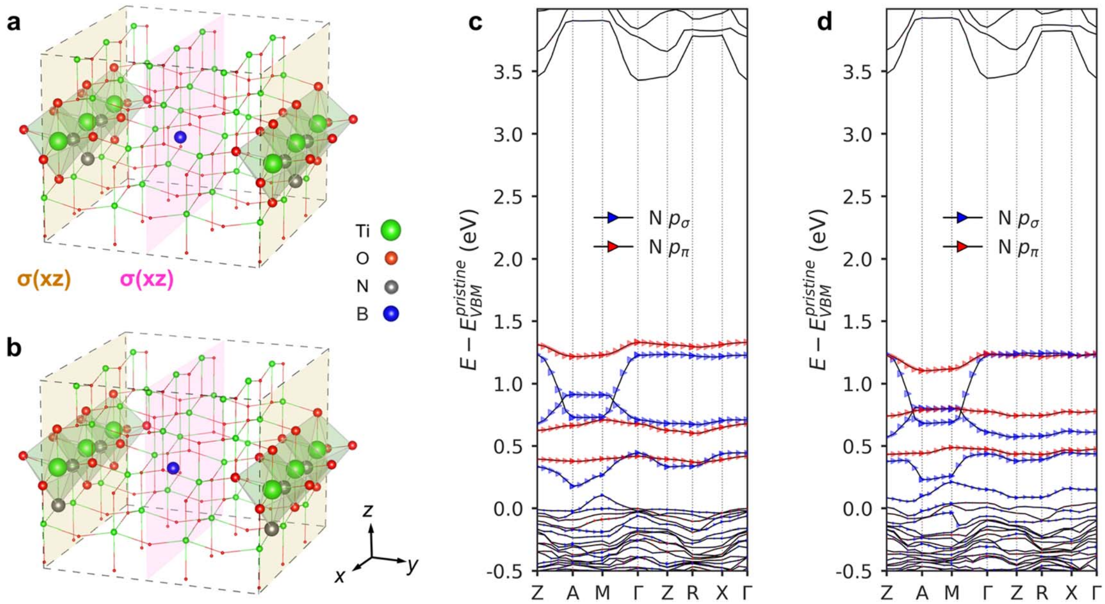

Fig. 1 Characteristics of the identified B/N-codoping configurations in anatase TiO_2. (a and b) Atomic structures of two arbitrarily selected B/N-codoping configurations (configID-809 and configID-862) from the 20 identified ones. (c and d) Band structures at the HSE06 level of the two B/N-codoping configurations illustrated in (a) and (b), respectively. The energy values are referenced to the VBM of pristine anatase TiO_2, which is set to be 0.0 eV.

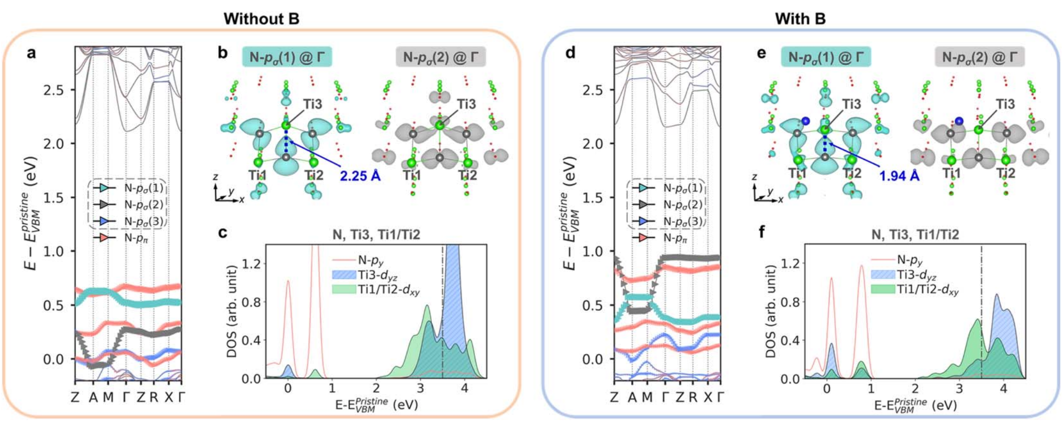

Fig. 2 Mechanism of the significant bandgap narrowing promoted by the interstitial B dopant. (a and d) Band structures at the PBE level of one identified B/N-codoping configuration (configID-809) without and with the interstitial B dopant. (b and e) Charge density isosurface plots (0.005 e $Å^{-3}$) of two N-p_σ doping levels labeled as N-p_σ(1) and N-p_σ(2) at Γ point. (c and f) Partial density of states (PDOS) of t_{2g} orbitals (d_{yz} and d_{xy}) of three Ti atoms labeled as Ti1, Ti2 and Ti3, and the PDOS of the p_y orbital of the N atom adjacent to the three Ti atoms. A vertical dashed line is added to highlight the energy shift in the antibonding region. The energy values are referenced to the VBM of pristine anatase TiO_2, which is set to be 0.0 eV.

can be attributed to the markedly enhanced p–d coupling between N-p_y (p_π) and Ti3-d_{yz} (t_{2g}) by introducing the interstitial B dopant. This is evident from the more pronounced hybridization peak of these two orbitals and the increased energy of Ti3-d_{yz} in the antibonding region.[43] In contrast, the p–d coupling between N-p_y and Ti-d_{xy} is much weaker as the hybridization peak of the two orbitals is barely changed by introducing the interstitial B dopant, indicating that the extra electron contributed by the interstitial B dopant for compensating the N^{3-}/O^{2-} charge imbalance[24,28] preferentially occupies the molecular orbital formed between N-p_y and Ti-d_{yz}, rather than the one between N-p_y and Ti-d_{xy}. This selective occupation leads to the shortening of the Ti3–N bond, thereby achieving the significant bandgap narrowing in the identified B/N-codoping configuration.

In order to understand the origin of the much stronger p–d coupling between p_y and d_{yz} than that between p_y and d_{xy}, one needs to move from a molecular picture to the crystal. Due to the translational symmetry of a crystal, an atomic orbital, for instance p_y, at position r should be expressed as a Bloch function $\psi = e^{ikR} p_y$, where R denotes a lattice vector and k represents a point in the reciprocal space. In the Bloch function, the k-

point determines the phase change of the atomic orbital upon translation between unit cells, and each *k*-point transforms under a specific point group due to the symmetry of the crystal.[41] To be specific, low-symmetry *k*-points, *i.e.*, those whose point group has fewer symmetry elements, result in the atomic orbital having amplitudes smaller than the maximum value in some unit cells, while high-symmetry *k*-points ensure that the atomic orbital has the maximum amplitudes in each unit cell.[44] Therefore, given a high-symmetry *k*-point, if two atomic orbitals, for instance p_y and d_{yz}, transform like the same irreducible representations of its associated point group,[45] the high-symmetry *k*-point generates Bloch function $\psi = e^{ikR}(p_y + d_{yz})$, which exhibits strong coupling between the two atomic orbitals within each unit cell.

Based on the above analysis, the much stronger p–d coupling between p_y and d_{yz} observed in the 20 identified B/N-codoping configurations indicates the existence of high-symmetry *k*-points that allow a strong p–d coupling in anatase TiO_2. In order to validate this speculation, we summarized the point groups of all high-symmetry *k*-points in the first Brillouin zone of anatase TiO_2 in Table S1.† From the table, the point group of the high-symmetry *k*-point X is C_{2v}, which allows the p–d coupling between p_y and d_{yz} according to its character table. As illustrated in Fig. 3a and b, under the C_{2v} point group, both p_y and d_{yz} transform like the same irreducible representation B_1, thus allowing strong p–d coupling between these two orbitals.[41]

In contrast, p_x/p_y and d_{xy} transform like different irreducible representations under the point groups associated with all high-symmetry *k*-points of anatase TiO_2. Taking C_{4v} as an example, in this case, (p_x, p_y) constitutes a two-dimensional irreducible representation E, while d_{xy} transforms like one-dimensional irreducible representation B_2, as illustrated in Fig. 3d and e. As a result, the p–d coupling between p_y and d_{xy} is much weaker than that between p_y and d_{yz}, well consistent with the PDOS results shown in Fig. 2c and f. In order to clearly visualize such a difference, we illustrated the charge density isosurface plot of the N doping level above the VBM of TiO_2, which is induced by a single N dopant in a 2 × 2 × 1 supercell of anatase TiO_2. As shown in Fig. 3c and f, the N doping level mainly originates from the p–d coupling between N-p_y and Ti-d_{yz} rather than between N-p_y and Ti-d_{xy}, further confirming the strong coupling between N-p_y and Ti-d_{yz}.

The existence of strong p–d coupling indicates that the N-p_y (p_π) are not isolated, since they can strongly couple with O-p_π and Ti-d_{yz} (one of three Ti-t_{2g} orbitals). In this case, a uniform distribution of the N dopants throughout the crystal supercell can ensure a similar N–N distance between different N dopants and prevent them from clustering, thus inducing delocalized molecular orbitals due to the strong coupling between N/O-p_π and Ti-t_{2g} orbitals. As these delocalized molecular orbitals correspond to the band-like doping levels above the VBM of pristine anatase TiO_2, the uniform N-doping can be intuitively expected for realizing band-like doping levels in N-doped anatase TiO_2. In order to confirm this, we constructed two B/N-codoping configurations with uniform N dopant spatial orderings (Fig. 4a and b) and calculated their band structures at the HSE06 level.[38–40] As shown in Fig. 4c and d, band-like doping levels with strong dispersion and apparent overlap with the VBM of pristine anatase TiO_2 can be observed for both of the two B/N-codoping configurations, thus achieving a significant bandgap narrowing as much as 0.70 eV.

Based on the above discussions, the existence of strong p–d coupling in anatase TiO_2 is essential for achieving band-like

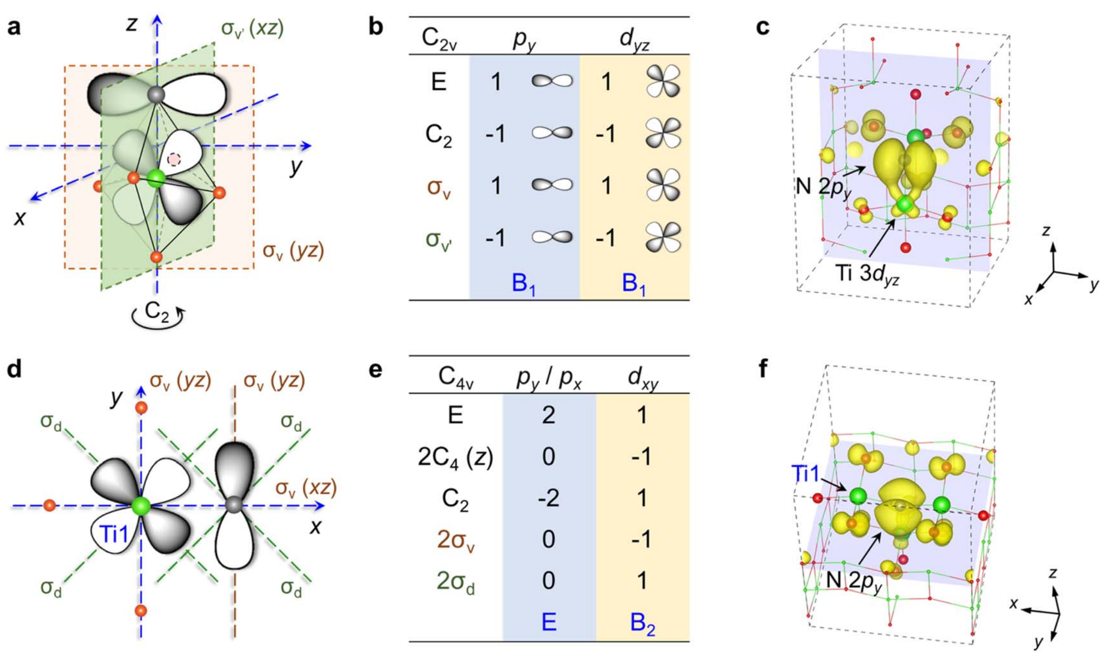

Fig. 3 Difference between d_{yz} and d_{xy} in coupling with p_y. (a) Symmetry operations of C_{2v} on both N-p_y and Ti-d_{yz}. (b) The character table of C_{2v}. (c) Charge density isosurface plot (0.004 e Å^{-3}) of the N doping level in a 2 × 2 × 1 anatase TiO_2 supercell with a single N dopant (some atoms are not shown for visual clarity). (d) Symmetry operations of C_{4v} on N-p_y and Ti-d_{xy}. (e) The character table of C_{4v}. (f) The illustration of the charge density isosurface plot in (c) from another perspective (some atoms are not shown for visual clarity).

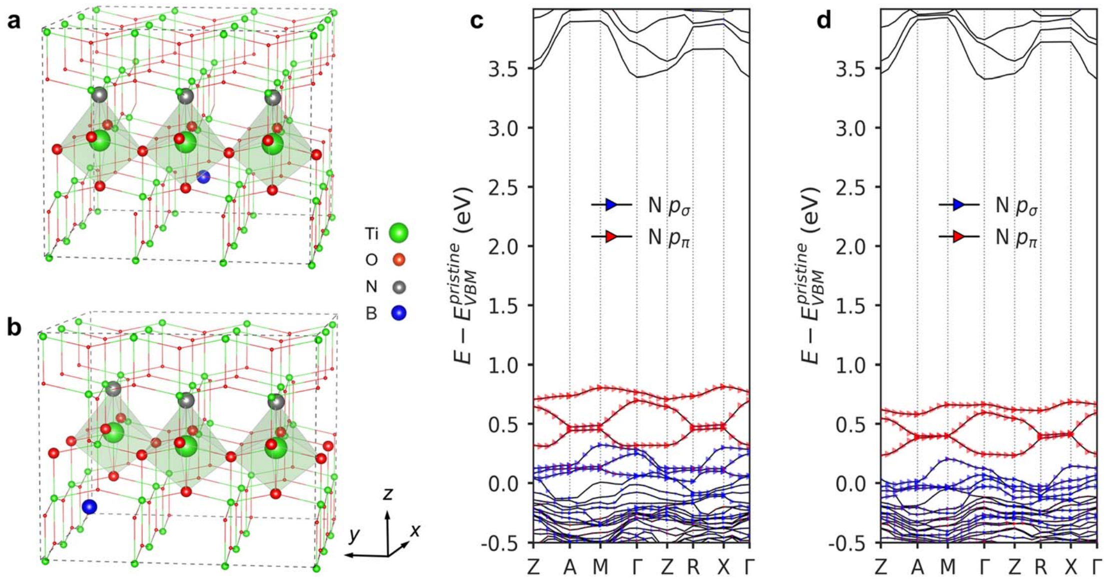

Fig. 4 B/N-codoping anatase TiO_2 configurations featuring uniform N dopant spatial orderings. (a and b) Atomic structures of two B/N-codoping configurations featuring uniform N dopant spatial orderings. (c and d) Band structures at the HSE06 level of the two B/N-codoping configurations illustrated in (a) and (b), respectively. The energy values are referenced to the VBM of pristine anatase TiO_2, which is set to be 0.0 eV.

doping levels. Due to the strong p–d coupling, the band-like doping levels can be achieved by dopant spatial orderings that are either uniform, or exhibit distinctive periodic features to enable apparent level crossing between N-p_σ and N-p_π doping levels. Both of these characteristic dopant spatial orderings match the experimental results, where both the substitutional N and interstitial B contents exhibit very small fluctuations throughout the TiO_2 microsphere.[25] Notably, as the existence of strong p–d coupling is determined by the crystal symmetry of anatase TiO_2, it enables us to effectively break out of the paradigm of correlating the origin of band-like doping levels with the electronic structures of a few specific B/N-codoping configurations. More significantly, it also enables the generalization of this p–d coupling criterion for the prediction of the feasibility of inducing band-like doping levels in other metal oxides. In order to validate this generalization, we further considered the other two polymorphs of TiO_2, namely rutile and brookite TiO_2.

As for rutile TiO_2, the point groups of the high-symmetry k-points in the first Brillouin zone are either D_{4h} or D_{2h} as summarized in Table S1.† Since p_π and t_{2g} orbitals transform like different irreducible representations in both of the point groups, the p–d coupling is much weaker in rutile TiO_2 compared to that in anatase TiO_2. Aiming to illustrate such a difference between rutile and anatase TiO_2, we have calculated the PDOS onto N-p_π and Ti-t_{2g} orbitals for both N-doped anatase and rutile TiO_2. As shown in Fig. S9,† the hybridization peak between N-p_π and Ti-t_{2g} orbitals is sharper and has a much higher value in N-doped anatase TiO_2 than in N-doped rutile TiO_2, demonstrating a much stronger p–d coupling between N-p_π and Ti-t_{2g} orbitals in anatase than in rutile TiO_2.

Therefore, it can be predicted that band-like doping levels are hard to be achieved in rutile TiO_2. In order to verify this, we also predicted the bandgap values of all possible B/N-codoping configurations in a 2 × 2 × 3 rutile TiO_2 supercell by using the established ML method. Specifically, by substituting three lattice O atoms with three N atoms and introducing one interstitial B atom, a total of 2576 symmetrically inequivalent B/N-codoping configurations can be obtained. Among these, 678 representative ones were uniformly sampled for DFT calculations, thus enabling the training of a highly accurate XGBoost model for the prediction of bandgap values, as summarized in Fig. S10.† Based on the ML prediction, we identified the 40 B/N-codoping configurations that exhibit the most significant bandgap narrowing, as shown in Fig. S11.† Besides, we also calculated the band structures of the two B/N-codoping configurations with the most significant bandgap narrowing at the HSE06 level.[38–40] As shown in Fig. S12,† the doping levels are largely localized in B/N-codoped rutile TiO_2, which is in accordance with the prediction made based on the p–d coupling criterion. Actually, this is in accordance with the previous DFT study on the effect of N doping on rutile TiO_2, where N doping typically induces localized doping levels above the VBM of rutile TiO_2 according to the PDOS plots.[46] Moreover, we also validated the prediction by synthesizing the B/N-codoped rutile TiO_2 using an approach similar to that for synthesizing the B/N-codoped anatase TiO_2.[24,47] As shown in Fig. S13,† only a shoulder-like absorption edge around 450 nm can be observed in the obtained B/N-codoped rutile TiO_2. This is typically induced by localized doping levels, since the localized doping levels are discrete within the bandgap, thereby only promoting light absorption at specific wavelengths rather than inducing a band-to-band redshift of the absorption edge. It should also be noted that the obtained B/N-codoped rutile TiO_2 has a greenish color rather than red, and a tail-like visible light absorption band can be observed in the UV-visible absorption spectra, which can be ascribed to the existence of a high concentration of oxygen vacancies.[48] However, in the N doped rutile TiO_2 where substitutional fluorine (F) was adopted as the charge compensator rather than the interstitial B, the obtained N/F-codoped rutile TiO_2 has an orange color,[49] indicating that the substitutional F can more effectively compensate for the charge imbalance between N^{3-} and O^{2-}, thereby substantially lowering the concentration of oxygen vacancies. Noting that, the N/F-codoping still induces localized doping levels above the VBM of rutile TiO_2, which is also in agreement with the prediction made based on the p–d coupling criterion.

For brookite TiO_2, its high-symmetry *k*-point S in the first Brillouin zone allows p–d coupling, as summarized in Table S1.† Thus, it can be predicted that band-like doping levels can be achieved by uniform N doping in brookite TiO_2. In order to validate this, by introducing one interstitial B dopant and three substitutional N dopants in a 2 × 2 × 1 brookite TiO_2 supercell, we constructed a B/N-codoping configuration with uniform N dopant spatial ordering (Fig. S12c†). As shown in Fig. S12f,† band-like doping levels with strong dispersion and apparent overlap with the VBM of pristine brookite TiO_2 can be observed for this B/N-codoping configuration, thus achieving a significant bandgap narrowing as much as 0.63 eV.

Since the p_π orbitals originate from sp^2 hybridization of the three-coordinated N/O atoms, and the t_{2g} orbitals arise from the splitting of metal d orbitals by the O_h point group symmetry of the MO_6 cluster, rather than being unique to Ti atom. Therefore, the physical picture of the p–d coupling criterion is in principle also applicable for other MO_6-cluster-based metal oxide semiconductors. To confirm this, we further studied the effect of N-doping on the electronic structures of four more wide bandgap metal oxides including $ScTaO_4$, WO_3, SnO_2 and $MgTa_2O_6$. By analyzing the point groups as summarized in Table S2,† strong p–d coupling is allowed in $ScTaO_4$ and WO_3, while it is absent in both SnO_2 and $MgTa_2O_6$. Therefore, it can be predicted that band-like doping levels can also be obtained in $ScTaO_4$ and WO_3 by uniform N-doping, while this is hard to be fulfilled in SnO_2 and $MgTa_2O_6$.

In order to validate the prediction for $ScTaO_4$, we constructed a uniform N-doping configuration based on a 2 × 2 × 2 $ScTaO_4$ supercell, and a B/N-codoping configuration by further introducing an interstitial B dopant into the N-doping configuration. The electronic structures of these two doping configurations are shown in Fig. S14.† Clearly, uniform N-doping results in band-like doping levels that merge apparently with the VBM of $ScTaO_4$. Moreover, the charge compensation effect of the interstitial B dopant can further broaden the band-like doping levels to achieve a larger bandgap narrowing of about 1.1 eV. These results well explain the significant overall bandgap narrowing (up to 1.32 eV) achieved experimentally in $ScTaO_4$ by codoping with N and charge compensators.[50] As for WO_3, we also constructed a uniform N-doping configuration based on a 2 × 2 × 1 WO_3 supercell. From the calculated electronic structures shown in Fig. S15,† band-like doping levels above the VBM of WO_3 can also be obtained, and this should be the real cause for the smooth shift of the absorption edge towards the visible light region in N-doped WO_3 observed in experiment.[51]

Notably, the demonstrated uniform doping mechanism has significant implications, since it explains, for the first time, why uniform N doping is essential for better performance in metal oxides. This can be highlighted by comparing the first red anatase TiO_2 (ref. 24) with the later improved one,[25] which features a more uniform N doping instead of gradient N doping. Due to the uniform N doping, not only the band-to-band light absorption but also the photocatalytic oxygen evolution activity was significantly improved. In fact, besides the above metal oxides, uniform N doping is also the key for achieving state-of-the-art photoelectrochemical and photocatalytic performances in other metal oxides, for instance $Cs_{0.68}Ti_{1.83}O_4$,[52] where strong p–d coupling exists in both metal oxides according to symmetry analysis, as summarized in Table S3.†

As for the other two metal oxides (SnO_2 and $MgTa_2O_6$) without strong p–d coupling, the absence of band-like doping levels has been firmly confirmed by considering a sufficient number of N-doping configurations with the most uniform N dopant spatial orderings. In specific, within a 2 × 2 × 3 SnO_2 supercell, we first identified the 24 doping configurations with the most uniform distributions of three N dopants, and calculated the corresponding band structures, as shown in Fig. S16.† Clearly, the doping levels are largely localized within the bandgap and exhibit no overlap with the VBM of SnO_2. Besides, as for the 24 N-doping configurations with the most uniform distributions of three N dopants within a 2 × 2 × 1 $MgTa_2O_6$ supercell, similar localized doping levels can also be observed, as shown in Fig. S17.† These results are in full accordance with the predictions based on the p–d coupling criterion.

Conclusion

In conclusion, based on the ML methods and DFT calculations, it was revealed that the existence of strong p–d coupling between O/N-p_π and Ti-t_{2g} orbitals in anatase TiO_2 is the prerequisite of inducing band-like doping levels. As the existence of the strong p–d coupling between p_π and t_{2g} orbitals in anatase TiO_2, as well as in other metal oxides, is entirely determined by crystal symmetry regardless of specific metal elements, we proposed a universal criterion based on p–d coupling to predict whether band-like doping levels can be induced in metal oxides through N doping. Notably, the accuracy of the criterion has been comprehensively verified on rutile/brookite TiO_2, $ScTaO_4$, WO_3, SnO_2 and $MgTa_2O_6$. Overall, our work provides a fundamental understanding and design guidelines for bandgap engineering of wide bandgap metal oxides.

Data availability

The data supporting this article have been included as part of the ESI.†

Author contributions

Conceptualization, G. L., L. C. Y. and K. Y. Z.; investigation, K. Y. Z.; writing – original draft, K. Y. Z.; writing – review & editing, K. Y. Z., L. C. Y., G. L. and X. Q. C.; funding acquisition, L. C. Y. and

G. L.; resources, L. C. Y., G. L. and H. M. C.; supervision, L. C. Y., G. L. and H. M. C.; experiments, G. Q. D. All authors discussed the results and commented on the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 52425201, 51972312), the National Key R & D Program of China (No. 2021YFA1500800), and the International Partnership Program of Chinese Academy of Sciences, Grant No. 172GJHZ2022010MI. The theoretical calculations in this work are performed on TianHe-1(A) at the National Supercomputer Center in Tianjin. Authors are thankful to Prof. Xinzheng Li, Prof. Junwei Luo and Prof. Yan Sun for valuable discussions. Gang Liu is thankful for the financial support from the New Cornerstone Science Foundation through the XPLORER PRIZE.

References

1 X. Chen and S. S. Mao, Titanium dioxide nanomaterials: Synthesis, properties, modifications, and applications, *Chem. Rev.*, 2007, **107**, 2891–2959.
2 J. Jia, *et al.*, Heterogeneous catalytic hydrogenation of CO_2 by metal oxides: defect engineering - perfecting imperfection, *Chem. Soc. Rev.*, 2017, **46**, 4631–4644.
3 X. G. Yu, T. J. Marks and A. Facchetti, Metal oxides for optoelectronic applications, *Nat. Mater.*, 2016, **15**, 383–396.
4 R. Asahi, T. Morikawa, T. Ohwaki, K. Aoki and Y. Taga, Visible-light photocatalysis in nitrogen-doped titanium oxides, *Science*, 2001, **293**, 269–271.
5 Z. D. Li, F. Wang, A. Kvit and X. D. Wang, Nitrogen doped 3D titanium dioxide nanorods architecture with significantly enhanced visible light photoactivity, *J. Phys. Chem. C*, 2015, **119**, 4397–4405.
6 R. Marschall and L. Z. Wang, Non-metal doping of transition metal oxides for visible-light photocatalysis, *Catal. Today*, 2014, **225**, 111–135.
7 A. H. Al-Naggar, N. M. Shinde, J. S. Kim and R. S. Mane, Water splitting performance of metal and non-metal-doped transition metal oxide electrocatalysts, *Coord. Chem. Rev.*, 2023, **474**, 214864.
8 R. Asahi, T. Morikawa, H. Irie and T. Ohwaki, Nitrogen-doped titanium dioxide as visible-light-sensitive photocatalyst: Designs, developments, and prospects, *Chem. Rev.*, 2014, **114**, 9824–9852.
9 W. Wang, M. O. Tade and Z. P. Shao, Nitrogen-doped simple and complex oxides for photocatalysis: A review, *Prog. Mater. Sci.*, 2018, **92**, 33–63.
10 M. Chiesa, S. Livraghi, M. C. Paganini, E. Salvadori and E. Giamello, Nitrogen-doped semiconducting oxides. Implications on photochemical, photocatalytic and electronic properties derived from EPR spectroscopy, *Chem. Sci.*, 2020, **11**, 6623–6641.
11 C. M. Ban, *et al.*, A novel codoping approach for enhancing the performance of $LiFePO_4$ cathodes, *Adv. Energy Mater.*, 2012, **2**, 1028–1032.
12 C. Di Valentin and G. Pacchioni, Spectroscopic properties of doped and defective semiconducting oxides from hybrid density functional calculations, *Acc. Chem. Res.*, 2014, **47**, 3233–3241.
13 J. Wang, *et al.*, Origin of photocatalytic activity of nitrogen-doped TiO_2 nanobelts, *J. Am. Chem. Soc.*, 2009, **131**, 12290–12297.
14 N. Umezawa and J. H. Ye, Role of complex defects in photocatalytic activities of nitrogen-doped anatase TiO_2, *Phys. Chem. Chem. Phys.*, 2012, **14**, 5924–5934.
15 M. N. H. Liton, *et al.*, Dual acceptor (N, Cu) doping effects on the electronic and optical properties of ZnO, *Mater. Chem. Phys.*, 2020, **242**, 122463.
16 E. Albanese, *et al.*, Nature of paramagnetic species in nitrogen-doped SnO_2: A combined electron paramagnetic resonance and density functional theory study, *J. Phys. Chem. C*, 2015, **119**, 26895–26903.
17 C. Zhang, *et al.*, Effect of non-metal elements (B, C, N, F, P, S) mono-doping as anions on electronic structure of $SrTiO_3$, *Comput. Mater. Sci.*, 2013, **79**, 69–74.
18 K. S. Yang, Y. Dai and B. B. Huang, Study of the nitrogen concentration influence on N-doped TiO_2 anatase from first-principles calculations, *J. Phys. Chem. C*, 2007, **111**, 12086–12090.
19 P. Li, *et al.*, Theoretical studies on the form and effect of N-doping in an $ZnGa_2O_4$ photocatalyst, *RSC Adv.*, 2016, **6**, 74483–74492.
20 F. F. Wu, *et al.*, Efficient photocatalytic oxygen production over nitrogen-doped $Sr_4Nb_2O_9$ under visible-light irradiation, *ChemCatChem*, 2016, **8**, 615–623.
21 H. F. Fu, *et al.*, Enhanced ethanol sensing performance of N-doped ZnO derived from ZIF-8, *Chin. Chem. Lett.*, 2023, **34**, 107425.
22 W. S. Liu, *et al.*, Nitrogen doping in Ta_2O_5 and its implication for photocatalytic H_2 production, *Appl. Surf. Sci.*, 2018, **459**, 477–482.
23 M. Batzill, E. H. Morales and U. Diebold, Influence of nitrogen doping on the defect formation and surface properties of TiO_2 rutile and anatase, *Phys. Rev. Lett.*, 2006, **96**, 026103.
24 G. Liu, *et al.*, A red anatase TiO_2 photocatalyst for solar energy conversion, *Energy Environ. Sci.*, 2012, **5**, 9603–9610.
25 X. X. Hong, *et al.*, Control of spatially homogeneous distribution of heteroatoms to produce red TiO_2 photocatalyst for visible-light photocatalytic water splitting, *Chem.–Eur. J.*, 2019, **25**, 1787–1794.
26 H. Kageyama, *et al.*, Expanding frontiers in materials chemistry and physics with multiple anions, *Nat. Commun.*, 2018, **9**, 772.
27 W. J. Yin, S. H. Wei, M. M. Al-Jassim and Y. F. Yan, Double-hole-mediated coupling of dopants and its impact on band gap engineering in TiO_2, *Phys. Rev. Lett.*, 2011, **106**, 066801.

28 C. H. Sun and D. J. Searles, Origin of the visible light absorption of boron/nitrogen co-doped anatase TiO_2, *J. Phys. Chem. C*, 2013, **117**, 26454–26459.
29 M. Niu, D. J. Cheng and D. P. Cao, Understanding photoelectrochemical properties of B-N codoped anatase TiO_2 for solar energy conversion, *J. Phys. Chem. C*, 2013, **117**, 15911–15917.
30 Q. Zhou, *et al.*, Learning atoms for materials discovery, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E6411–E6417.
31 M. Kaneko, M. Fujii, T. Hisatomi, K. Yamashita and K. Domen, Regression model for stabilization energies associated with anion ordering in perovskite-type oxynitrides, *J. Energy Chem.*, 2019, **36**, 7–14.
32 T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.*, 2018, **120**, 145301.
33 P. G. Ghanekar, S. Deshpande and J. Greeley, Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis, *Nat. Commun.*, 2022, **13**, 5788.
34 R. E. A. Goodall and A. A. Lee, Predicting materials properties without crystal structure: deep representation learning from stoichiometry, *Nat. Commun.*, 2020, **11**, 6280.
35 D. V. Anand, Q. Xu, J. Wee, K. L. Xia and T. C. Sum, Topological feature engineering for machine learning based halide perovskite materials design, *npj Comput. Mater.*, 2022, **8**, 203.
36 K. Y. Zhang, L. C. Yin, G. Liu and H. M. Cheng, Accurate structural descriptor enabled screening for nitrogen and oxygen vacancy codoped TiO_2 with a large bandgap narrowing, *J. Mater. Sci. Technol.*, 2022, **122**, 84–90.
37 T. Q. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System. Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
38 J. Heyd and G. E. Scuseria, Efficient Hybrid Density Functional Calculations in Solids: Assessment of the Heyd-Scuseria-Ernzerhof Screened Coulomb Hybrid Functional, *J. Chem. Phys.*, 2004, **121**, 1187–1192.
39 J. Heyd, G. E. Scuseria and M. Ernzerhof, Hybrid Functionals based on a Screened Coulomb Potential, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
40 J. Heyd, G. E. Scuseria and M. Ernzerhof, Erratum: "Hybrid Functionals based on a Screened Coulomb Potential" [J. Chem. Phys. 118, 8207 (2003)], *J. Chem. Phys.*, 2006, **124**, 219906.
41 M. S. Dresselhaus, G. Dresselhaus and A. Jorio, *Group Theory: Application to the Physics of Condensed Matter*, Springer, Heidelberg, 2008.
42 C. Di Valentin, G. Pacchioni and A. Selloni, Origin of the different photoactivity of N-doped anatase and rutile TiO_2, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2004, **70**, 085116.
43 J. Ma and S. H. Wei, Origin of novel diffusions of Cu and Ag in semiconductors: The case of CdTe, *Phys. Rev. Lett.*, 2013, **110**, 235901.
44 A. H. Slavney, B. A. Connor, L. Leppert and H. I. Karunadasa, A pencil-and-paper method for elucidating halide double perovskite band structures, *Chem. Sci.*, 2019, **10**, 11041–11053.
45 L. D. Yuan, H. X. Deng, S. S. Li, S. H. Wei and J. W. Luo, Unified theory of direct or indirect band-gap nature of conventional semiconductors, *Phys. Rev. B*, 2018, **98**, 245203.
46 A. Miyoshi, A. Kuwabara and K. Maeda, Effects of nitrogen/fluorine codoping on photocatalytic rutile TiO^{-2} crystal studied by first-principles calculations, *Inorg. Chem.*, 2021, **60**, 2381–2389.
47 T. T. Wu, Y. P. Xie, L. C. Yin, G. Liu and H. M. Cheng, Switching photocatalytic H_2 and O_2 generation preferences of rutile TiO_2 microspheres with dominant reactive facets by boron doping, *J. Phys. Chem. C*, 2015, **119**, 84–89.
48 Y. Q. Yang, *et al.*, An unusual strong visible-light absorption band in red anatase TiO_2 photocatalyst induced by atomic hydrogen-occupied oxygen vacancies, *Adv. Mater.*, 2018, **30**, 1704479.
49 A. Miyoshi, *et al.*, Nitrogen/fluorine-codoped rutile titania as a stable oxygen-evolution photocatalyst for solar-driven Z-scheme water splitting, *Sustainable Energy Fuels*, 2018, **2**, 2025–2035.
50 L. Pei, *et al.*, A novel visible-light-responsive semiconductor $ScTaO_{4-x}N_x$ for photocatalytic oxygen and hydrogen evolution reactions, *ChemCatChem*, 2021, **13**, 180–184.
51 M. Takeuchi, Y. Shimizu, H. Yamagawa, T. Nakamuro and M. Anpo, Preparation of the visible light responsive N^{3-}-doped WO_3 photocatalyst by a thermal decomposition of ammonium paratungstate, *Appl. Catal., B*, 2011, **110**, 1–5.
52 G. Liu, *et al.*, Band-to-band visible-light photon excitation and photoactivity induced by homogeneous nitrogen doping in layered titanates, *Chem. Mater.*, 2009, **21**, 1266–1274.