Chapter 2
Electronic Structure of Transparent Conducting Oxides

J. Robertson and B. Falabretti

2.1 Introduction

Metallic oxides are a materials class showing one of the greatest range of properties – superconducting, ferroelectric, ferromagnetic [1], multiferroic, magneto-resistive, dielectric, or conducting. Of particular interest are the so-called transparent conducting oxides (TCOs) and amorphous semiconducting oxides (ASOs). The TCOs are heavily used for flat panel displays, photovoltaic cells, low emissivity windows, electrochromic devices, sensors and transparent electronics [2–4]. Oxides are of particular interest because the metal-oxide bond is strong so that the oxides have a combination of a high heat of formation and a wide band gap, compared to any similar compound.

This chapter describes the basic electronic structure of oxides that allows this to occur, why they can be doped, what controls the polarity of the doping, and the effect of disorder on their properties. The majority of the TCOs are n-type electron conductors. A few p-type hole conductors have been discovered following the break through of Kawazoe et al. [5].

2.2 Band Structures of n-Type Oxides

There are numerous n-type TCOs. We will focus here on the electronic structure of a subset of them, SnO_2, In_2O_3, ZnO, Ga_2O_3 and CdO, which illustrate their main properties. These all are oxides of group IIB-IVB metals. They have smaller ions and are not as electropositive as the corresponding alkaline earth metals of groups

J. Robertson (✉)
Department of Engineering, Cambridge University, Cambridge CB2 1PZ, UK
e-mail: jr@eng.cam.ac.uk

D.S. Ginley (ed.), *Handbook of Transparent Conductors*,
DOI 10.1007/978-1-4419-1638-9_2,

IIA. They are predominantly ionic bonded except for ZnO. Their crystal structures are summarised in Fig. 2.1.

SnO_2 is perhaps the simplest of the TCOs. It has the rutile structure, in which each tin atom is surrounded by six oxygens in an octahedral array, and each oxygen is surrounded by three tin atoms in a planar array. Figure 2.2 shows the band

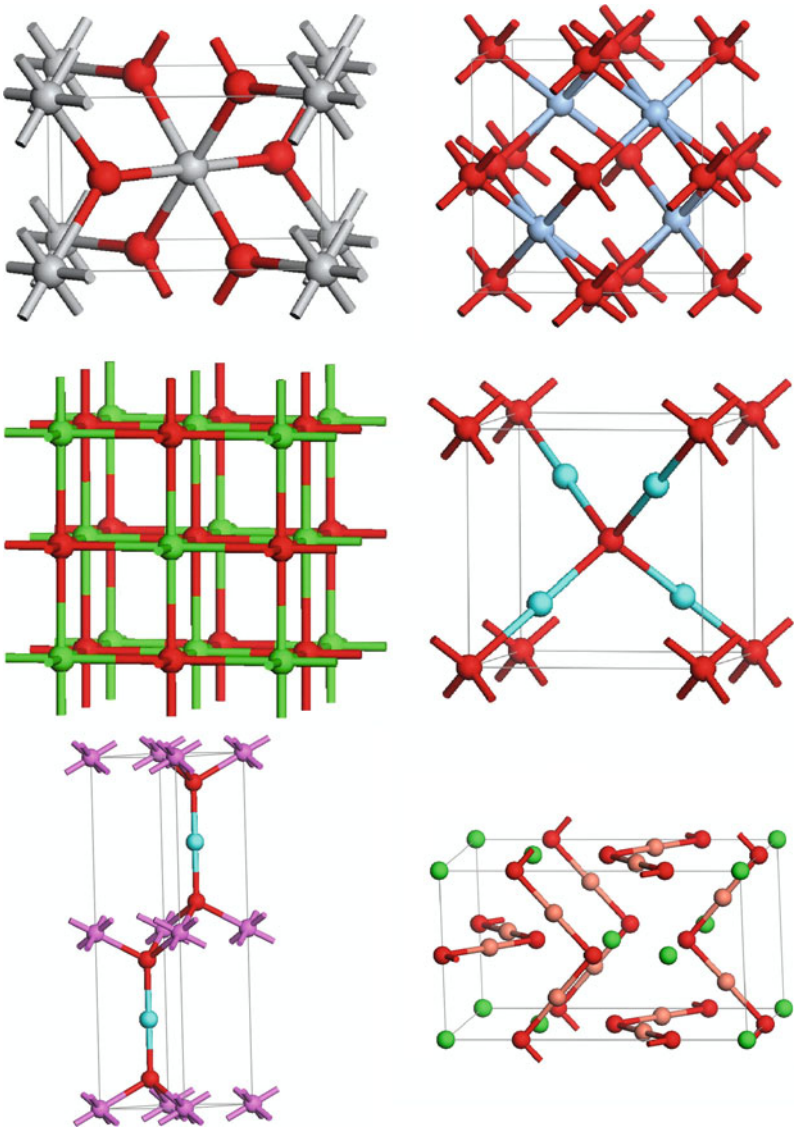

Fig. 2.1 Crystal structures of SnO_2, In_2O_3, CdO, 2H-$CuAlO_2$ and $SrCu_2O_2$ (oxygen = *darker balls*)

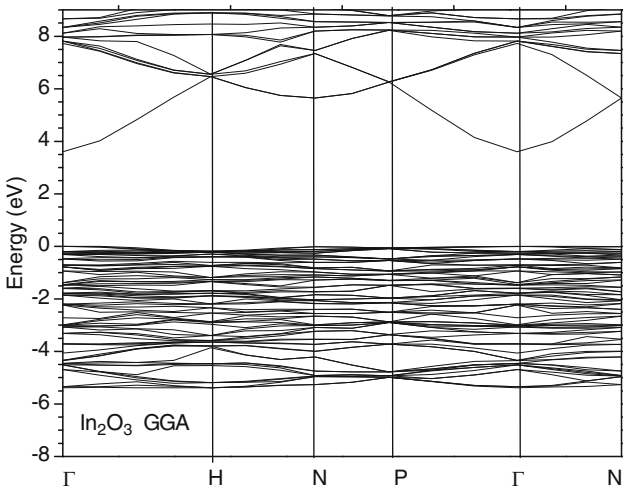

Fig. 2.2 Band structure of SnO_2. Band gap fitted

structure of SnO_2. The band gap is 3.6 eV and direct. The band structure shown here was calculated by the plane wave pseudopotential method, using the generalised gradient approximation (GGA) of the local density formalism (LDF). The GGA functional represents the exchange-correlation energy of the electron gas. These LDF and GGA methods under-estimate the band gap. This error has been corrected in the band structure shown by the "scissors operator," a rigid upward shift of the conduction bands. The band structure of SnO_2 was first calculated correctly by Robertson [6], followed by numerous calculations using improved methods [7–9].

The most obvious feature in Fig. 2.2 is the free-electron-like conduction band minimum at the zone centre Γ. It is noticeable that this state from 3.6 eV upwards is a single minimum, without any subsidiary minima leading to indirect gaps. This main minimum is formed out of Sn 5s states [6]. In a tight-binding description, it consists of 96% of Sn s states. The band gap is direct. The electron effective mass is 0.23–0.3, reasonably small, but not 0.1 like for example Si.

The upper valence band from 0 eV down to −8.1 eV consists mainly of O 2p states, mixed with some Sn s and p states. The ionicity of SnO_2 is about 60%, so this sets the Sn content of the valence band, averaged over the Brillouin zone, as rather low. Finally, at −16 eV, there are O 2s states which do not contribute to the bonding. Any Sn 4d states lie below this, and can be ignored.

The upper valence band in SnO_2 is typical of many oxides. It is relatively flat, and thus has a large effective mass, which does not favour conduction by holes. The valence band maximum has a Γ_2^- symmetry, which leads to a direct forbidden band gap [10]. The valence band is consistent with the experimental ultraviolet photo-emission spectra [11].

The next most important TCO, In_2O_3, has the bixbyite structure, in which the oxygens form a close packed lattice and the In ions lie at sixfold and fourfold

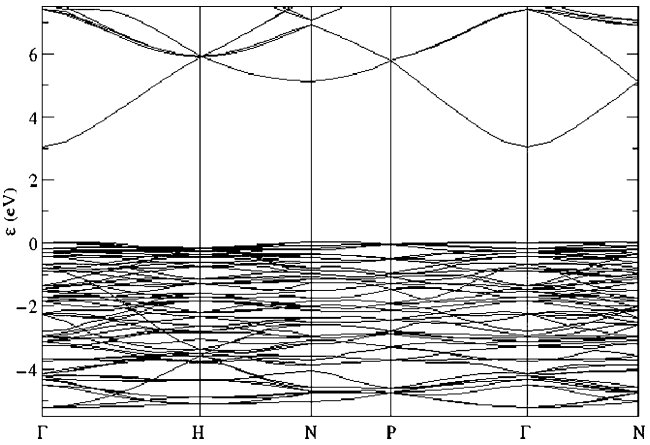

Fig. 2.3 Band structure of In_2O_3 by screened exchange

interstices. The In sites are sixfold coordinated by oxygen. The overall symmetry is cubic, but the unit cell is large, 40 atoms. Figure 2.2 shows our calculated band structure of In_2O_3. The band gap was originally believed to be 3.7 eV. The first good band calculation of In_2O_3 was by the Shigesato et al. [12], followed by Mi et al. [8], and Mryasov and Freeman [13].

We see again in Fig. 2.3 that In_2O_3 has a single free-electron-like conduction band minimum, formed from In s states. The minimum band gap is direct. The effective mass is 0.3 m. Thus, the conduction band minimum of In_2O_3 has the same nature as that of SnO_2. The main valence band is 5.2 eV wide, less than that of SnO_2, indicating that In_2O_3 is more ionic than SnO_2. This is consistent with the ultra violet photoemission spectra of Christou et al. [14] and Klein [15]. Below the O 2p states come In 4d states, then O 2s states.

The band gap of In_2O_3 has recently been re-appraised. A smaller indirect gap was once proposed [16]. However, there can be no indirect gap due to the parabolic nature of the conduction band. It is now realised by Walsh et al. [17–19] that the minimum band gap is 2.9 eV, direct and forbidden. The screened exchange band structure in Fig. 2.3 gives this value. The upper valence bands all have the wrong symmetry for allowed optical transitions to the conduction band, as in SnO_2. The first allowed optical transition is 0.8 eV below the valence band top [17]. This now gives a consistent view of the band structure, optical gap and surface band bending of In_2O_3.

A third important TCO, especially for photovoltaic applications, is ZnO. Interest in ZnO is wider than just as a TCO, because it is a prototype direct-gap, wide band gap optoelectronic semiconductor in competition with GaN [20]. It has been an important phosphor. It is also easy for make as “nanorods.” ZnO typically has the hexagonal wurzite structure in which each Zn or O atom is surrounded by four neighbours of the other type. There is also a hypothetical zinblende polymorph of ZnO, with the same bonding. The band structure of this zincblende phase is shown

in Fig. 2.4, as the gaps and band widths are the same. It band gap is 3.35 eV and direct, and the conduction band minimum is again a single broad minimum formed from Zn s states [8]. There are numerous band calculations of ZnO [21, 22], some at high levels of accuracy such as GW [23]. Note that the uncorrected GGA band gap of ZnO is only ~0.9 eV, very small compared to experiment. Donors such as interstitial Zn or substitutional Al are shallow in ZnO, but other defects like the O vacancy are deep.

Ga_2O_3 is the least studied of the binary oxides. It has more complex crystal structure such as the β-Ga_2O_3 structure. In this, the Ga sites are both fourfold or sixfold coordinated. Its band structure is shown in Fig. 2.5. The band structure has a minimum direct band gap of 4.52–4.9 eV [24, 25]. The broad conduction band

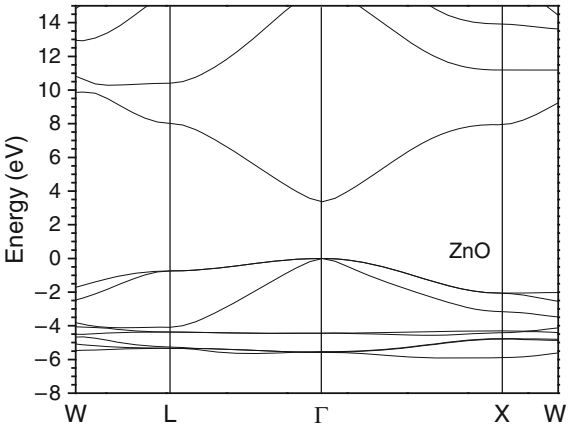

Fig. 2.4 Band structure of Ga_2O_3

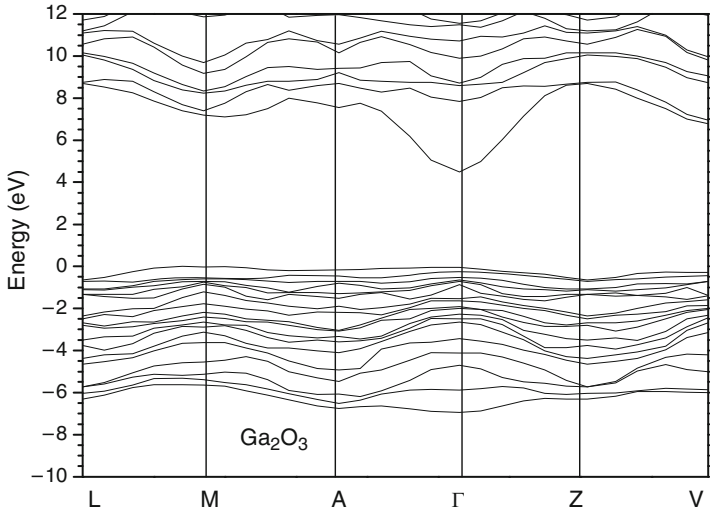

Fig. 2.5 Band structure of zincblende (*cubic*) ZnO

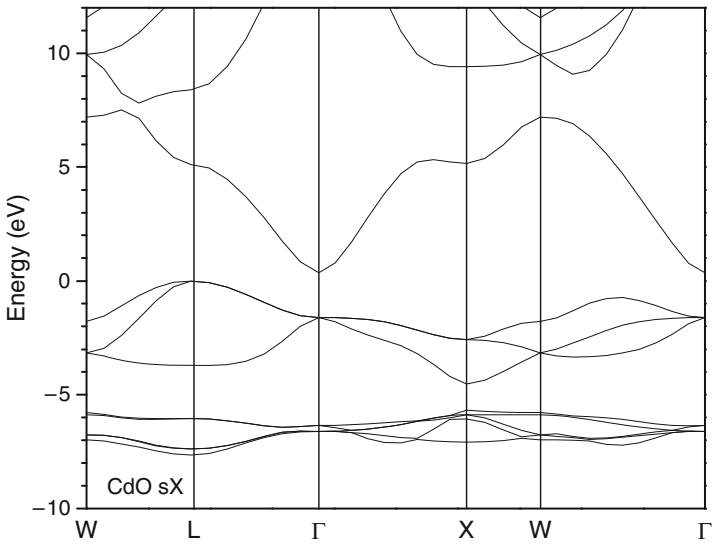

Fig. 2.6 Band structure of CdO using screened exchange

minimum at Γ is formed of Ga s states. The valence band maximum is very flat, a slightly indirect gap and a maximum at M 0.05 eV above Γ.

The last n-type oxide considered is CdO. This has the rock-salt structure, in which each Cd or O ion is surrounded by 6 neighbours. Its band structure is shown in Fig. 2.6. This is similar to that found by others [22, 26, 27]. In contrast to the other TCOs, the minimum gap of CdO is indirect, at 0.8 eV [28, 29]. The conduction band minimum is at Γ, free-electron-like, and is formed from Cd s states. However, the valence band maximum is not at Γ. It is displaced to the zone boundary at the L point and along ΓW, due to a repulsion of the O p states in the upper valence band by Cd d states lying at −7 eV. The three upper valence bands in CdO consist mainly of O 2p states. However, instead of Cd p states leading to a downward repulsion of these states away from Γ, the upward repulsion of Cd d states is stronger. This leads to a calculated minimum indirect gap of 0.6 eV and a minimum direct gap of 2.1 eV, compared to experimental values of 0.8 and 2.3 eV, respectively. The valence band width is consistent with that seen experimentally by photoemission [30].

When these various oxides are doped with donors, the free electrons lie in the lowest conduction band. The next available empty state after this is not the conduction band minimum itself, but higher unoccupied conduction states. This increases the energy of the lowest optical transition, the optical gap. This is the Moss-Burstein shift, and varies inversely with the effective mass [3].

An aspect that is not so important for the overall understanding of TCOs, but nevertheless relevant to this chapter, is the question of the calculated band gaps. The standard method of calculating electronic structure uses the local density functionals such as the GGA to represent the electron's exchange-correlation

energy. These so-called ab initio methods are able to give the structure, bond lengths and total energies quite well. However, they under-estimate the band gap of semiconductors and insulators. It is generally found that this under-estimate is of order 30%, as is the case of Si. However, for the oxides of interest here, the under-estimate is very large, typically 70%. For SnO_2 the GGA band gap is 1.2 eV compared to the experimental value of 3.6 eV. The largest problem is for CdO, where LDF gives a negative band gap of −0.5 eV.

In Figs. 2.2 and 2.7–2.11 we have corrected the GGA band gaps by the scissors operator, rigidly shifting the conduction bands upwards. The conduction band dispersions in GGA are correct. Thus, once corrected for the error, the band structures shown are correct.

There are a number of methods beyond LDF which do give better band gaps. The best known of these is the GW method [31], but this is computationally very expensive. Other popular methods are the self-interaction correction (SIC) [32] and the B3LYP functional [33–35]. B3LYP is a hybrid functional, that is a LDF-type functional of the exchange-correlation energy containing a fraction of the Hartree-Fock function, which can give the correct total energy and reasonable eigenvalues. Other hybrid functionals are PBE0, HSE and screened exchange.

A hybrid functional which is valuable is the method of screened exchange. This was first implemented by Kleinman [36]. It was then taken up by Freeman et al. [37, 38] and applied to various oxides of interest here [26]. We have also used screened exchange for TCOs [39, 40]. The CdO band structure shown in Fig. 2.6 is that calculated by the screened exchange method. It gives a reasonable band gap compared to experiment. In other cases, the sX or WDA band structures can be used to verify that the conduction bands formed by the scissors operator are indeed correct, and that gaps do have the character as shown in Figs. 2.2–2.10.

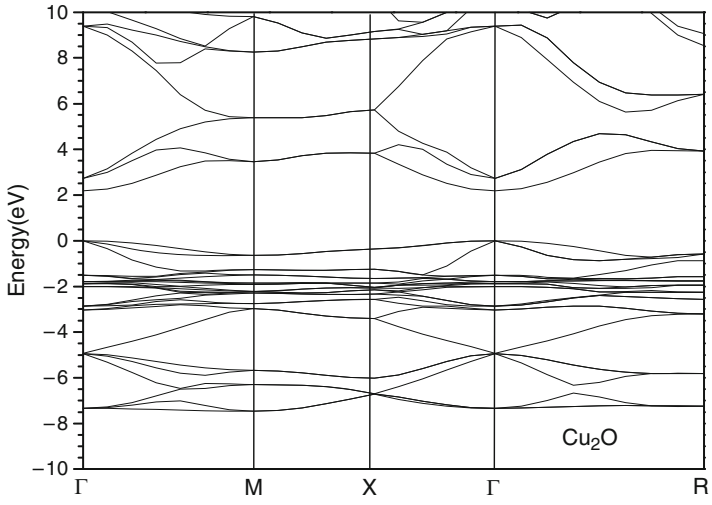

Fig. 2.7 Band structure of Cu_2O

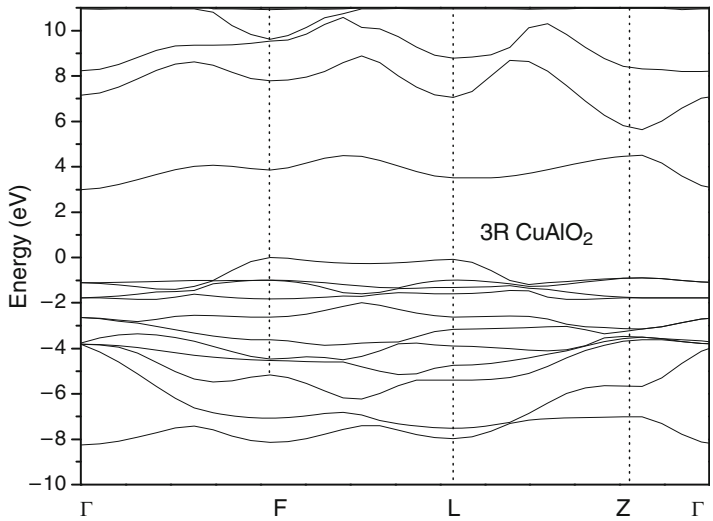

Fig. 2.8 Band structure of 3R-$CuAlO_2$

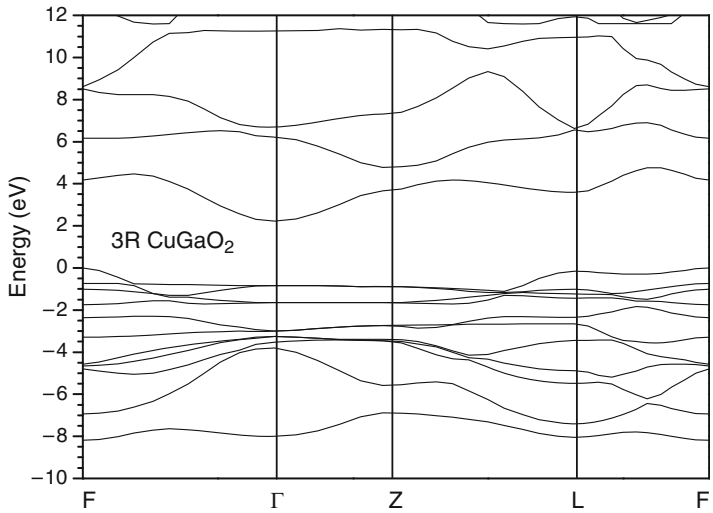

Fig. 2.9 Band structure of 3R-$CuGaO_2$

2.3 Band Structures of p-Type Oxides and Other Cu Oxides

For many years, the only apparent p-type conducting oxide was Cu_2O. This was used since early days in Cu-Cu_2O rectifiers. However, its band gap is only 2.17 eV. Its band structure calculated in the GGA method is shown in Fig. 2.7. The Cu_2O

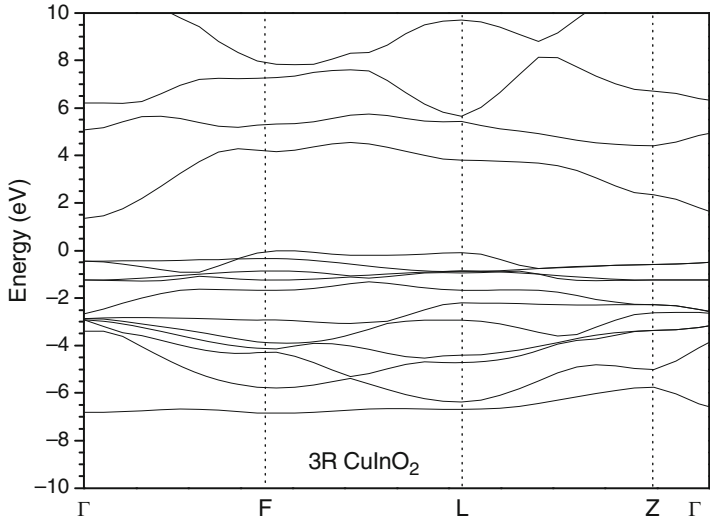

Fig. 2.10 Band structure of 3R-$CuInO_2$

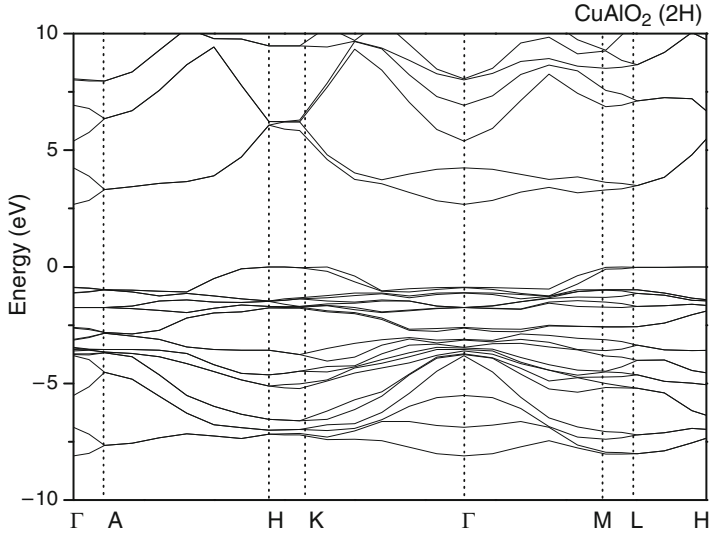

Fig. 2.11 Band structure of 2H-$CuAlO_2$

crystal structure is unusual, the oxygens are fourfold coordinated, and the Cu atoms are only twofold coordinated, in a linear configuration (Fig. 2.1). In the band structure, there is again a broad conduction band minimum at Γ due to Cu s states [41–45]. The Cu 3d states are all filled. They lie as a mass of narrow bands at around

−1 to −3 eV. The Cu d states mix with the O 2p states and form a continuous band, which extends down to −7 eV.

Kawazoe et al. [5] noted that the main problem with oxides is the high effective mass of their holes. Thus acceptors would not ionise. The high hole effective mass arises in SiO_2 or SnO_2 because the valence band maximum states are nonbonding pπ states, directed perpendicular to the bonding direction. Thus, they have a small dispersion. ZnO is one case without nonbonding p states, but it still has a rather large hole mass. Kawazoe et al. [5] noted that we should try to increase the dispersion of the valence band top states, and the way to do this is to hybridise (mix) them with d states of a cation at a similar energy. The Cu d states are the best case of this. Thus, the interaction of Cu d and O 2p states should reduce the effective hole mass in Cu_2O.

However, the problem with Cu_2O itself is that its 2.17 eV band gap is too narrow to be transparent across the optical spectrum. This arises because its conduction band is too broad, and its conduction band minimum falls too low. The dispersion of the Cu s band is mainly due to Cu-Cu interactions. The width of the Cu s band is proportional to the number of Cu neighbours. By using $CuAlO_2$ in the defossalite structure, we can reduce the Cu-Cu coordination from 12 to 3, and this should reduce the conduction band dispersion and increase the band gap [5]. This indeed occurs [46]. Of course, the largest interest in Cu based oxides arose from high temperature superconductivity, so the role of holes in Cu-O layers is well known.

$CuAlO_2$ has a layered structure in which the Cu ions a linearly coordinated to two O atoms, as in Cu_2O. The Al ions are surrounded by six oxygens. These AlO_6 units form a layer of hexagonal symmetry. The AlO_6 and Cu layers can be stacked in various patterns. A two-layer repeat gives the 2H (P63/mmc) form and a three-layer repeat gives the 3R (R-3m) rhombohedral form. Despite having more layers, the primitive cell of the 3R form has the fewer atoms. Figure 2.8 shows the band structure of 3R-$CuAlO_2$ calculated using the GGA functional. The indirect band gap has been adjusted to the experimental value. The bands are very similar to those given by Ingram et al. [47]. The other important calculations of $CuAlO_2$ are by Yanagi [48] and Zhang [49]. We see that 3R-$CuAlO_2$ is an indirect gap semiconductor. The conduction band minimum is at Γ and consists of Cu s states, while the valence band maximum is at F. The upper valence band from 0 eV down to −8 eV consists of a mixture of Cu d and O 2p states, with the Cu d states tending to lie higher. There are O 2s states at −20 eV.

The minimum indirect gap is 3.0 eV, and the minimum direct gap is 3.5 eV at L. Ingram et al. [47] and Zhang et al. [49] have discussed the nature of the optical transitions. We see that the conduction band has subsidiary minima at F and L, so it is not as simple as those in Cu_2O or SnO_2.

In Fig. 2.8, when using the scissors operator, we fitted the main direct gap at 3.5 eV, not the minimum gap, which is not well known experimentally. The resulting band gap is nevertheless consistent with that found by the screened exchange method or weighted density approximation [39, 40].

Figures 2.9 and 2.10 show the band structures of 3R-$CuGaO_2$ and 3R-$CuInO_2$ calculated using the GGA functional. In each case, the main direct gap was fitted [49]. We see that the conduction band minimum of $CuInO_2$ is mainly at Γ, and there are no other subsidiary minima as was found in $CuAlO_2$. The valence band of $CuInO_2$ is reasonably similar to that of $CuAlO_2$. The band width is slightly smaller and the band dispersions flatter, so the hole effective mass is larger. There is also a mass of flat bands at −14 eV due to In d states.

The minimum indirect band gap of $CuInO_2$ is now 1.4 eV, and the minimum direct gap is 3.9 eV at L. Thus, the indirect gap decreases from $CuAlO_2$ to $CuInO_2$, but the direct gap increases, which is unusual [47, 49]. This gives the appearance that the optical gap increases when changing from $CuAlO_2$ to $CuInO_2$, which is against the trend in ionic radii.

$CuCrO_2$ is an oxide with the defossalite structure which also shows p-type behaviour [50]. Now the band gap appears in the middle of the Cr d states. This is consistent with the insulating property of Cr_2O_3. The alloy $CuAl_{1-x}Cr_xO_2$ can be doped p-type [51] by substitutional Mg_{Al}.

Figure 2.11 shows the band structures of the 2H polymorph of $CuAlO_2$ [35]. It confirms that the stacking does not alter the main band gaps so much.

The final oxide considered is $SrCu_2O_2$. This is an amipolar oxide [52–54]. $SrCu_2O_2$ has the body-centred tetragonal $D_{4h}{}^{10}$, 4/mmm space group. Its structure consists of O-Cu-O zig-zag chains in the x and y direction, separated by SrO_6 octahedra. Cu is again twofold coordinated by O. There are four Cu atoms in the primitive unit cell. The band structure of $SrCu_2O_2$ is shown in Fig. 2.12. The conduction band has a minimum at Γ and the valence band has a weak maximum at Γ. The bands have been calculated by Ohta et al. [53], Robertson et al. [35] and

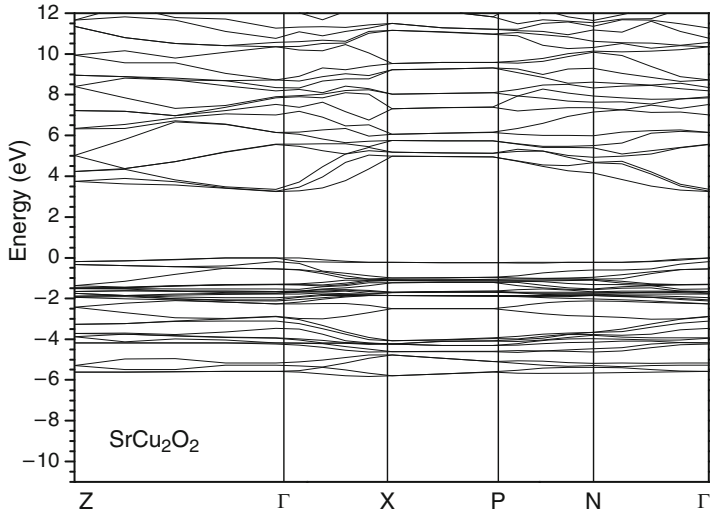

Fig. 2.12 Band structure of $SrCu_2O_2$

Nie et al. [44]. The different appearance of some of the published bands arises because not all authors used the primitive cell. There are a number of other ternary oxides of interest which are covered elsewhere [47, 48].

2.4 Band Line-ups and Work Functions

An important application of TCOs is as electrodes on semiconductor devices, such as solar cells or organic light emitting diodes. A key criterion is that these electrodes form ohmic rather than injecting contacts. This depends on the band alignment of the conducting oxides to the semiconductor, the Schottky barrier height. The barrier height can be measured by photoemission, internal photoemission or by electrical means. It can be estimated by a variety of theoretical means, which we now discuss.

An often-used method is to compare the work function of the conductor ϕ_M with the electron affinity (EA) of the semiconductor (for the n-type case), χ, each measured from the vacuum level. The barrier height is then taken as the difference,

$$\phi_n = \phi_M - \chi \tag{2.1}$$

This approximation is the so-called electron affinity rule. It can work for very wide gap semiconductors, or those with van der Waals bonding such as organic semiconductors.

Figure 2.13 plots the work functions of the various TCOs, data taken from Minami et al. [56]. It is interesting that the n-type TCOs have very high work functions, their conduction band minima lying well below the vacuum level. The work functions of the oxide films are often maximised by ensuring an oxygen-rich outer surface, by processing. It is interesting in Fig. 2.13 that the conduction band minima of SnO_2, and In_2O_3 lie *deeper* than the Fermi level of the *parent metal*. This behaviour distinguishes the best n-type TCOs from normal oxides.

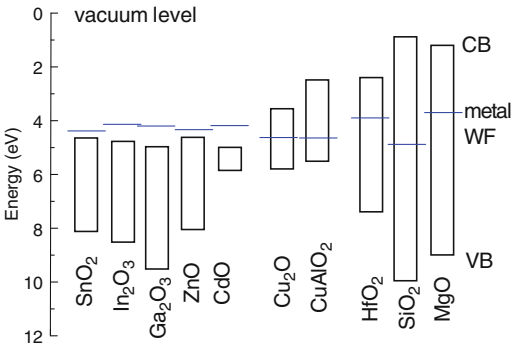

Fig. 2.13 Work functions and electron affinities of various oxides, compared to the work functions of their parent metal. Note how the metal WF of n-type TCO lies above the oxide conduction band

In fact, the electron affinity rule does not work for metals on typical semiconductors [57–59]. This is because the semiconductor interfaces possess mid gap states which tend to pin the metal Fermi level from changes in barrier height. The semiconductor's mid gap states on the neutral surface are filled up to some energy, which we will call the charge neutrality level (CNL). The effect of these states is to try to pin the metal work function towards this CNL. The degree of pinning depends on the density of these states and their extent into the semiconductor.

The net effect is that the Schottky barrier heights tend to follow the equation

$$\phi_n = S(\Phi_M - \Phi_S) + (\Phi_S - \chi_s) \tag{2.2}$$

Here Φ_S is the CNL energy measured from the vacuum level. S is the Schottky barrier pinning factor. S = 1 for the strongly pinned case, a narrow gap semiconductor, and S = 0 for the unpinned case, like SiO_2, as in the electron affinity rule.

Monch [57] found that S follows an empirical dependence on the electronic dielectric constant, ε_∞,

$$S = \frac{1}{1 + 0.1(\varepsilon_\infty - 1)^2} \tag{2.3}$$

This model can also be applied to the interfaces between two semiconductors [60], to derive their band offsets, where the electron barrier ϕ_n or conduction band (CB) offset between semiconductors *a* and *b* is given by

$$\phi_n = (\chi_a - \Phi_{S,a}) - (\chi_b - \Phi_{S,b}) + S(\Phi_{S,a} - \Phi_{S,b}) \tag{2.4}$$

There are more detailed methods of calculating the band offsets, as for example given by van de Walle et al. [61], or by Zunger et al. [62], based on calculations of explicit interface structures.

The CNL can be calculated from the oxide band structure as the energy at which the simple Greens function is zero;

$$G(E) = \int_{BZ} \int_{-\infty}^{\infty} \frac{N(E')dE'}{E - E'} = 0 \tag{2.5}$$

The CNL can be calculated from the bands calculated by the local density approximation/pseudopotential method, after adjusting the band gap to the experimental value.

The CNL normally lies near the centre of the band gap. For most ionic oxides, the CNL energy tends to vary with the metal valence, because the large number of oxygen-related valence states repels the CNL up in the gap [60, 63]. The transparent oxides are different. The CNL is effectively the mid point of the *average* gap over the Brillouin zone. But the s-band oxides have a broad CB minimum, which comes

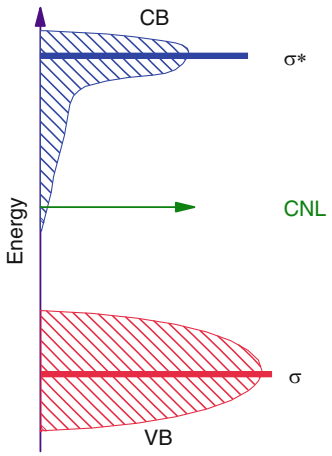

Fig. 2.14 Schematic of the charge neutrality level in an oxide with a deep s-like conduction band minimum of low density of states

well below the average CB energy. This causes the CNL to lie either close to the CB, as in ZnO, or even above the CB minimum, as in SnO_2 and In_2O_3 (Fig. 2.14). This is an unusual situation.

Parameters are listed in [64]. The S value is calculated from their experimental refractive index ($\varepsilon_\infty = n^2$). The electron affinities are taken from experiment for poly-crystalline oxide films; those for the dielectrics are tabulated previously [60]. The work function of doped SnO_2 is large, 4.5 eV or more [50]. This is partly because the surface is treated to be O-rich, to maximise the work function.

2.5 Ability to Dope

Substitutional doping is a key requirement for a semiconductor to be used in practical devices. It is often stated that doping of both polarities is required. This not strictly accurate; thin film transistors of amorphous hydrogenated silicon (a-Si:H) are the main-stay of the flat panel display industry and only use n-type doping.

There are three requirements for successful doping of a given polarity;

- Solubility of the dopant in the lattice
- Shallowness of the dopant level
- Lack of compensation of the dopant by an intrinsic defect

When designing the p-type TCOs, Kawazoe et al. [5] only considered point 2. Point 1 is usually easily satisfied, it is only a problem in cases like diamond where there is a large mismatch between the atomic radii of a shallow donor (e.g. Sb) and the small diamond lattice.

Compensation, point 3, is the most interesting case [65–68]. In wide gap semiconductors, a donor electron can lower its energy if it falls into an empty intrinsic defect state (such as a vacancy state) at the bottom of the gap. Indeed the energy gain can be so much that this energy gain exceeds the cost of creating the defect. In that case, the donor action is completely compensated by the defect, if there is thermal equilibrium.

Another way to express this is that the creation energy of the intrinsic defect (say the vacancy) depends on the Fermi energy, E_F as [66].

$$\Delta H(E_f) = qE_f + \Delta E$$

This is shown schematically in Fig. 2.15. It means that there will be some Fermi energy at which the cost of the vacancy is zero. If a dopant would move the Fermi level to this energy, called the dopant pinning energy, then vacancies will be spontaneously created at no cost. There will be two pinning energies, one for n- and one for p-type doping. Thus, if equilibrium holds, it will be impossible to shift the Fermi level beyond these two pinning energies $E_{pin,\ n}$ and $E_{pin,\ p}$. Practical doping will only occur if these pinning energies lie in the conduction or valence band, respectively, and not in the gap. If for example, $E_{pin,\ p}$ lies above the valence band edge E_v, then there will be no p-type doping, because it will not be possible to shift E_F to the valence band edge. Note that the type of defect doing the compensation will differ for n- and p-type doping.

The NREL group have applied these ideas to study the limits to doping of the tetrahedral semiconductors such as III-Vs, ZnO and the chalcopyrites [66–68]. The bands of the various semiconductors are aligned with each other using the calculated or observed band offsets, and the pinning energies are found to lie at a roughly constant energy across the series, Fig. 2.16. Zunger [68] prefers to reference band offsets to the vacuum level E_{vac}. In their picture, a semiconductor cannot easily be doped n-type if its conduction band energy lies too high towards E_{vac}, i.e. its electron affinity is too small. AlN would be an example. On the other hand, a semiconductor cannot easily be doped p-type if its valence band lies too far below E_{vac}, if its photoelectric threshold is too large. ZnO is a good example.

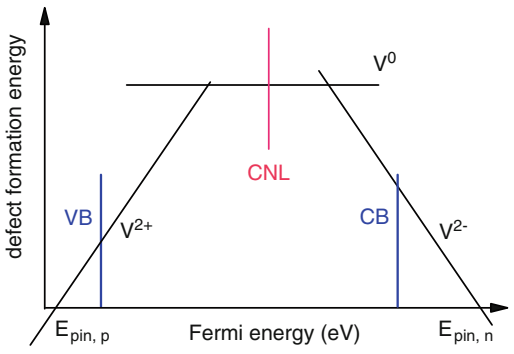

Fig. 2.15 Concept of the doping pinning levels for n- and p-type dopants, on the band diagram. If the pinning level lies inside the bands, then the semiconductor can be doped; if it falls in the gap the semiconductor cannot easily be doped to that polarity

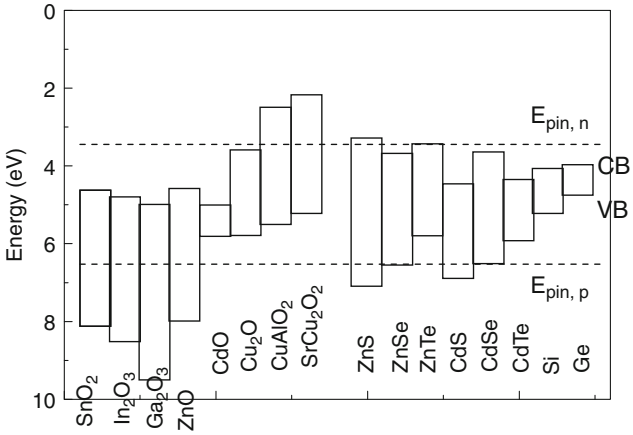

Fig. 2.16 Band diagrams of various oxides and comparative semiconductors aligned according to the vacuum level, with dopant pinning levels indicated

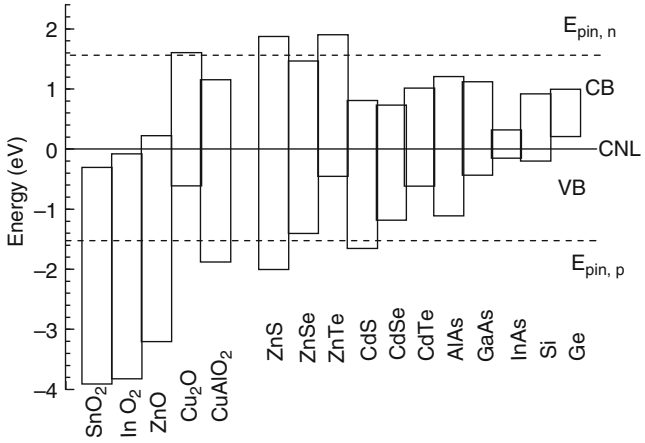

Fig. 2.17 Band diagrams of various oxides and comparative semiconductors aligned according to their charge neutrality levels, with dopant pinning levels indicated

This method gives a good view of doping possibilities in TCOs. The n-type TCOs SnO_2, In_2O_3, ZnO stand out as having very large work functions (when n-type). As they have the same band gap, their valence band maxima are very deep below E_{vac}. In this model, p-type ZnO is only possible by inhibiting thermal equilibrium occurring.

There is a second, related view of doping limits, which is more consistent with band offset models. The semiconductor oxides of interest here have $\varepsilon_\infty \sim 3.7$–$4.0$, giving $S \sim 0.5$. Thus, the bands should be aligned [55] using CNLs and (2.3) not just electron affinities. This is done in Fig. 2.17.

In addition, van de Walle and Neugebauer [69] note that for AB compounds the CNL tends to equal the average of the dangling bond energies of the cation and anion species. This arises because defect levels ultimately depend on bulk band structures. Turning to AB_n compounds, the CNL will be a weighted average of cation and anion site defect levels. Thus, in our view [70] shown in Fig. 2.17, each semiconductor is aligned according to the band offset, using primarily the CNL, with no reference to the vacuum level. The doping pinning levels then lie at some energy above and below the CNL. (This energy is not necessarily constant.) In this case, n-type doping is difficult if the conduction band edge lies too far above the CNL, and p-type doping becomes difficult if the valence band edge lies too deep below the CNL.

This is consistent with experiment. SnO_2, In_2O_3 are very unusual in that their CNLs actually lie *above* the conduction band edge [18, 70], rather than in the gap as normal. In ZnO, the CNL lies close to the conduction band edge. This accounts for their ease of n-type doping. On the other hand, that the CNL is so far above the valence band edge in SnO_2, In_2O_3 and ZnO accounts for why these oxides are difficult to dope p-type.

In the case of $CuAlO_2$ and Cu_2O, the CNL is calculated to lie closer to the valence band, and this is consistent with their p-type behaviour. In contrast, in $CuInO_2$, the CNL lies higher in the gap [64], and now this oxide can be doped both n- and p-type. $SrCu_2O_2$ also has a CNL near midgap, allowing ambipolar doping.

What controls the CNL energy? From (2.5), the CNL lies between the density of states (DOS) of valence band (VB) and conduction band (CB). A larger DOS repels the CNL away [60, 64]. A large VB DOS repels the CNL to the upper gap. A high valence and large O content gives a large VB DOS and a high CNL energy. But a second factor is the nature of the conduction band minimum. The CNL lies midway in the *indirect* gap. A deep CB minimum with only a small DOS as in SnO_2 makes only a small contribution to the integral in (2.5) and has little effect on the CNL. The CNL lies high in SnO_2, In_2O_3 and ZnO because the DOS in their CB minima is small. On the other hand, in $CuAlO_2$ its multiple CB minima push the CNL down.

A complicating factor which is not discussed here is the "self" doping of TCOs, in which vacancies are generated in multi-component oxides such as Cd_xSnO_3.

2.6 Effect of Disorder–Disorder in Amorphous Semiconductors

We now turn to a different question, why n-type TCOs work well even when they are amorphous [71, 72]. The ineffectiveness of disorder scattering has been noted for some time [73, 74]. The n-type TCOs consist of the oxides of post-transition metals, Sn, In, Ga, Zn, Cd, etc. As we noted above, the conduction band minima of these oxides are free-electron-like states, localised on the metal s states. Unlike the alkaline earth metal oxides, because of their lower ionicity, these metal oxides can be made amorphous [71].

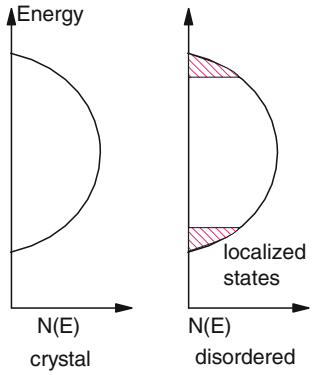

Fig. 2.18 Effect of disorder on an s band

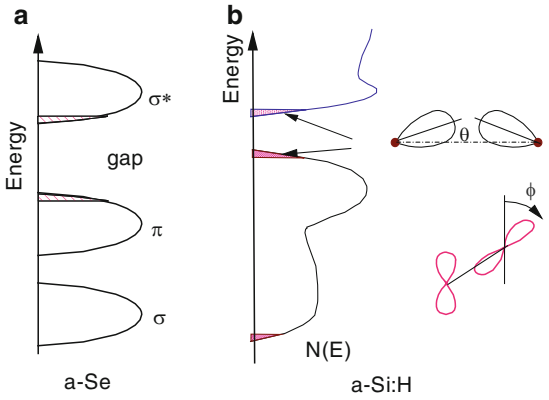

Fig. 2.19 (**a**) Schematic density of states of a-Se, (**b**) a-Si, and the effect of bond angle and dihedral angle disorder

Since the work of Anderson [75], it is known that disorder will cause a localization of electron states in the band structure. Mott [76, 77] then showed that disorder first localizes states at the band edges, and that the extended states and localized tail states are separated by an energy called the mobility edge. With increasing disorder, the mobility edges move further into the bands, as in Fig. 2.18. Eventually, the whole band becomes localized. These results were worked out for s states, which are spherically symmetric.

The first amorphous semiconductors to be studied in depth experimentally were the amorphous chalcogenides, such as amorphous Se (a-Se). The chalcogenides are in fact p bonded [78], and their simplified band diagram is shown in Fig. 2.19a. Disorder introduces localized band tail states.

The next and most important amorphous semiconductors are a-Si and hydrogenated amorphous silicon (a-Si:H). Its bonds are sp^3 states, but the states around its band gap are p states [78]. Figure 2.19b shows a schematic of its density of states. The valence band maximum consists of pure p states, whereas the conduction band minimum consists of mixed s,p states. The effect of disorder has been considered in

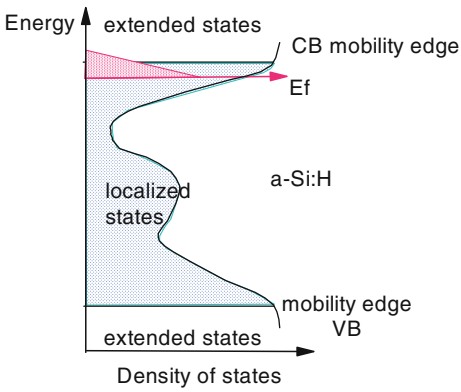

Fig. 2.20 Schematic density of states and mobility edges of a-Si:H

more detail in a-Si than in most other amorphous semiconductors [78–83]. The Si-Si bond length is relatively fixed in a-Si:H. On the other hand, the bond angle θ varies by 10° and the dihedral angle ϕ varies by 180°. The nearest neighbour V (ppπ) interaction equals 0.7 eV and it varies directly with ϕ, Fig. 2.19b, and therefore this is a strong source of disorder in the valence band edge [79]. This causes a strong tailing of the valence band edge, giving a characteristic tail width of at least 60 meV. The conduction band is less affected by dihedral angle disorder, but is affected by bond angle disorder. This also gives quite strong tailing, but less than for the valence band edge [80–82]. This gives rise to density of states in the gap, as shown schematically in Fig. 2.20.

The overall result of this strong effect of angular disorder is that the electron mobility of a-Si:H is quite low, $\sim$1 cm^2/V s. The hole mobility is very low, 10^{-3} cm^2/V s.

A second limitation of a-Si:H is that its substitutional doping is very inefficient. Whereas every substitutional atom in c-Si produces a carrier, in a-Si:H there is an unusual self-compensation mechanism which limits the doping efficiency [78, 83]. The net effect of this is that the Fermi level can never move up to the donor level, or above the mobility edge. Thus, E_F is stuck in localised states. Conduction is in extended states, but only after thermal excitation of the carriers.

2.7 Disorder in Oxide Semiconductors

The amorphous n-type TCOs are in fact the first practical examples of disorder in an s band. The effect of disorder on s states is rather weak, compared to p states. As the conduction band minimum state is 90–95% localized on metal s states, then its energy depends mainly on the V(ss) interaction between second neighbour metal sites, and not much on the V(sp) interaction between metal s and oxygen p states, Fig. 2.21. A two-centre Slater-Koster interaction V(l,m) between orbitals

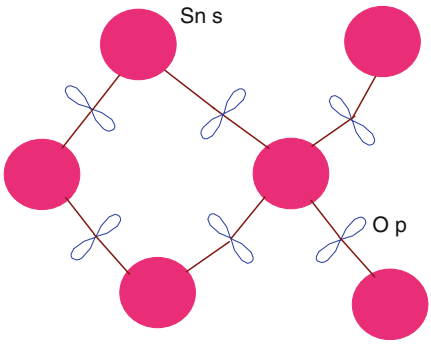

Fig. 2.21 s-Like atomic orbitals of CB states disordered SnO_2

on atoms l and m [84] would normally depend on their distance (r), the angles θ between the orbitals and the separation vector r, and their dihedral angle (ϕ),

$$V(l, m) = V(r,\ \theta, \phi)$$

However, for an interaction between two s states, this reduces to

$$V(ss) = V(r) \tag{2.6}$$

because of their spherical symmetry. Thus, the only source of disorder is the variation of the metal-metal distance, and any angular disorder has no effect on s states.

The first known effect of this was on the conduction band edge on amorphous SiO_2 (silica) [85]. SiO_2 has a wide band gap, 9 eV, and low screening. The conduction band minimum of SiO_2 has an effective mass of 0.5 m and is formed from Si s states and O 3s states [86]. The effect of disorder theoretically is small because the O-O distance is relatively fixed due to the small disorder of the O-Si-O bond angle. Holes form polarons and have a very low mobility. In contrast, unexpectedly, free electrons have a high mobility, indicating an absence of disorder effects and localized states. (Of course, there are not many free electrons, due to the band gap).

The same effect occurs in SnO_2 and related TCOs [87]. The large metal ion radius means that ion packing keeps the metal-metal distance rather constant, s states mean that the angular disorder has no effect, so the effects of disorder on the s-like conduction band minimum is very small. This is the case for the pure oxide. The density of states is summarised in Fig. 2.22. There is no mobility edge.

Experimentally, these s-like TCOs differ very strongly from a-Si:H. The electron mobility is large, of order 10–40 cm^2/V s. The Fermi level can be moved far into the conduction band without any problem, creating large free carrier concentrations [71]. Thin film transistors of n-type oxides show high field effect mobilities, 10–40 cm^2/V s [87–95], much higher than a-Si:H.

It is interesting that the recent oxide-based TFTs use mixed oxides of Ga, In and Zn [87–89] to control the off-current and vacancy concentration. The carrier

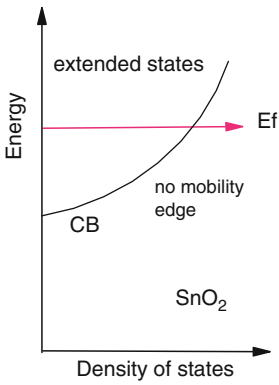

Fig. 2.22 Schematic density of states of disordered SnO_2, with no mobility edge, and the Fermi level able to enter the extended conduction band states

mobility is still high. Clearly the conduction band edge is still very delocalized and it is not even affected by compositional disorder. The reason for this is that Al, Ga or In form shallow donor states in ZnO [96], while ZnO forms resonant states in Ga or In oxide. Ga forms a shallow bound state below the ZnO CB edge, but at high concentrations, this forms a continuous band with the ZnO states in the alloy. Similarly, in SnO_2, substitutional Sb gives a shallow state [97]. The absence of deep states due to aliovalent impurities means that there are no localized states at the conduction band edge, and no effects that break the delocalisation of the CB states.

The absence of localized states due to aliovalent dopants also means that disorder does not introduce localized tail states below E_c. Thus, in a TFT, the field effect mobility is given essentially by the free carrier mobility or Hall mobility. The free carrier mobility in s states is much higher than in p states, which partly accounts for the higher FE mobility. A second factor is that the Fermi wavevector is large. This contrast with a-Si where the Hall coefficient has the opposite sign to the carrier.

A very narrow conduction band tail of order 0.1 eV was found in a calculation of the alloy $InGaZnO_x$ [98]. Note that the Urbach energy of these disordered oxides is ~0.2 eV [99], because of valence band tailing. Extended X-ray Fine Structure shows the local bonding [100].

2.8 Summary

The band structures of the various transparent conducting oxides are given, and discussed in terms of their band edge properties. The reasons for the ability to dope them n- or p-type given. Finally, the oxides are shown to be able to support a higher electron mobility than amorphous silicon due to their s-like conduction band minima.

Acknowledgements We thank Dr. S. J. Clark for many band calculations.

References

1. C H Ahn, J M Triscone, J Mannhart, Nature 424 (2003) 1015
2. D S Ginley, C Bright (eds) MRS Bull 25 (2000)
3. I Hamberg, C G Gramquist, J App Phys 60 (1986) R123
4. M Batzill, U Diebold, Prof Surf Sci 79 (2005) 47
5. H Kawazoe, N Yasukawa, H Hyodo, M Kurita, H Yanagi, H Hosono, Nature 389 (1997) 939
6. J Robertson, J Phys C 12 (1979) 4767
7. K C Mishra, K H Johnson, P C Schmidt, Phys Rev B 51 (1995) 13972
8. Y Mi, H Odaka, S Iwata, Jpn J App Phys 38 (1999) 3453
9. C Kilic, A Zunger, Phys Rev Lett 88 (2002) 095501
10. V T Agekyan, Phys Stat Solidi A 43 (1977) 11
11. R G Egdell, W R Flavell, P Tavener, J Solid State Chem 51 (1984) 345
12. H Odaka, S Iwata, N Taga, S Ohnishi, Y Kaneta, Y Shigesato, Jpn J App Phys 36 (1997) 5551
13. O N Mryasov, A J Freeman, Phys Rev B 64 (2001) 233111
14. R L Weiher, R P Ley, J App Phys 37 (1966) 299
15. V Christou, M Etchells, O Renault, P J Dobson, OV Salata, J App Phys 88 (2000) 5180
16. A Klein, App Phys Lett 77 (2000) 2009
17. A Walsh, J L F DaSilva, S H Wei, C Korber, A Klein, L F J Piper, A DeMast, K E Smith, G Panaccione, P Torelli, D J Payne, A Bourlange, R G Egdell, Phys Rev Lett 100 (2008) 167402
18. P D C King, T D Veal, D J Payne, A Bourlange, R G Egdell, C F McConville, Phys Rev Lett 101 (2008) 116808
19. A Bourlange et al, App Phys Lett 92 (2008) 092117
20. U Ozgur, Y I Alikov, C Liu, A Teke, M A Reshchikov, H Morkoc, J App Phys 98 (2005) 041301
21. A F Kohan, G Ceder, D Morgan, C G van de Walle, Phys Rev B 61 (2000) 15019
22. A Schleife, F Fuchs, J Furthmuller, F Bechstedt, Phys Rev B 73 (2006) 245212
23. M Usuda et al, Phys Rev B 66 (2002) 125101
24. N Ueda, H Hosono, R Waseda, H Kawazoe, App Phys Lett 71 (1997) 933
25. M Orita, H Ohta, M Hirano, H Hosono, App Phys Lett 77 (2000) 4166
26. R Asahi, A Wang, J R Babcock, N L Edleman, A W Metz, M A Lane, V P Dravid, C R Kannewurf, T J Marks, Thin Solid Films 411 (2002) 101
27. J E Medvedeva, A J Freeman, Europhys Lett 69 (2005) 583
28. P Koffyberg, Phys Rev B 13 (1976) 4470
29. R C Whited, C J Flaten, W C Walker, Solid State Commun 13 (1973) 1903
30. Y Dou, R G Egdell, D S L Law, N M Harrison, B G Searle, J Phys Cond Mat 10 (1998) 8447
31. M S Hybertsen, S G Louie, Phys Rev B 34 (1986) 5390
32. A Filippetti, N A Spaldin, Phys Rev B 67 (2003) 125109
33. A D Becke, J Chem Phys 98 (1993) 5648
34. J Muscat, A Wander, N M Harrison, Chem Phys Lett 342 (2001) 397
35. J Robertson, P W Peacock, M D Towler, R Needs, Thin Solid Films 411 (2002) 96
36. B M Bylander, L Kleinman, Phys Rev B 41 (1990) 7868
37. R Asahi, W Mannstadt, A J Freeman, Phys Rev B 59 (1999) 7486
38. C B Geller, W Wolf, S Picozzi, A Continenza, R Ashi, W Mannstadt, A J Freeman, E Wimmer, App Phys Lett 79 (2001) 368
39. J Robertson, K Xiong, S J Clark, Thin Solid Films 496 (2006) 1
40. J Robertson, K Xiong, S J Clark, Phys Stat Solidi B 243 (2006) 2054
41. L Kleinman, K Mednick, Phys Rev B 21 (1980) 1549
42. J Robertson, Phys Rev B 28 (1983) 3378
43. W Y Ching, Y N Xu, K Wong, Phys Rev B 40 (1989) 7684

44. X Nie, S H Wei, S B Zhang, Phys Rev B 65 (2002) 075111
45. M Nolan, S R Elliott, Phys Chem Chem Phys 8 (2006) 5350
46. A Buljan, P Alemany, E Ruiz, J Phys Chem B 103 (1999) 8060
47. B J Ingram, T O Mason, R Asahi, K T Park, A J Freeman, Phys Rev B 64 (2001) 155114
48. H Yanagi, S Inoue, K Ueda, H Kawazoe, N Hamada, J App Phys 88 (2000) 4159
49. X Nie, S H Wei, S B Zhang, Phys Rev Lett 88 (2002) 066405
50. R Nagarajan, A D Draeseke, A W Sleight, J Tate, J App Phys 89 (2001) 8022
51. D O Scanlon, A Walsh, B J Morgan, G W Watson, D J Payne, R G Egdell, Phys Rev B 79 (2009) 035101
52. A Kudo, H Yanagi, H Hosono, H Kawazoe, App Phys Lett 73 (1998) 220
53. H Ohta, M Orita, M Hirano, I Yagi, K Ueda, H Hosono, J App Phys 91 (2002) 3074
54. D Segev, S H Wei, Phys Rev B 71 (2005) 125129
55. S B Zhang, S H Wei, App Phys Lett 80 (2002) 1376
56. T Minami, T Miyaia, T Yamamoto, Surf Coating Technol 108 (1998) 583
57. W Mönch, Phys Rev Lett 58 (1987) 1260
58. W Mönch, Surface Sci 300 (1994) 928
59. J Tersoff, Phys Rev B 30 (1984) 4874
60. J Robertson, J Vac Sci Technol B 18 (2000) 1785
61. C G Walle, J Neugebauer App Phys Lett 70 (1997) 2577
62. S H Wei, A Zunger, App Phys Lett 72 (1998) 2011
63. P W Peacock, J Robertson, J App Phys 92 (2002) 4712
64. J Robertson, B Falabretti, J App Phys 100 (2006) 014111
65. W Walukiewicz, Physica B 302 (2001) 123
66. S B Zhang, S H Wei, A Zunger, J App Phys 83 (1998) 3192
67. S B Zhang, S H Wei, A Zunger, Phys Rev Lett 84 (2000) 1232
68. A Zunger, App Phys Lett 83 (2003) 57
69. C G van de Walle, J Neugebauer, Nature 423 (2003) 626
70. B Falabretti, J Robertson, J App Phys 102 (2007) 123703
71. H Hosono, J Non-Cryst Solids 352 (2006) 851
72. J Robertson, Phys Stat Solidi B 245 (2008) 1026
73. J R Bellingham, W A Phillips, C J Adkins, J Phys Condens Mat 2 (1990) 6207
74. J R Bellingham, W A Phillips, C J Adkins, J Mats Sci Lett 11 (1992) 263
75. P W Anderson, Phys Rev 109 (1958) 1492
76. N F Mott, Phil Mag 19 (1969) 835
77. N F Mott, Phil Mag 22 (1970) 7
78. J Robertson, Adv Phys 32 (1983) 361
79. J Singh, Phys Rev B 23 (1981) 4156
80. T Tiedje, J M Cebulka, D L Morel, B Abeles, Phys Rev Lett 46 (1981) 1425; Solid State Commun 47 (1983) 493
81. G Allan, C Delerue, M Lannoo, Phys Rev B 57 (1998) 6933
82. R Atta-Fynn, P Biswas, P Ordejon, D A Drabold, Phys Rev B 69 (2004) 085207
83. R A Street, Phys Rev Lett 49 (1982) 1187
84. W A Harrison, Electronic structure, W A Freeman, San Francisco, 1979, p 480
85. N F Mott Adv Phys 58 (1977) 363; Phil Mag B 58 (1988) 369
86. J R Chelikowksy, M Schluter, Phys Rev B 15 (1977) 4020
87. K Nomura, T Kamiya, H Ohta, K Ueda, M Hirano, H Hosono, App Phys Lett 85 (2004) 1993
88. K Nomura, H Ohta, K Ueda, T Kamiya, M Hirano, H Hosono, Science 300 (2003) 1269
89. K Nomura, H Ohta, A Takagi, T Kamiya, M Hirano, H Hosono, Nature 432 (2004) 488
90. H Q Chiang, J F Wager, R L Hofmann, J Jeong, A Keszler, App Phys Lett 86 (2005) 013503
91. W B Jackson, R L Hoffman, G S Herman, App Phys Lett 87 (2005) 193503
92. E Fortunato, A Pimentel, A Goncalves, A Marques, R Martins, Thin Solid Films 502 (2006) 704
93. B Yaglioglu, H Y Yeon, R Beresford, D C Paine, App Phys Lett 89 (2006) 062103

94. P F Garcia, R S McLean, M H Reilly, G Nunes, App Phys Lett 82 (2003) 1117
95. Navamathavan R et al, J Electrochem Soc 153 (2006) G385
96. S B Zhang, S H Wei, A Zunger, Phys Rev B 63 (2001) 075205
97. J Robertson, Phys Rev B 30 (1984) 3520
98. K Nomura, T Kamiya, H Ohta, T Uruga, M Hirano, H Hosono, Phys Rev B 75 (2007) 035212
99. K Nomura, T Kamiya, H Yanagi, H Hosono, App Phys Lett 92 (2008) 202117
100. D Y Cho, J Song, K D Na, App Phys Lett 94 (2009) 112112

http://www.springer.com/978-1-4419-1637-2

Handbook of Transparent Conductors

Ginley, D.; Hosono, H.; Paine, D.C. (Eds.)

2011, XIV, 534 p. 107 illus. in color., Hardcover
ISBN: 978-1-4419-1637-2