UNIVERSITY
OF OSLO

Defects and Transport in Crystalline Solids

Per Kofstad† and Truls Norby

Compendium for the advanced level course
Defect Chemistry and Reactions in Solids
KJM 5120 and KJM 9120 (formerly KJM 4120 and KJ-MV 417)

at
Department of Chemistry
Faculty of Mathematics and Natural Sciences
University of Oslo

2007

Version: September 2007

Preface

In 1972 Per Kofstad published his "Non-stoichiometry, diffusion and electrical conductivity in binary metal oxides". It has been a popular textbook in defect chemistry of oxides worldwide, not least because it contained a comprehensive review of defect structure and defect-related properties of all binary metal oxides. It followed Kofstad's equally well-recognized book "High temperature oxidation of metals" from 1967, revised and published under the title "High temperature corrosion" in 1987.

In the 25 years that have passed since 1972 the field of defect chemistry in oxides has grown and developed in many ways. New and improved measurement techniques, new synthesis techniques, as well as skilled and devoted investigators have produced better data for many oxides, partly encouraged by industrial developments in the use of solid electrolytes (solid oxide fuel cells, gas separation membranes, ceramic electrodes, catalysts, etc). Computer modelling enables detailed analysis of defect chemistry and transport processes, and helps us interpret experimental data. Although the increasing accuracy and complexity is welcome and unavoidable, the sound overview and ability to analyse a situation may be obscured or lost. The newcomer may find the field difficult to enter,

Along with the development in the application of theory and modelling, the field has changed focus from binary oxides to ternary or higher oxides, much due to the growing awareness that perovskite and related structures tolerate very large substitutions and variations in valence and stoichiometry. Here we find lanthanum aluminate insulators, titanate and zirconate electroceramics with ferroelectricity, piezoelectricity, etc, the high-T_c cuprate and other superconductors, high-temperature electronic conductors such as doped lanthanum cobaltate and chromite, high temperature ionic conductors such as barium indate and doped lanthanum gallate, mixed conductors, and alkaline earth cerates which dissolve protons and become proton conductors at high temperatures.

A revision of the book from 1972 as such is thus utopic in many respects. In the early 1990s Kofstad decided to author a new text that represented a development of the general and theoretical first part of the former book, while he intended to include only a few examples of data from important example oxide systems. After having sketched the book he was struck by cancer in 1993. He fought the cancer for four years, with the obligation to finish the book as one of the driving forces. However, deprived of much of his usual strength, he was unable to finish the work. I became engaged as a co-author at that time with the purpose of completing the book.

I used the new text as a curriculum when I took over the teaching of Kofstad’s course in defect chemistry at the University of Oslo, thereby being able to adjust and improve it step by step. However, the burdens of taking over the research group has kept me from finishing the book. Furthermore, a textbook on defect chemistry by Kofstad’s close friend and colleague, Donald M. Smyth was published that took away some of the immediate need for our textbook. All in all I have felt that the new book needs more work and refinement. With the present text I am still in the middle of that process, and feedback is thus as welcome as ever.

Truls Norby, February 2006

Copyrights and use of graphic material

Graphic material is original, except where stated. Reproduction from other sources have in most instances not yet been authorised in the present manuscript which is not sold commercially. Such authorisations will be collected – or original artwork created - before any commercial version of this text is printed.

Truls Norby, February 2006.

Contents

Preface .... 1—3

Copyrights and use of graphic material .... 1—4

Contents... 1—5

1. STRUCTURE AND DEFECTS ....1—10

Introduction.. 1—10

Common crystal structures... 1—13
Close-packing and voids.. 1—13
Less close-packed arrangements... 1—15
Some simple structures for oxides and other ionic compounds................ 1—16

Point defects... 1—21
Notation for description of point defects.. 1—21
Notation for charged defects... 1—22

Electroneutrality .... 1—24

Stoichiometry and nonstoichiometry.. 1—25
Defects in stoichiometric compounds. .... 1—25
Defects in nonstoichiometric oxides. .... 1—27

Associations of point defects, point defect clusters, extended defects...... 1—28
Example cases of defect clusters... 1—29
Extended defects, shear structures. .... 1—29
Modulated structures/infinitely adaptive structures.................................. 1—30

Summary .... 1—31
Literature.. 1—31
Exercises... 1—32

2. DEFECT REACTIONS...2—1

Introduction.. 2—1

The three rules for writing defect reactions... 2—1
Mass balance.. 2—1
Electroneutrality... 2—2
Ratios of regular lattice sites.. 2—2

Examples of reaction equations for defects.. 2—3
Stoichiometric compounds – instrinsic disorders .... 2—3

Nonstoichiometry .... 2—5
Dissolution of foreign elements .... 2—9

Ternary and higher compounds .... 2—14

Summary .... 2—16
Exercises .... 2—16

3. DEFECT EQUILIBRIA 3.1

Introduction 3.1

Point defects and equilibrium in an elemental solid - a simple statistical thermodynamic approach 3.2

Thermodynamics of chemical reactions 3.5

Thermodynamics and point defects 3.8
Virtual chemical potentials of point defects 3.8
Ideal and non-ideal models 3.9

Examples of defect equilibria in pure, stoichiometric metal oxides 3.10
Schottky defects 3.10
Frenkel defect pairs 3.12
Intrinsic ionisation of electrons. 3.12

Defect equilibria in nonstoichiometric oxides 3.14
Oxygen deficient oxides 3.14
Oxide with excess metal 3.17
Simultaneous presence of oxygen vacancies and interstitial metal ions 3.18
Metal-deficient oxide 3.20
Metal oxides with excess oxygen. 3.22

Examples of defect structures involving both oxygen deficiency and excess 3.22
Oxygen deficit. 3.23
Oxygen excess 3.23
Stoichiometric condition 3.24

Summary 3.26
Literature 3.27
Exercises 3.27

4. IMPURITIES AND DOPANTS 4.1

Introduction 4.1

Constant concentration of impurity/dopant 4.3
Schottky equilibria 4.3
Frenkel equilibria 4.4
Oxygen-deficient oxides doped with lower valent cations 4.5

Oxygen deficient oxides doped with higher valent cations 4.6
Oxides with excess metal 4.6
Metal-deficient oxides doped with lower valent cations 4.7
Metal deficient oxides doped with higher valent cations 4.8
Doping of oxides which may have regions with both oxygen and metal deficit 4.8

Variable concentration of impurity/dopant 4.11
Doping M_2O_3 with excess MlO 4.12
Doping M_2O_3 with excess of MhO_2 .... 4.13

Hydrogen defects in metal oxides 4.15
Effect of water vapour on oxygen-deficient M_2O_3. .... 4.17
Effect of water vapour on acceptor-doped M_2O_3 .... 4.18

Association between impurities/dopants and other point defects 4.21

Summary 4.22
Literature (protonic defects only) 4.22
Exercises 4.22

5. DIFFUSION 5.1

Introduction 5.1

Models of diffusion 5.1
Fick's first law 5.2
Potential gradients as the driving force 5.3
Example of chemical potential gradient acting on neutral particles 5.4
Simplified model for one-dimensional diffusion 5.5
Random diffusion 5.8

Fick's second law 5.10
Measurements of the diffusion coefficients by tracer techniques 5.12

Diffusion mechanisms 5.14
Vacancy mechanism 5.14
Interstitial mechanism 5.15
Interstitialcy mechanism 5.15
Diffusion of protons in metal oxides. 5.17

Further considerations of factors that affect the diffusion coefficient in crystalline solids 5.18
Free proton transport 5.20

Temperature and oxygen pressure dependence of diffusion in metal oxides 5.21
Dependencies related to the concentration of defects 5.21
Temperature dependence of the attempt frequency ω 5.22
Resulting analysis of the diffusion coefficient D_r 5.24

Diffusion coefficients of point defects **5.27**
Excercises 5.28

6. ELECTRICAL CONDUCTIVITY 6.1

Introduction **6.1**

Transport in an electrical potential gradient **6.1**

Charge carriers in ionic compounds **6.3**

The Nernst-Einstein relation between mobility and diffusion coefficient **6.4**

Electronic conductivity in oxides **6.7**
The band theory 6.8
Effects of charged impurities or defects. Extrinsic semiconductors 6.12
The Fermi level and chemical potential of electrons 6.17
Charge mobilities of electrons and electron holes 6.18
Nonstoichiometric semiconductors 6.21

Ionic conductivity **6.25**
Correlation effects: tracer diffusion and ionic conduction 6.26

Literature **6.28**

Exercises **6.29**

7. ELECTROCHEMICAL TRANSPORT 7.1

Electrochemical potential **7.1**

Flux equations – Wagner theory **7.1**
General derivation of Wagner-type flux expressions 7.1
From charged to neutral species: the electrochemical reaction 7.3
The voltage over a sample 7.4
Flux of a particular species 7.5

Wagner-type transport theory case studies **7.6**
Oxides with transport of oxide ions and electrons 7.6
Ambipolar diffusion and conduction 7.11

Chemical diffusion and tracer diffusion **7.12**
General case 7.12
Chemical diffusivity in the case of mixed oxygen vacancy and electronic conductor 7.14

Surface and interface kinetics limitations **7.16**

Ionic transport of both anions and cations **7.22**
General expressions 7.22
High temperature oxidation of metals; the Wagner oxidation theory 7.26

Sintering 7.34

Oxides with additional transport of protons 7.36

Other cases 7.37
Excercises 7.39

1. Structure and defects

Introduction

Classical chemistry and crystallography gave an idealised picture of the composition and crystal structure of inorganic compounds. During the 19th century crystallographers concluded that the atoms or structural units in crystals were arranged in ideal structures where all the structural sites were occupied. Chemists, on their side, expressed the compositions in terms of the law of definite proportions and generally believed that inorganic compounds had definite, unvarying, stoichiometric compositions determined by the valence of the constituent atoms. However, there was still considerable discussion regarding these concepts and it was alternatively proposed that inorganic compounds can have variable compositions. This was clearly established in the beginning of the 20th century and the question was theoretically resolved in the 1930's when Wagner and Schottky (1930) showed through statistical thermodynamic treatments of mixed phases that crystal structures are not ideal. Some lattice sites will be empty (vacant) and extra atoms may occupy the interstitial space between the atoms on the lattice sites. The empty lattice sites are termed vacancies and the extra atoms interstitial atoms. Following Wagner and Schottky all crystalline solids will at any temperature contain vacancies and extra atoms and will as such exhibit deviations from the ideal structure. Furthermore, all inorganic compounds may in principle have variable composition and thus be nonstoichiometric. In fact, the very exact stoichiometric composition is an exception rather than a rule and can under equilibrium conditions only be achieved at definite sets of temperature and activities of the components in a crystal. However, for many inorganic compounds the deviation from stoichiometry (the nonstoichiometry) is so small that it is difficult or impossible to measure directly with presently available experimental techniques.

These deviations from the ideal structures are present at any temperature and occur naturally in all crystalline compounds. Even so these deviations or imperfections are called *defects*. The reason for this is that *by convention the ideal structure is used as the reference state*, and *any deviation from this ideal state is termed a defect*.

Crystalline solids contain different types of structural defects. If the imperfection is limited to one structural or lattice site and its immediate vicinity, the imperfection is termed a point defect. Vacancies and interstitial atoms are point defects. An impurity atom present in a crystal and that either occupies a lattice site or an interstitial site is also termed a point defect. But in addition to the point defects the structural defects also comprise line and plane defects. The line defects are dislocations which are characterised by displacements in the structure in certain directions. The plane defects comprise stacking faults, grain boundaries, internal and external surfaces.

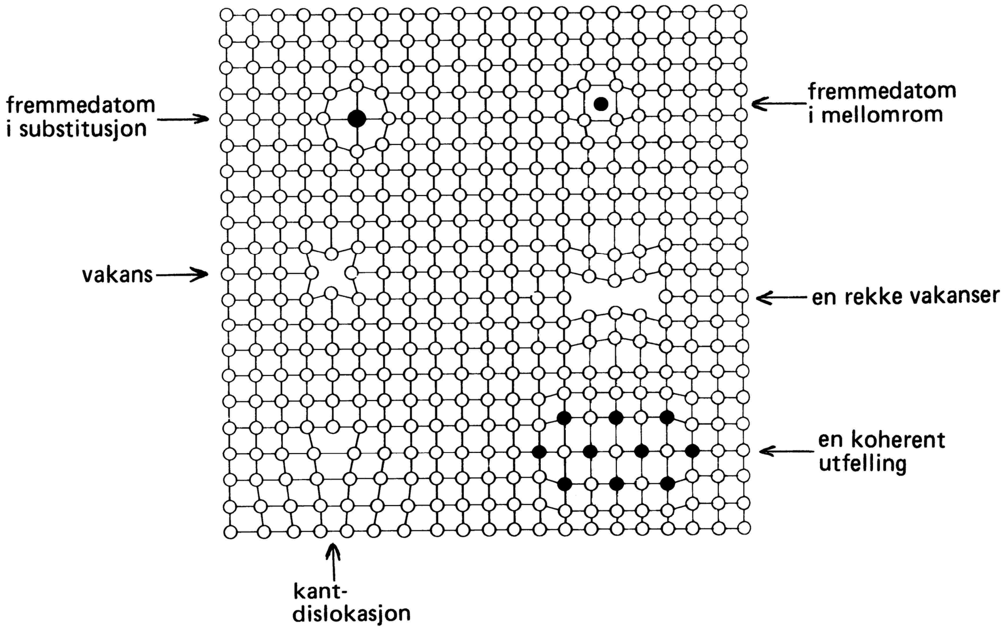

Figure 1-1. Schematic representation of some defects in an elemental solid. From A. Almar-Næss: Metalliske materialer.

In addition to the structural defects, crystals also contain electronic defects, i.e. electrons and electron holes that are relatively free to move in the crystal. The electronic defects may either be formed through internal excitation of valence electrons or they may be formed in association with point defects. If these electronic defects are localised (trapped) at regular sites in the structure, the electronic defects are termed polarons or - from a chemical point of view - valence defects. Defect electrons or electron holes trapped at point defects often make otherwise transparent materials coloured, and composite defects involving point defects and trapped electronic defects are termed colour centres.

A complete description of the point and electronic defects in a compound and their concentrations as a function of the activities (partial pressures) of the constituents and the temperature is often termed the *defect structure* of the compound.

The reason for our interest in and concern with defects and imperfections is that they determine a number of properties of crystalline solids. Thus diffusion or ionic conductivity in crystalline solids takes place because of the presence of defects. Point defects are responsible for lattice diffusion, dislocation diffusion takes place along line defects and grain boundary and surface diffusion along planar defects. Solid state diffusion, in turn, determines or strongly influences a number of properties or processes such as mass transport in solids, solid state reactions, sintering, high temperature creep, gas-metal reactions resulting in the formation of compact layers of reaction products, etc. Electronic defects determine properties such as electronic conductivity, electro-optical properties, etc. Heterogeneous catalysis is closely related to the defect chemistry of catalysts.

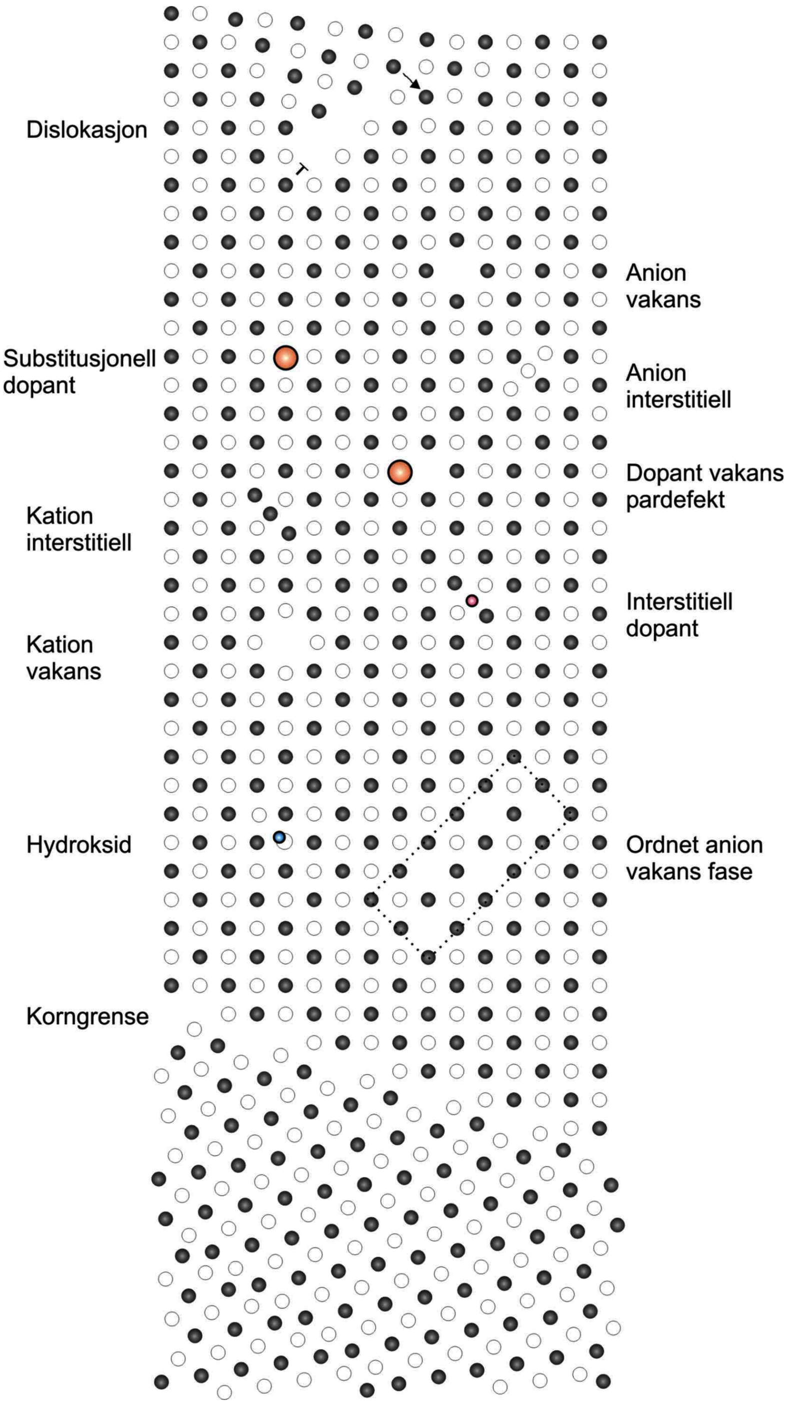

Figure 1-2. Schematic representation of some defects in an ionic binary compound.

As a basis for further considerations of defect structures and defect-dependent properties to be discussed later in the book, the remainder of this chapter briefly discusses some commonly occurring crystal structures in simple metal oxides, the notation of point defects, and qualitative descriptions of different types of defect structure situations.

Common crystal structures

Close-packing and voids

The crystal structures of many metals can be described by close-packing of spheres (atoms). Similarly, many simple oxides to be discussed later can conveniently be considered to consist of close-packing of oxygen ions while the metal ions occupy voids in the close-packed structures. Let us therefore briefly discuss close-packing of spheres (atoms or ions).

Consider an atom as a small hard sphere and make a layer of identical atoms so that the empty space between the atoms is minimum (layer A). Let us now add a second layer (B) such that the atoms in this second layer sit in one set of the hollows of the first layer. When we add a third layer of spheres, the spheres can be placed in two different positions: the spheres in the third layer can be placed directly over the spheres in layer A, and if we continue this stacking, the stacking sequence becomes ABABAB... as illustrated in (a) of the figure below. This type of close-packed stacking is called hexagonal close packing (hcp) of the spheres because it gives crystal structures with hexagonal symmetry.

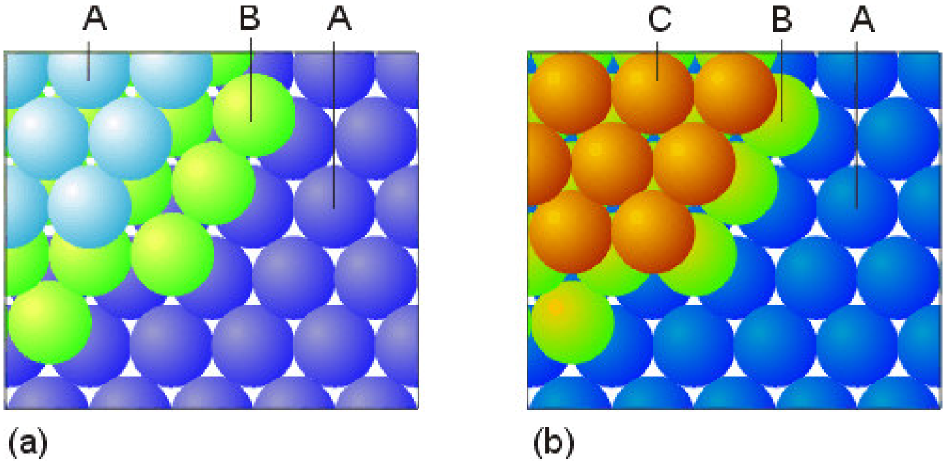

Figure 1-3. ABABAB...(hcp) clos-packing (a) and ABCABC...(fcc) close-packing (b) of spheres. From Shriver and Atkins; Inorganic Chemistry.

The third layer (layer C) could alternatively be placed such that the spheres in this layer will not be directly over either the atoms in the A or B layers, see (b) in the figure above. The stacking sequence is now ABCABC... and this close packing is known as cubic close packing (ccp) as this gives rise to a cubic (face-centred) crystal structure. The following figure shows how the two close-packed structures appear when their heaxonal and face-centred cubic unit cells are emphasized.

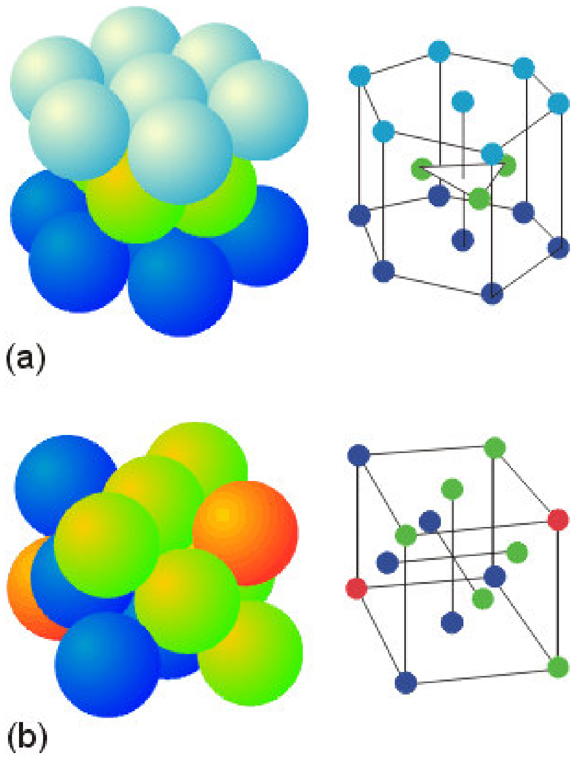

Figure 1-4. Schematic views of the hcp (a) and fcc (b) structure types. From Shriver and Atkins; Inorganic Chemistry.

The close-packed structures contain two different types of empty space or voids between the spheres, cf. the figures below. One type of void is termed a tetrahedral void because each of these voids is surrounded by four spheres which centres are at the corners of a tetrahedron. The other type of void is called an octahedral void. In this case each void is surrounded by six spheres, three in one layer and three in the neighbouring layer, and the centres of these six spheres lie at the corners of an octahedron. In the close-packed structures there are two tetrahedral voids and one octahedral void per atom (sphere). It may be noted that an octahedral void will fit a sphere with radius 0.414r where r is the radius of the bigger close-packed spheres. A tetrahedral void is smaller and may fit a sphere with radius 0.225r. In different crystal structures of simple metal oxides which can be considered to consist of a close-packing of oxygen atoms, the metal ions often occupy all or part of the tetrahedral and/or octahedral interstices in regular patterns.

Figure 1-5. Octahedral and tetrahedral hole in close-packed layers of speheres. From Shriver and Atkins: Inorganic Chemistry.

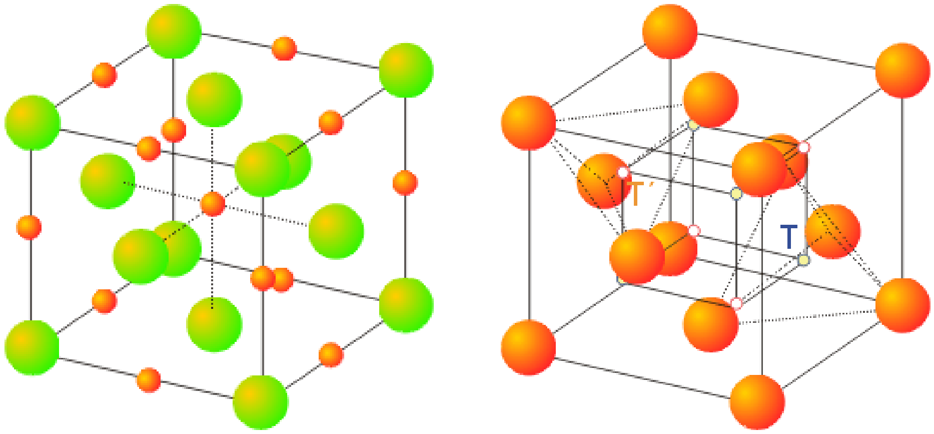

Figure 1-6. Small spheres show locations of octahedral (left) and tetrahedral (right) voids in fcc structure. From Shriver and Atkins: Inorganic chemistry.

Less close-packed arrangements

Some metals and ionic compounds choose for various reasons to arrange themselves in less close-packed structures, as exemplified in the following two figures.

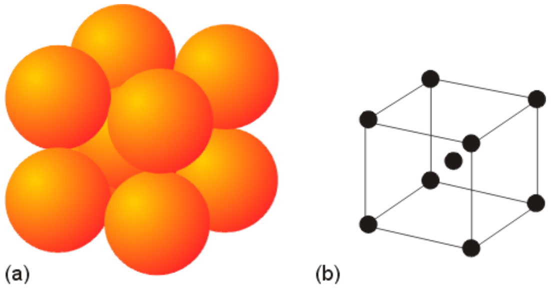

Figure 1-7. Body-centered cubic (bcc) packing of spheres. From Shriver and Atkins: Inorganic Chemistry.

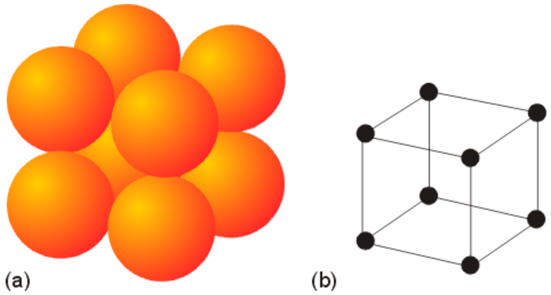

Figure 1-8. Simple cubic (sc) packing of spheres. From Shriver and Atkins: Inorganic Chemistry.

Some simple structures for oxides and other ionic compounds

As mentioned earlier, the structures of ionic compounds can often be seen as close-packing of the large anions, while the normally smaller cations occupy the interstitial voids. (Note that once the structure is established, the interstices occupied by cations are not any longer considered interstitial; they are part of the ideal (reference) structure.)

In judging the packing of ions it is useful to recall some important principles of ionic radii; the size of the elements increase down a group of the periodic table (resulting from the larger orbital of the outermost shell). Further, the size of the elements as a rule of thumb decreases from left to right through a period of the periodic table (resulting from increased nuclear charge). Finally, and most importantly in this context: Negatively charged ions (anions) are much larger than their neutral atoms and positively charged ions (cations) are much smaller than their neutral atoms. One may also recall that the effective size increases with the coordination number. In the following we look at some example structures possessed by oxides.

The NaCl (rock salt, halite) structure

In the NaCl structure the anions are cubic close-packed, and the smaller metal ions occupy all the octahedral interstices. Each metal ion is thus surrounded by six anions and vice versa. The NaCl structure is illustrated below. Note the similarity with the illustration of the octahedral voids in the fcc structure shown previously.

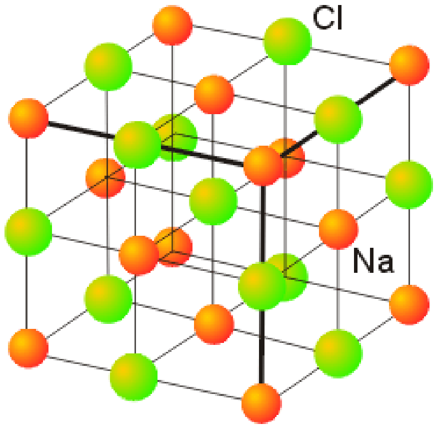

Figure 1-9. The NaCl structure. Note that the centre anion is six-coordinated, as are also all other ions. From Shriver and Atkins: Inorganic Chemistry.

The oxides MgO, CaO, SrO, BaO, CdO, CoO, NiO, FeO, TiO, NbO, and VO possess this crystal structure. The last six of these oxides may be highly nonstoichiometric and as such their NaCl structures highly defective.

In the sulfides FeS and NiS, the structure is similar, but the close-packing of anions is hexagonal rather than cubic.

The ZnS (zinc blende and wurtzite) structures

These structures are named after two different mineral forms of zinc sulphide (ZnS). The zinc blende structure can be considered as a cubic close-packing of sulphide ions with the zinc ions occupying every other tetrahedral void. Each zinc ion is thus tetrahedrally coordinated by four sulphide ions and vice versa.

The wurtzite structure is composed of hexagonal close-packed sulphide ions and where zinc ions occupy every other tetrahedral void. The metal oxides BeO and ZnO adopt the wurtzite structure.

The CaF_2 (fluorite) structure

The fluorite structure is named after the mineral calcium fluoride (CaF_2) and is illustrated below. This structure may geometrically be considered to constitute a close-packing of the metal ions while the anions occupy all the tetrahedral sites. The calcium ions are smaller than the fluoride ions and the fluoride ions are thus not able to fit into the interstices of the calcium ion array, but the model describes the relative positions of the ions. It is nevertheless more correct to view the structure as a simple cubic packing of anions (not close-packed), with cations filling every second cubic void, as also depicted below; each cation is surrounded by eight anions.

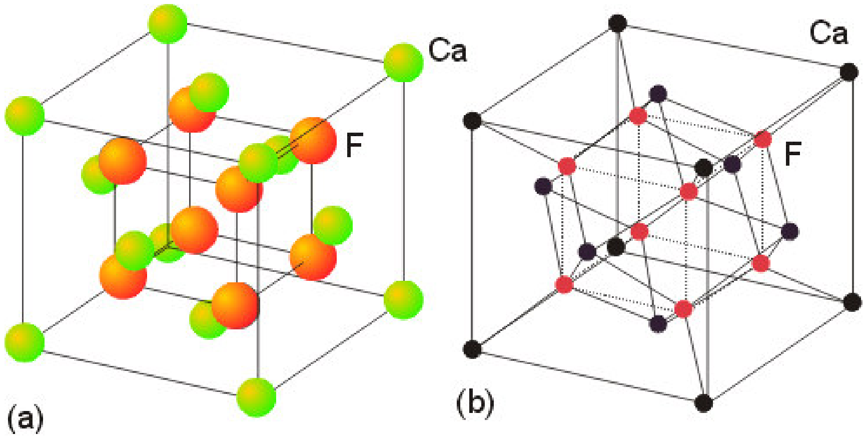

Figure 1-10. The CaF_2 structure. (a) The cations form an fcc sublattice, while the anions form a simple cubic arrangement. Every second anion cube has a cation in its centre. (b) The anions can be taken to occupy tetrahedral voids in the cation sublattic.

The fluorite structure is typical for AX_2 compounds where A is a sufficiently large cation that it forces the anions apart and into a less than close-packed arrangement. The oxides ZrO_2, HfO_2, CeO_2, ThO_2, PuO_2, etc. possess the fluorite structure. All the octahedral interstices are unoccupied for the stoichiometric compositions, but interstitial ions may be accommodated in these interstices and form oxides with excess oxygen (hyper-stoichiometric oxides), e.g. UO_{2+x}.

Pyrochlore ($ABO_{3.5}$) and C-type rare earth sesquioxide (A_2O_3) structures can be viewed as fluorite-derived structures with, respectively, 1/8 or 1/4 of the oxygen ions missing (in an ordered fashion).

The alkali metal oxides Li_2O, Na_2O, K_2O, and Rb_2O possess the so-called anti-fluorite structure. In this structure the oxygen ions can be considered to be cubic close-packed (fcc) while the metal ions occupy all the tetrahedral voids.

The rutile (TiO_2) structure

The rutile crystal structure has its name after the rutile modification of titanium dioxide, TiO_2. It is based on a hcp packing of the oxygen ions, with Ti ions occupying every second octahedral void, so that the structure can be viewed as consisting of TiO_6 octahedra (slightly distorted) which share edges and corners in such a way that each oxygen atom belongs to three neighbouring octahedra. The structure is tetrahedral, and when the structure is viewed along the c-axis it may be seen that the structure contains channels in the c-direction, i.e. rows of octaedra not filled with cations.

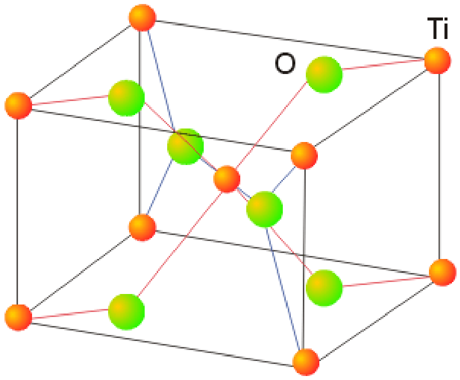

Figure 1-11. The rutile TiO_2 structure. From Shriver and Atkins: Inorganic chemistry.

In addition to rutile itself, the oxides SnO_2, MnO_2, VO_2, MoO_2, RuO_2, GeO_2 a.o. possess regular or distorted rutile structures.

The corundum (α-Al_2O_3) and ilmenite ($FeTiO_3$) structures

The corundum structure is named after the high temperature modification of alumina, α-Al_2O_3. This structure may be described as a hexagonal close-packing (hcp) of oxygen ions with the trivalent aluminium ions occupying 2/3 of the octahedral sites. As the metal ions occupy octahedral sites, each metal ion is octahedrally coordinated and surrounded by six oxygen atoms, while each oxygen atom is surrounded by four metal ions. Two and two metal ions occupy neighbouring interstitial octahedral sites, and the two corresponding AlO_6 octahedra are linked by common faces. Other oxides possessing the corundum structure are α-Fe_2O_3, Cr_2O_3, Ti_2O_3, and V_2O_3.

A number of oxides with the nominal formula ABO_3 have the corundum structure when A and B have an average valence of 3 and are of approximately the same size. The ilmenite structure with the nominal formula ABO_3 and named after the mineral $Fe^{II}Ti^{IV}O_3$ is similar to the corundum structure in that the oxygen ions can be described as hexagonal close-packed, the valences of the two cations have an average valence of 3 and occupy 2/3 of the octahedral sites. However, in this case the two metal ions (Fe^{2+} and Ti^{4+}) are arranged in alternate layers.

The spinel structures

The spinel structure is named after the mineral spinel, $MgAl_2O_4$. The spinels have the general formula AB_2O_4, where A is most often a divalent ion, A^{2+}, and B trivalent, B^{3+}. The oxygen ions can be considered to be cubic close packed (fcc), and the metal ions occupy both tetrahedral and octahedral sites. The unit cell contains 32 oxygen atoms, 8 AB_2O_4 units. It contains 32 octahedral and 64 tetrahedral voids. In the normal spinel structure one half of the octahedral voids are occupied by the B^{3+}-ion while the A^{2+} ions occupy 1/8 of the tetrahedral voids.

Other distributions of the cations occur. When the eight tetrahedral sites per unit cell are occupied by trivalent cations and the octahedral sites are shared by the divalent and trivalent ions, the structure is termed an inverse spinel. Magnetite, Fe_3O_4, is an example of an inverse spinel. An inverse spinel may be better written as $B(AB)O_4$ as this indicates that half of the B^{3+} ions occupy tetrahedral sites and the other half together with the A^{2+} ions occupy the octahedral sites. Another example of an oxide with inverse spinel structure is $Fe(MgFe)O_4$.

If the general formula of the spinel is written AB_2O_4, the cations may have the valence A^{2+} and B^{3+} as discussed above. This is termed a 2-3 spinel. In addition one may have spinels with other cation valences, e.g. A^{4+} and B^{2+} (4-2) spinel) and with A^{6+} and B^{1+}. In all cases the sum of the cation valences equals 8.

The perovskite ($CaTiO_3$) structure

The perovskite structure has its name after the mineral perovskite, $CaTiO_3$. The general formula may be written ABO_3 in which A is a large cation (e.g. alkali, alkaline earth, rare earth ions) and B a small or medium sized metal ion often a transition metal ion. In this structure the large A atoms and the oxygen atoms together form a close-packed lattice, such that each A-atom is coordinated to twelve oxygen atoms. The B-atoms occupy octahedral sites surrounded by six oxygen ions.

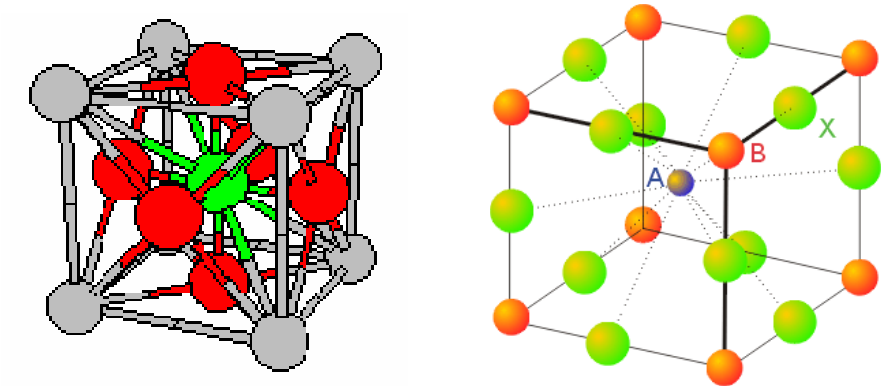

Figure 1-12. Perovskite ABX_3 structure. Left: Shown as a fcc packing of A (corners) and X (faces) atoms, with the B in the center. Right: The A cation is 12-coordinated.(The A cation is depicted small, but is in reality much bigger than the B cation). From Shriver and Atkins: Inorganic chemistry.

The average valence of the A and B ions is +3. The perovskite structure is remarkable in being very tolerant to variations (as long as A is large and B is small). Thus, A+B valences can be 3+3 (e.g., $LaAlO_3$), 2+4 (e.g. $CaTiO_3$), 1+5 (e.g. $KTaO_3$) and even 0+6 (where the A site is in fact empty, as in WO_3).

The perovskite structure tolerates relatively large nonstoichiometries and substitutions, while a number of perovskite-related structures (with perovskite building blocks) are formed when excesses, deficiencies or substitutions beyond the solubility limits lead to ordered defects and new phases. These comprise, for instance, oxygen deficiency (brownmillerite, $ABO_{2.5}$), and A excess (K_2NiF_4-type structure, A_2BO_4) and ordered intermediate phases.

Summary of relations between structures, radii and packing arrangements.

In oxides with small cations only, the structure is dominated by close-packing of the oxygen ions and various occupancies of interstices by the small cations. The table below summarises relations between structures and close-packed arrangements of anions or cations in oxides (and some sulphides).

Table 1-1. Some structures of oxides based on close-packed oxygen ion sublattices.

Formula	Cation:anion coordination	Type and number of occupied interstices	fcc of anions	hcp of anions
MO	6:6	1/1 of octahedral sites	NaCl, MgO, CaO, CoO, NiO, FeO a.o.	FeS, NiS
MO	4:4	1/2 of tetrahedral sites	Zinc blende: ZnS	Wurtzite: ZnS, BeO, ZnO
M_2O	8:4	1/1 of tetrahedral sites occupied	Anti-fluorite: Li_2O, Na_2O a.o.	
M_2O_3, ABO_3	6:4	2/3 of octahedral sites		Corundum: Al_2O_3, Fe_2O_3, Cr_2O_3 a.o. Ilmenite: $FeTiO_3$
MO_2	6:3	½ of octahedral sites		Rutile: TiO_2, SnO_2
AB_2O_4		1/8 of tetrahedral and 1/2 of octahedral sites	Spinel: $MgAl_2O_4$ Inverse spinel: Fe_3O_4	

When the cations get somewhat larger they force the anions apart, and we get non-close-packed structure types like the fluorite and its derivatives (pyrochlore, rare-earth oxide structures, etc.)

When we involve even larger cations, being of the same size as the anions, the large cations take part in the close-packing together with the anions. In this way we get the perovskite structure and its derivatives.

More detailed treatments on crystal structures of oxides are found in general texts on solid state chemistry (e.g. Smart and Moore (1992) or West (1984)) and in more specialised literature, such as Náray-Szabó (1969) and Rao and Raveau (1995).

Point defects

In order to understand and describe properties of inorganic compounds that are dependent on the presence of point defects, it is necessary to be able to express the concentrations of the defects under various conditions. This, first of all, requires a system of notation to describe the point defects and the constituent atoms of the inorganic compounds and their charges.

Notation for description of point defects

In the development of the field of defect chemistry of inorganic compounds various systems of notation have been proposed and used to describe point defects. However, the most widely adopted system is that due to Kröger and Vink (1956) (see also Kröger (1964)), and this will be used in this book. This system describes crystals in terms of structural elements, and an imperfection is indicated by a major symbol describing its chemical content and a subscript that indicates the site that it occupies.

In an oxide the metal ions on the regular lattice positions are written M_M, where the subscript thus describes the type of lattice site which is occupied. Correspondingly, the oxygen ions on normal lattice sites are written O_O.

The native point defects in an oxide include M and O vacancies and M and O interstitial atoms or ions. Vacancies are denoted by v with a subscript M or O referring to vacant metal or oxygen sites, respectively. Interstitial ions or sites are described with a subscript "i". Vacancies and interstitials in an oxide are thus written

v_O oxygen vacancy,

v_M metal vacancy,

O_i oxygen interstitial,

M_i metal interstitial.

Correspondingly, an unoccupied or vacant interstitial site can be written v_i.

While the notation traditionally has used V for vacancy, confusion with the element vanadium (V) can be avoided by using lower-case v or italic *V*. Here, we will use the lower case v.

Component atoms of a crystal may find themselves on the wrong site; the A and B cations in perovskite, pyrochlore, spinel, etc. may thus swap sites so as to obtain A_B and B_A. An oxygen atom in a cation site is similarly written O_M, but is an example of a very unstable defect.

Real crystals always contain impurities. When a foreign metal ion Mf occupies a regular M site, the foreign atom is denoted by Mf_M. If it occupies an interstitial site, it is described by Mf_i.

Notations for other or more complex defects will be introduced when needed.

Notation for charged defects

In an ionic compound the atoms are charged, and the cations and anions may be assigned a definite integer valence (formal oxidation number). This is in fact a fully valid model in defect chemistry, as will be argued in a discussion on the influence of the varying degree of ionicity, or distribution of valence electrons, in a later paragraph.

In the following, oxygen ions on regular sites in an oxide will be considered to have a formal valence of -2. The cations have a positive charge so as to make the sum of all positive and negative charges in the compound equal to zero.

The point defects may be neutral or charged. Their charges may be described in terms of their actual charge or valence. However, it is generally more convenient in writing defect reactions to use the ideal, perfect crystal as the reference state and consider the charge on the defects relative to this perfect crystal. This relative charge is termed the effective charge of the defect. With the perfect crystal as reference, the normal atoms on regular lattice sites have zero effective charge. To emphasise this zero effective charge, an "x" is often added as a superscript to the symbol. The normal cations and anions in an oxide are then written M_M^x and O_O^x.

When vacancies are present, part or all of the actual charge of the missing atom may be excited away from the vacant site. Let us, for instance, consider the formation of an oxygen vacancy by removing an oxygen atom from the crystal. If the two electrons of the oxygen ion are localised at or near the empty site, the vacancy and its immediate neighbourhood have two negative actual charges. The charge at this location is thus the same as in the perfect crystal and the oxygen vacancy then has zero effective charge. As such it is termed a neutral oxygen vacancy, which is written v_O^x.

Now that the electronegative oxygen atom is no longer present to hold the extra electrons, one or both of the two electrons may be excited and transferred away from the oxygen vacancy, and the oxygen vacancy then becomes singly or doubly charged. Since electrons are removed, the ionised vacancy has an effective positive charge relative to the perfect crystal, compensated by the two electrons which are now negatively charged defects somewhere out in the crystal.

To distinguish effective from actual charges, different symbols are used. While the actual charges are written + or -, the effective positive charge is indicated by a supscript dot ($^\bullet$) and the effective negative charge by a superscript prime ($^/$). Singly and doubly charged oxygen vacancies are, for instance, written $v_O^\bullet$ and $v_O^{\bullet\bullet}$, respectively.

Similarly, cation vacancies may be neutral or have negative effective charges. To illustrate this let us remove a metal atom from the oxide MO and create a vacant metal ion site. The M^{2+} ion picks up two electrons to leave as an atom, leaving behind two positive charges in the form of holes or valence defects. If these are localised at neighbouring oxygen ions (as O^-) or metal ions (as M^{3+}) the actual charge at the vacancy and its immediate neighbourhood is then the same as in the perfect crystal, and the effective charge on the metal vacancy is considered to be zero, i.e. the metal vacancy is neutral, v_M^x. One or both of the positive charges may be excited and transferred away from the vacancy and distributed in the oxide, and the vacancy then has one or two negative effective charges and is written as $v_M^{/}$ or $v_M^{//}$, respectively.

Similarly, interstitially dissolved atoms will have zero real and effective charge, but depending on their nature they can be ionised to cations with a real and effective positive charge or to anions with real and effective negative charge.

Substitutionally dissolved foreign ions which tend to have a valence different from that of the normal ions (aliovalent foreign ions) are easily ionised to get an effective charge. For instance, if the valence of a substitutionally dissolved foreign cation is higher than that of the normal host cation, the foreign cation gets a positive effective charge. Conversely, if the valence of the substitutionally dissolved foreign ion is smaller than the host cation, the foreign cation gets a negative effective charge. Common examples are found in the semiconductor field. Boron dissolved in silicon is denoted B_{Si}^x in the unexcited state, where we have B^0. But since boron has one valence electron less than the surrounding Si^0 atoms it easily accepts one from the valence band of the Si lattice and becomes B^- or, in Kröger-Vink notation, $B_{Si}^{/}$. Boron is called an electron acceptor (acceptor dopant) and the lacking electron in the valence band is called a hole. In a similar way, phosphorous forms a donor in silicon; P_{Si}^x and the excited state is P_{Si}^x.

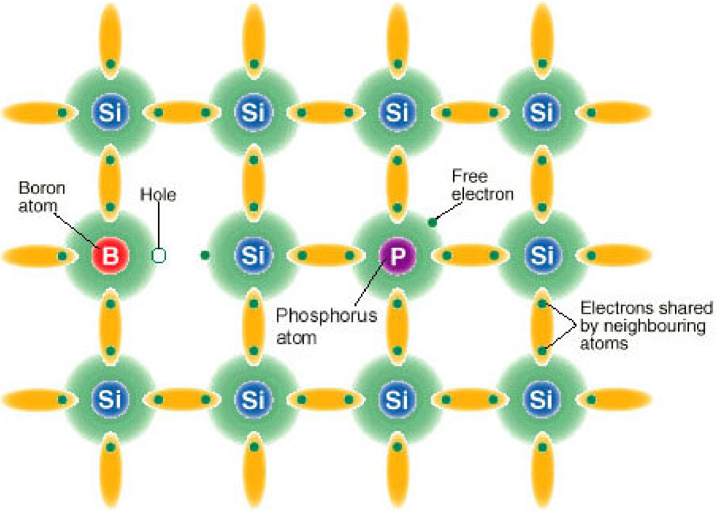

Figure 1-13. Schematic view of electronic structure of boron- and phosphorous-doped silicon. From http://acre.murdoch.edu.au/refiles/pv/text.html

Defect electrons and electron holes that are free to move in the oxide have effective negative and positive charges, respectively. They are written $e^{/}$ and $h^{\bullet}$. If the electron, for instance, is associated with a cation on a regular site - and may as such be considered a valence defect - the defect may be written $M_M^{/}$.

We have mentioned but a few types of defects to illustrate the principles of notation. Many other defects can be described in the same way; we will see more defects later on.

In addition to the notation, the important thing to remember about charges is the following: *The effective charge of a species is the real charge of the species minus the real charge the perfect reference lattice would have had in the same volume element.*

Electroneutrality

Crystals that we will deal with are considered to be electrically neutral. Then the sum of all positive charges must equal the sum of all negative charges:

Σ positive charges = Σ negative charges

We may thus sum up the real charges of all cations and anions and defects. However, it is equivalent and more convenient to consider only the effective charges:

Σ positive effective charges = Σ negative effective charges

It is important to note that the number of positive and negative charges must be compared on equal bases, namely as concentrations (numbers or moles per volume). One may in general not use site fractions since the concentrations of different types of sites are not necessarily equal. Thus, the electroneutrality condition can be expressed as

$$\sum_S z[S^z] = 0 \tag{1.1}$$

where the brackets denote concentration and z is the number and sign of charges on the species S^z. For instance, in an oxide MO where the major defects are oxygen vacancies, metal interstitials, and defect electrons, the electroneutrality condition can be written

$$2[v_O^{\bullet\bullet}] + 2[M_i^{\bullet\bullet}] - [e^/] = 0 \quad \text{or} \quad 2[v_O^{\bullet\bullet}] + 2[M_i^{\bullet\bullet}] = [e^/] \tag{1.2}$$

Despite its simplicity, and despite it's being used in e.g. traditional aqueous chemistry, the electroneutrality condition tends to cause both conceptual difficulties and occasional errors (by misplacing the coefficients z). In order to help, we suggest that one learns it by heart and uses it accordingly. Secondly, that one realises that it is *not* a chemical reaction, but a mathematical relationship. It thus maintains no mass balance or charge balance, it simply counts positive and negative charges from whatever concentrations of charged species that are present and requires them to be equal.

Stoichiometry and nonstoichiometry

The compositions and chemical formulae for metal oxides and other inorganic compounds are usually written with a definite ratio of cations to anions, e.g. M_aO_b where a and b are usually small integers determined by the valence of the constituent atoms. In crystalline compounds this also reflects that the structure contains different types of sites (e.g. close-packed sites and tetrahedral or octahedral interstices) in simple ratios and that these are selectively and systematically filled with cations or anions. When the oxide M_aO_b contains M and O atoms in the exact ratio a:b, it is said to have a stoichiometric composition.

As mentioned above, an exact stoichiometric composition in inorganic compounds is in principle the exception rather than the rule. Oxides in equilibrium with their surroundings are generally nonstoichiometric, except under specific conditions of temperature and activities of the components. However, within our experimental ability to measure the ratio of the constituent atoms, many inorganic compounds may be considered to be so near stoichiometry over large temperature and activity ranges that minor deviations from stoichiometry may often be neglected in a discussion of defect concentrations and defect-controlled properties.

Numerous oxides will exhibit marked deviations from stoichiometry. Some oxides are even unstable when stoichiometric. Wustite, for instance, which is nominally written FeO, has under all conditions a deficiency of iron which varies with temperature and oxygen activity, and its formula should correspondingly be written $Fe_{1-y}O$. In many systems of the transition metal oxides there exist phases which show large variations from the apparent stoichiometric composition in that the average valences of the metal ions in these phases are intermediate between the integral valences of the metal ions. These arise due to the elimination of defects in the form of shear planes. One often finds homologous series of oxides of this kind, with small differences in average valence and composition. This will be further discussed below.

In the following is given a qualitative description of different defect structure situations. The corresponding defect equilibria that express the concentration of the different point defects as a function of temperature and oxygen activity are treated in a later chapter.

Defects in stoichiometric compounds.

Let us consider a stoichiometric crystal with composition MX. If a charged point defect is formed in such a crystal, a complimentary point defect with opposite effective charge must be formed to conserve the electroneutrality of the stoichiometric crystal. Two types of defect structures have been found to be important in stoichiometric metal oxides and these are termed Schottky and Frenkel defects, respectively, honouring early contributions of two of the many German scientists who pioneered the development of defect chemistry (Schottky (1935), Frenkel (1926)).

Schottky disorder

A stoichiometric crystal with Schottky disorder contains equivalent concentrations of cation and anion vacancies. A stoichiometric oxide MO contains equal concentrations of metal and oxygen ion vacancies. In a stoichiometric oxide MO_2 the concentration of oxygen vacancies is twice as large as that of the metal ion vacancies. It may be noted that the formation of Schottky defects can only occur at outer and inner surfaces or dislocations and will diffuse into the crystal until equilibrium is reached.

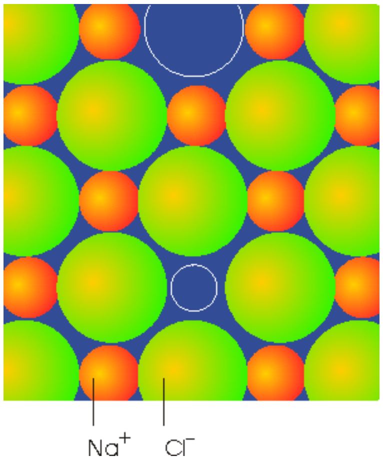

Figure 1-14 Schottky defect pair in NaCl. From Shriver and Atkins: Inorganic Chemistry.

Frenkel disorder

A stoichiometric crystal with Frenkel disorder contains the same concentrations of metal vacancies and metal interstitial ions. Contrary to the Schottky defects, Frenkel defect pairs can be formed directly inside the crystal.

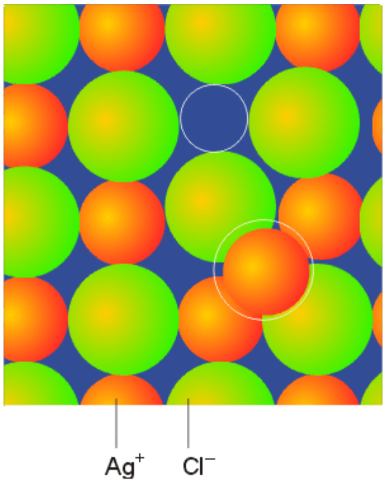

Figure 1-15. Cation Frenkel defect pair in AgCl. From Shriver and Atkins: Inorganic Chemistry.

Although Schottky and Frenkel disorder may be simultaneously present in stoichiometric compounds, one type of disorder usually predominates. As a rough rule

Schottky disorder is favoured in crystals where the cations and anions are of comparable size, while Frenkel disorder predominates when the sizes of the cations and anions are appreciably different. Another factor is that Schottky disorder tends to dominate when the structure is very effectively packed so that the interstitials that are part of Frenkel pairs are hard to form.

Pairs of interstitials have not been found as a dominant defect structure. Anti-site disorder (swapping of sites) is common in intermetallics and between cations on different sites (or anions on different sites). However, swapping between cations and anions in ionic compounds is mostly prohibited by the very unfavourable effective charges that would arise.

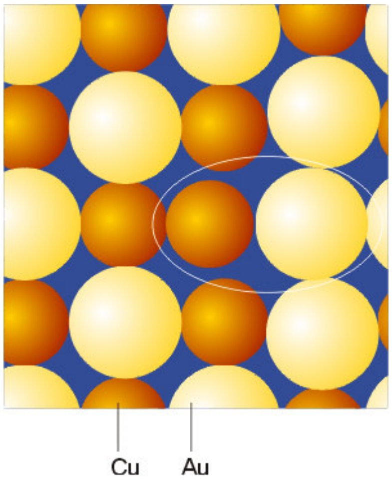

Figure 1-16. Anti-site defect pair in CuAu intermetallic. From Shriver and Atkins: Inorganic Chemistry.

Defects in nonstoichiometric oxides.

The deviation from the exact stoichiometric composition is directly related to the presence of point defects. While complimentary point defects are formed in stoichiometric crystals, the electrical neutrality of nonstoichiometric compounds is conserved through the formation of point defects and charge compensating electronic defects.

Nonstoichiometric oxides may - depending on the oxide, temperature and activities of the components - have an excess or deficit of metal or oxygen. In view of this, nonstoichiometric oxides may be divided in four limiting groups:

- metal deficient oxides, e.g. $M_{1-y}O$, where y is usually a small fraction of 1. Metal vacancies or complex defects based on metal vacancies are the majority defects, and examples with oxides with metal deficit are $Co_{1-y}O$, $Ni_{1-y}O$, and $Fe_{1-y}O$;
- metal excess oxides, e.g. $M_{1+y}O$, in which metal interstitials are the prevalent defects. $Cd_{1+y}O$ is an example of an oxide with this type of nonstoichiometry;

- oxygen deficient oxides, e.g. MO_{2-y}. In these oxides oxygen vacancies prevail and examples of such oxides are CeO_{2-y} and other oxides with the fluorite structure;

- oxygen excess oxides, e.g. MO_{2+y} such as, for instance UO_{2+y}. Oxygen interstitials or complex defects based on these are the predominating defects.

Although a particular type of defect predominates in an oxide and as such constitutes the majority defect in the crystal, it is important to emphasise that the crystal will contain all other defects in varying concentrations in the form of minority defects. Furthermore, the majority defects in the crystal may also change with temperature or activity of the components in the crystal. By way of example, in an oxide MO_2 the predominant defects may in principle change from oxygen vacancies to interstitial metal ions on changing the oxygen activity, and under these conditions the nonstoichiometry may be written $M_{1+x}O_{2-x}$.

Associations of point defects, point defect clusters, extended defects

Point defects are often assumed to be present as single, unassociated point defects that are randomly distributed in the structure. However this is in principle only applicable for small defect concentrations, that is, typically when the fraction of defects (relative to the number of normal atoms in the crystal) is smaller than 0.001-0.01 (but depending a lot on temperature).

For larger defect concentrations one expects that defects begin to interact. The interactions can be coulombic or mechanical and be attractive or repulsive. These interactions may change the activity coefficients and formation enthalpies of defects, and they may lead to associations between defects which attract each other. A simple example is the association of a native point defect and a dopant or impurity ion with opposite effective charges. Energy minimisation may for large defect concentrations lead to clusters of different defects or elimination of defects in lines or planes, so as to form a new structure (e.g. shear structures).

Order or disorder of defects is a matter of a fine balance between the configurational entropy of the defects and the exothermic enthalpy gained in association of certain point defects to form various types of larger defect agglomerations. Systems with relatively small defect concentrations are entropy controlled and consist of randomly distributed point defects. As concentration increases the enthalpy takes control and leads successively to formation of larger defect entities; randomised associated defects or defect clusters, point defects assimilated into randomised structure elements of a new phase, superlattice ordering and extended defects, shear planes, and discrete intermediate phases.

Example cases of defect clusters

Defect clusters in a metal-deficient oxide: $Fe_{1-y}O$

A well known example of an oxide where the point defects form defect clusters is wustite, $Fe_{1-y}O$. As mentioned above, under normal equilibrium conditions the oxide is never stoichiometric; the value of y ranges from 0.05 to 0.15 depending on the temperature and oxygen activity. Following the composition of the oxide it was originally believed the point defects were iron vacancies. However, extensive structural investigations by x-ray, electron and neutron diffraction have over the years shown that the defects not only comprise Fe-vacancies but also Fe-interstitials. The earliest studies of complexes indicated that for each vacancy created as a result of nonstoichiometry, a Frenkel defect pair was also formed. The combination of these defects gave a complex defect consisting of an interstitial ion (on a tetrahedral site) adjoined by two vacancies (vacant octahedral sites). Further studies have shown that the complexes comprise combinations of larger number of vacancies and interstitials with ratios of vacancies to interstitials ranging from 3 to 4. The basic unit of such clusters consists of four vacancies and one interstitial as illustrated below. Such units may further combine to form more complex clusters through side-, edge-, and corner-sharing and for these the ratios of vacancies to interstitials are close to 3. It is noteworthy that the larger corner-sharing clusters have the same local atomic arrangement as magnetite, Fe_3O_4, and these can be considered as nuclei of Fe_3O_4 to which $Fe_{1-y}O$ transforms when the defect concentrations exceed critical values. The defect cluster in question might be denoted $(4v_{Fe}Fe_i)^{6/}$.

Defect clusters in an oxide with excess oxygen: UO_{2+x}

Uranium dioxide has the fluorite structure and may have a deficit or excess of oxygen depending on the temperature and oxygen activity. Under conditions when it has excess oxygen, UO_{2+x}, the maximum value that x may have is 0.25. This composition may suggest that the predominant point defects are oxygen interstitials, however, as shown by Willis through neutron diffraction studies the defects are not single unassociated oxygen interstitials, but comprise clusters of oxygen interstitials and vacancies. The real composition of the clusters and detailed location of the individual point defects in the clusters are still a matter of discussion, but the formation of clusters may qualitatively be visualised as follows: When an oxygen atom is dissolved interstitially an oxygen atom on a neighbouring normal site is simultaneously displaced to neighbouring interstitial site leaving behind a vacant normal oxygen site. This qualitative model suggests a defect cluster consisting of two oxygen interstitials and an oxygen vacancy. Later studies have shown that clusters consist of larger number of oxygen interstitials and vacancies.

Extended defects, shear structures.

Extended defects are planar defects and may be considered to be formed by ordering and elimination of point defects, e.g of oxygen vacancies, along specific crystallographic planes. This process may be illustrated below. The predominant defects are in this case assumed to be oxygen vacancies and are considered to be aligned as illustrated. By shear of these crystals the aligned vacancies are eliminated, and this produces a fault in the cation sublattice. The formation of shear phases gives rise to the homologous series of oxide structures, e.g W_nO_{3n-1} and Ti_nO_{2n-1}. In this way the crystals may be able to have variable composition and may as such be termed nonstoichiometric.

The schematic depictions of defective lattices shown initially in this chapter contain some other extended defects, like dislocations, aggregates and precipitates. Contrary to point defects, extended defects can often be imaged directly in electron microscopy, as seen in the figure below.

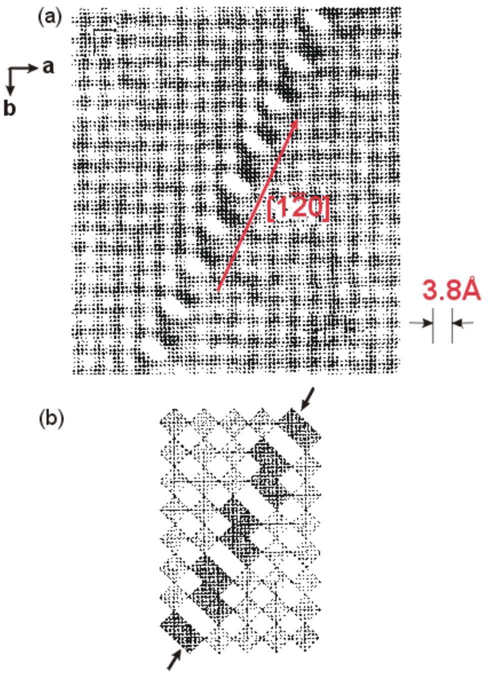

Figure 1-17. Real and simulated image of shear plane in WO_{3-x}. From Shriver and Atkins: Inorganic Chemistry.

Modulated structures/infinitely adaptive structures

Some oxide systems may accommodate nonstoichiometry by continuous adjustment of a parent structure without forming point defects. Such structures are now called modulated nonstoichiometric compounds, but these have also in the literature been termed vernier structures or infinitely adaptive structures. An example of such an oxide is the low-temperature modification of tantalum pentoxide, L-Ta_2O_5. The oxide exchanges oxygen with the ambient atmosphere, but the monophasic structure is maintained through the creation or elimination of oxygen sites in the oxygen sublattice while the number of tantalum sites remains unchanged.

Summary

Due mainly to the gain in enthalpy of electrostatic attractions and electron orbital overlap, solids usually organise their constituent ions, atoms, or molecules in a crystal lattice. In such lattices the species take specific positions relative to each other, and this pattern is repeated in 1, 2 or 3 dimensions. Different compounds take on different geometrical arrangements depending on the properties of the constituents. This is a complex function of the charge distribution of each atom or ion, in turn a function of nuclear charge, ionisation energies, electron affinity and shape of the valence electron orbitals. However, it is often useful to view the structures as close-packing or other packings of rigid spheres, with systematic filling by smaller spheres of voids so created. Because of the preferred occupation of specific sites in the lattice, together with the tendency of distinct preferences in valence for the various elements, most compounds are basically stoichiometric, that is, the ratio of the different atoms is a simple ratio of small integers. In competition with this minimisation of enthalpy with perfect, stoichiometric lattices, we have the tendency of increasing the entropy at $T > 0$ K and this is done by the creation of defects. At any temperature there is thus an equilibrium concentration of defects. The defects comprise point defects (vacancies, interstitials and substitution), electronic defects (conduction electrons, holes, valence defects), and associates, clusters, lines or planes of defects. The overall defect structure is a result of energy minimisation by the enthalpy and entropy of formation and mixing, the electroneutrality condition (charge balance), mass balance, and preservation of the structure (site ratio balance). A defect is denoted by its constituent (a chemical element symbol or a vacancy (v)), a subscript indicating lattice site or interstitial site (i) and a superscript indicating effective positive or negative charge ($^{\bullet}$ or $^{\prime}$).

Literature

Frenkel, J. (1926) *Z. Phys.*, **35**, 652.

Kröger, F.A. (1964) *The Chemistry of Imperfect Crystals*, North-Holland, Amsterdam, and Wiley, New York.

Kröger, F.A. and Vink, H.J. (1956) in *Solid State Physics* (eds F. Seitz and D. Turnbell) **3**, 307, Academic Press, New York.

Náray-Szabó, I. (1969) *Inorganic Crystal Chemistry* Akadémiai Kiadó, Budapest.

Rao, C.N.R. and Raveau, B. (1995) *Transition Metal Oxides* VCH, New York.

Schottky, W. (1935) *Z. Phys. Chem.*, **B29**, 335.

Shriver, D.F., and Atkins, P.W. (1999) *Inorganic Chemistry, 3rd ed.* Oxford Univ. Press., Oxford.

Smart, L. and Moore, E. (1992) *Solid State Chemistry; an Introduction*, Chapman & Hall, London.

Wagner, C. and Schottky, W. (1930) *Z. Physik. Chem.*, **B11**, 163.

West, A.R. (1984) *Solid State Chemistry and Its Applications*, Wiley, New York.

Exercises

1. Translate the defect names used in Fig. 1-1 and 1-2 into your first (or second) language.
2. Calculate the number of atoms in the bcc and sc unit cells of Figs. 1-7 and 1-8. (If you are not familiar with the procedures: Count all atoms, but divide by 2, 4, or 8 for atoms that are shared with neighbouring cells by being in faces, edges or corners.)
3. Calculate the number of atoms in the hcp and fcc unit cells of Figs. 1-4 (a) and (b).
4. Calculate the number of ions of each kind in the NaCl and ABX_3 cells in Figs. 1-9 and 1-12.
5. Calculate the number of atoms and voids in the cells in Fig. 1-6.
6. Calculate the volume filling ratio of balls in the structures of Figs. 1-4 (b), 1-7, and 1-8. What would it be in 1-4 (a)?
7. Why would you say are metals generally heavier (have higher densities) than ionic compounds (salts and ceramics)?
8. Sketch the periodic table of the elements, but include where possible the typical binary oxides of the elements in their normal (most common) oxidation states, in terms of formulae (stoichiometry) and structure type. Indicate also, based on what you know about their tendency to take on neighbouring oxidation states, whether the oxide is expected to be stoichiometric, have oxygen-deficiency, or oxygen-excess.
9. List the main types of 0-, 1-, 2-, and 3-dimensional defects in crystalline solids.
10. Write the Kröger-Vink notation for the following fully charged species in MgO: Cation and anion on their normal sites, oxygen vacancy, magnesium vacancy, interstitial magnesium ion.
11. Write the Kröger-Vink notation for the following species in ZrO_2: Cation and anion on their normal sites, oxygen vacancy, zirconium vacancy, yttrium dopant substituting Zr, interstitial carbon atom, cluster of yttrium dopant and oxygen vacancy, nitrogen ion (N^{3-}) substituting for oxygen ion.
12. Write the Kröger-Vink notation for the following fully charged species in $CaTiO_3$: Calcium vacancies, titanium vacancies, oxygen vacancies, Ti ions on Ca sites and vice versa, Ti interstitials.
13. Write the Kröger-Vink notation and for the possible species and defects in a material of your choice – preferably a compound you are working with.

14. For the materials and list of defects in Exercises 10-13, write the electroneutrality condition.
15. Write the electroneutrality condition for defects in boron-doped silicon. Write the electronuetrality condition for defects in phosphorous-doped silicon. Write the electroneutrality condition for pure (undoped) silicon.
16. Suggest an electroneutrality condition for Y-doped ZrO_2.
17. Write an electroneutrality condition for MO_{1-x} (hint: includes an oxygen defect type and an electronic defect type).
18. Write an electroneutrality condition for MO_{1+x}.
19. Write an electroneutrality condition for $M_{1-x}O$.
20. Write an electroneutrality condition for $M_{1+x}O$.

Answers to selected exercises, Ch. 1

1. -
2. 2, 1
3. 6, 4
4. 4:4, 1:1:3
5. 4:4, 4:8
6. 1-4b: $\frac{\pi}{3\sqrt{2}}$=0.74, 1-7: $\frac{\pi\sqrt{3}}{8}$=0.68, 1-8: ______, 1-4a: $\frac{\pi}{3\sqrt{2}}$=0.74.
7. -
8. -
9. -
10. -
11. -
12. -
13. -
14. -
15. -
16. -
17. -
18. -
19. -
20. -

2. Defect reactions

Introduction

As discussed in the previous chapter several different types of point defects may be formed in metal oxides and other inorganic compounds. In principle, all types of defects will be present, but in general, only a small number of different defects will predominate.

When defect-dependent properties are to be interpreted, it is important that the defect concentrations as a function of temperature and the activities of the crystal components are known. For this purpose it is necessary to formulate reactions and write equations for the formation (or annihilation) of the defects.

The defect formation may either occur internally in the solid or through reactions with the environment. In the following, the rules for formulating defect reactions will be described and applied to different defect structure systems, while in the next chapter, conditions for equilibrium and equations relating equilibrium defect concentrations to temperature, activities (partial pressures) of the components in a compound, impurity concentrations, etc., will be discussed.

The three rules for writing defect reactions

From a thermodynamic point of view a solid containing point defects constitutes a solid solution where the point defects are dissolved in the solid. In analogy with liquid solutions, the solid may be considered to be the solvent and the point defects the solute. Similarly, the defect equilibria may be treated in terms of the thermodynamics of chemical reactions and solutions.

In analogy with reactions in aqueous solutions, the rules for writing defect reactions include requirements of electroneutrality and mass balance. But for defect reactions in a crystalline compound it is also necessary to require that the ratio of regular lattice sites in the metal and oxygen sublattices must be maintained, even if the total number of sites may increae or decrease during the reaction. These rules may be summarised as follows:

Mass balance

The defect reaction must balance with respect to the mass, i.e. the number and types of atoms involved in the defect reaction must be the same before and after the defect formation or annihilation. Vacancies, which only represent empty sites, have zero mass and do not count. Also electronic defects are considered not to count in the mass balance.[1]

Electroneutrality

The compounds are and should remain electrically neutral. With the perfect crystal as reference, this means that the total effective charge is the same before and after the formation of the defects.

This means that the net charge on the left and right hand sides of a reaction equation must be the same. This charge may be counted in terms of effective or real charges. However, effective and real charges may in general not be balanced against each other, and one must therefore avoid using both effective and real charges in the same equation.

Ratios of regular lattice sites.

The ratio(s) of the number of regular cation and anion lattice sites in a crystalline compound is constant. For instance, in an oxide MO the ratio of regular M and O sites is 1:1 regardless of whether the actual composition is stoichiometric or nonstoichiometric. Correspondingly, in a compound M_2O_3 the ratio of regular cation to anion sites is 2:3, and if 3 oxygen sites are created in M_2O_3 through a defect reaction, two regular M-sites - vacant or filled - in the metal sublattice must also simultaneously be created.

This rule is special to defect chemistry in crystalline solids, and requires special attention and some practice to avoid its pitfalls. As is discussed below, this rule does not apply to oxides with infinitely adaptive structures.

Despite the requirement of the constancy in the ratio of regular sites, the total number of regular lattice sites may change in a defect reaction, and therefore, the defect equation may include the formation or annihilation of lattice sites as long as the proper ratios are maintained.

It should be noted that no sites are created in the formation of electronic defects.[2]

In the many defect equations and equilibria which will be discussed, surface atoms will not be considered separately. All atoms will be considered to be bulk atoms, and this means that the treatment only applies to sufficiently large

---

[1] One may choose to count and balance also the mass of electrons, but then one must be sure to include also non-defect electrons (electrons in the valence band).

[2] One may choose to operate with sites – or energy levels - for electronic defects, and one must then take into account the emty levels of the conduction band and the occupied levels of the valence band.

crystallite sizes for which the number of surface atoms is insignificant to that of the bulk atoms.

Defect reactions generally do not result in significant changes in the number of surface atoms. In the formation of a vacancy, for instance, a bulk atom is transferred to the surface, but in the same process a previous surface atom, in turn, becomes a bulk atom. (It is worth mentioning, however, that the creation of the new bulk atom has consequences for the energy involved, as we shall discuss in a later chapter.)

Examples of reaction equations for defects

In view of the many types of point defects that may be formed in inorganic compounds and that each type of defect may have varying effective charge, numerous defect reactions may in principle be formulated. In the following, a few simple cases will be treated as examples. First, we will consider defect structure situations in stoichiometric compounds (Schottky, Frenkel and intrinsic electronic disorders) and then defect structure situations in nonstoichiometric oxides will be illustrated. Finally, examples of defect reactions involving foreign elements will be considered.

It should be noted that both the stoichiometric and nonstoichiometric defect structures may prevail in the same compound, depending on the activities (partial pressures) of the components in the compound.

The treatment and examples of defect reactions will not only provide training in applying the rules of defect reactions, but also broaden the description of defect structures and of the individual point defects.

Stoichiometric compounds – instrinsic disorders

As described in the previous chapter, the defect structures in stoichiometric compounds contain equivalent concentrations of negatively and positively charged point defects. These are formed as a result of internal equilibria in the crystal and do not involve reactions with the surroundings. For this reason the defect structures in stoichiometric compounds are also termed internal disorder.

Schottky disorder

As described in the previous chapter, the Schottky disorder involves the presence of equivalent amounts of cation and anion vacancies. In an oxide MO this means that the crystal contains equal concentrations of metal and oxygen vacancies. The overall formation of such a defect pair within the crystal involves the transfer of a pair of cations and anions on regular lattice sites from the bulk to the surface. In reality the defects are formed at external and internal surfaces or dislocations and subsequently diffuse into the crystal until they are randomly distributed. In writing the equation, one is only interested in the initial and final states, and one disregards the kinetics of the defect reaction.

If one starts with a pair of cations and anions on regular lattice sites within the crystal, M_M and O_O, one must also take into account that the formation of the Schottky pair results in the formation of two new lattice sites, and the overall equation may thus be written

$$M_M + O_O = v_M^{//} + v_O^{\bullet\bullet} + M_M + O_O \tag{2.1}$$

However, in this equation M_M and O_O on both sides may be cancelled, and the net equation therefore becomes

$$0 = v_M^{//} + v_O^{\bullet\bullet} \tag{2.2}$$

where 0 (nil) designates a perfect crystal.

In Eqs. 2.1 and 2.2 it is assumed that doubly charged vacancies predominate and that neutral and singly charged vacancies can be neglected. The equations could have been written in more detailed steps starting with the formation of neutral vacancies and the subsequent, consecutive ionisation of these to singly and doubly charged vacancies.

Frenkel disorder

For the Frenkel disorder the predominant defects are either limited to the cations and anions, and the disorder involves the presence of equal numbers of vacancies and interstitial ions in a sublattice in a crystal. In the formation of a Frenkel defect pair, a cation on a normal site is transferred to an interstitial site, and no new lattice sites are created in the process. If, for the sake of illustration, the interstitial ion and the resulting vacancy are assumed to be doubly charged, the formation of a Frenkel defect pair may be written

$$M_M = v_M^{//} + M_i^{\bullet\bullet} \tag{2.3}$$

A corresponding equation may be written for the formation of an anion Frenkel defect pair. This latter defect situation is also often termed an anti-Frenkel defect structure.

Intrinsic ionisation of electrons

The excitation of an electron from the valence band to the conduction band, thereby leaving an electron hole in the valence band, is written

$$0 = e^{/} + h^{\bullet} \tag{2.4}$$

As mentioned above we can choose to apply the mass balance to electrons in which case we might rewrite the equation as follows:

$$e^{x} = e^{/} + h^{\bullet} \tag{2.5}$$

We can also choose to keep some kind of track of the sites – or energy levels – for electrons;

$$e_v^{x} + h_c^{x} = e_c^{/} + h_v^{\bullet} \tag{2.6}$$

but both the two latter reactions represent levels of detail that we normally need not consider.

As mentioned earlier, the electronic defects may be localized (valence defects). In these cases the reactions are connected with individual atomic sites. For instance the intrinsic ionization (disproprtionation) of Fe^{3+} ions into Fe^{2+} and Fe^{4+} ions would be written

$$2Fe_{Fe}^{x} = Fe_{Fe}^{/} + Fe_{Fe}^{\bullet} \tag{2.7}$$

Similarly, intrinsic ionization may also take place by charge transfer. In ilmenite, $FeTiO_3$, we may for instance observe the reaction $Fe^{2+} + Ti^{4+} = Fe^{3+} + Ti^{3+}$, which in terms of defects is written:

$$Fe_{Fe}^{x} + Ti_{Ti}^{x} = Fe_{Fe}^{\bullet} + Ti_{Ti}^{/} \tag{2.8}$$

Nonstoichiometry

The ratio of cation to anion lattice sites is the same whether a compound is stoichiometric or nonstoichiometric. But as nonstoichiometry means that there is an excess or deficit of either cations or anions, nonstoichiometry also means that there is an excess of a certain type or types of defects relative to that in the stoichiometric condition. If the predominating type of defects are charged, complementary electronic defects are created in order to conserve electrical neutrality. The extent of nonstoichiometry and the defect concentrations in inorganic compounds are functions of temperature and activities (partial pressures) of their components.

First we consider, as an example the formation of oxygen deficiency in an oxide MO_2. The overall reaction may be written

$$MO_2(s) = MO_{2-x}(s) + \tfrac{x}{2} O_2(g) \tag{2.9}$$

From this equation it is qualitatively seen using le Chatelier's principle that the oxygen deficit increases with decreasing oxygen pressure. Conversely, for oxides with excess oxygen the nonstoichiometry increases with increasing oxygen pressure.

In the following various defect reactions which are encountered in oxides with oxygen or metal excess or deficit will be considered. As the activity of the metal component is usually negligibly small compared to that of the oxygen activity under most experimental conditions, the nonstoichiometry in metal oxides is correspondingly a result of the interaction and exchange of oxygen between the metal oxide and the surrounding gas atmosphere. In the following examples of formation of nonstoichiometric defects, only cases where the metal oxides interact with gaseous oxygen are illustrated. However, it should be borne in mind that the corresponding nonstoichiometry and defects may be formed by interaction with metal if it is experimentally feasible to control the activity of the metal component in the surroundings of the crystal.

Oxygen-deficient oxides

An oxygen vacancy is formed by the transfer of an oxygen atom on a normal site to the gaseous state. No change in the number of lattice sites takes place. This defect reaction may be written

$$O_O^x = v_O^x + \tfrac{1}{2} O_2(g) \tag{2.10}$$

In this equation it is assumed that the oxygen vacancy is neutral, i.e. the two electrons of the O^{2-} ion are associated with the vacancy or its immediate neighbourhood and the vacancy has zero effective charge. As described in Chapter 1, the two electrons trapped at or near the vacancy may, depending on the temperature and vacancy concentration, be excited and transferred away from the vacancy. Correspondingly, the oxygen vacancy acts as a donor and becomes singly and doubly charged:

$$v_O^x = v_O^{\bullet} + e^{/} \tag{2.11}$$

$$v_O^{\bullet} = v_O^{\bullet\bullet} + e^{/} \tag{2.12}$$

The formation of the doubly ionized oxygen vacancy can be written as the total reaction:

$$O_O^x = v_O^{\bullet\bullet} + 2e^/ + \tfrac{1}{2}O_2(g) \tag{2.13}$$

In these equations the free electrons are considered delocalised in the conduction band. If they are localised at a metal ion on a normal lattice site the last equation could be reformulated as follows:

$$O_O^x + 2M_M^x = v_O^{\bullet\bullet} + 2M_M^/ + \tfrac{1}{2}O_2(g) \tag{2.14}$$

If this takes place in the oxide MO_2 the valence of the M atoms then partly reduce from +4 to +3 when the oxide becomes non-stoichiometric (MO_{1-x}).

Oxides with excess metal

The oxygen deficiency in an oxide may alternatively be equivalent to the presence of excess metal relative to stoichiometric composition, and in this case the predominant defects constitute interstitial atoms. Correspondingly the composition of the oxide MO_2 should then be written $M_{1+x}O_2$. The formation of an interstitial M atom in this oxide involves the transfer of a regular M_M atom to an interstitial site. As a metal lattice site is annihilated in this process, two oxygen lattice sites must simultaneously be annihilated in the oxygen sublattice and this is achieved by transferring two oxygen atoms to the gas phase. The defect equation then becomes

$$M_M^x + 2O_O^x = M_i^x + O_2(g) \tag{2.15}$$

The neutral interstitial M_i atoms may be successively ionised to singly, doubly, triply or quadruply charged interstitial ions, e.g.

$$M_i^x = M_i^\bullet + e^/ \text{, etc.} \tag{2.16}$$

From this we can state that the metal interstitial, like the oxygen vacancy, is an electron donor. It may be noted that in both cases it is the effectively neutral or incompletely ionized defects that is a donor, not the fully ionized defects.

We have until now written the formation of metal defects by exchange with oxygen in the surrounding atmosphere. We may in principle choose to write the same in terms of exchange with the metal component in the surroundings. For instance, metal insterstitials in MO can be formed in equilibrium with the metal vapour:

$$M(g) = M_i^{\bullet\bullet} + 2e^{/} \qquad (2.17)$$

Metal-deficient oxides

In metal-deficient compounds metal vacancies are the predominating point defects. In an oxide MO a metal vacancy is formed by reacting oxygen gas with the oxide. This creates new oxygen lattice sites and therefore an equivalent number of new metal lattice sites - which are vacant - is also formed:

$$\tfrac{1}{2}O_2(g) = v_M^x + O_O^x \qquad (2.18)$$

In this reaction it is assumed that the vacancy that is formed is effectively neutral, that is, two electrons are taken from the surroundings of the vacancy in order to form the O^{2-} ion. The lacking electrons may be taken up from the valence band, forming electron holes here;

$$v_M^x = v_M^{//} + 2h^{\bullet} \qquad (2.19)$$

The formation of electron holes may alternatively be expressed in terms of valence defects:

$$v_M^x + 2M_M^x = v_M^{//} + 2M_M^{\bullet} \qquad (2.20)$$

This correspondingly means that the valence for some M atoms changes from +2 for to +3.

The metal vacancy, whether it is effectively neutral (electrons missing from its neighbourhood) or ionized (electrons supplied from the rest of the crystal) it illustrates that the effective size of a point defect may extend beyond the site itself; the charge may be distributed over at least the closest neighbours of the defect. Accordingly, defect chemistry works fine under the ionic model and with the assumption of integer charges, even if the compound is far from ideally ionic – the point defect is just a little bigger than one site. It works because in defect chemical as in other chemical reactions, the electrons have to choose to go or stay – no half electrons are involved. It then does not matter how close that electron was of the point defect – as long as it was associated with it. For most defect chemical considerations, it then also suffices to consider the defect as a true point defect, and that the electrons are associated with the point defect itself.

Oxides with excess oxygen

In metal oxides with excess oxygen the predominating point defects are interstitial oxygen atoms or ions. The formation of a neutral interstitial oxygen atom through reaction of oxygen with the oxide is written

$$\tfrac{1}{2}O_2(g) = O_i^x \tag{2.21}$$

No new lattice sites are formed in this reaction. The neutral interstitial oxygen atoms may in principle be ionised to yield electron holes and interstitial oxygen ions with negative effective charges, so that the total reaction becomes, for instance,

$$\tfrac{1}{2}O_2(g) = O_i^{//} + 2h^{\bullet} \tag{2.22}$$

Dissolution of foreign elements

The presence of impurities or dopants may significantly affect or even control the concentrations of the native defects in a compound. The effects are to a large extent dependent upon the relative valences of the ions in the parent compound and those of the impurities or dopants and what sites the impurities and dopants occupy. When the principal valence is the same as that of the host site, we say that the foreign element is homovalent. When it is unequal to that of the host site, we say that the foreign element is heterovalent – or aliovalent.

Foreign aliovalent elements may dissolve in the host structure with zero effective charge, However, their difference from the host atom makes them easily donate or accept one or more electrons – they are donors or acceptors. We will briefly analyse the reason for this and describe the processes in terms of defect reactions.

When boron, B, or phosphorous, P, dissolve as dopants in elemental silicon, Si, they have, respectively one less or one too many valence electrons, compared to Si. However, they are in this state effectively neutral since the electron number in each case is compensated by the nuclear charge. Nevertheless, the missing or extra electrons form local states in the band structure that easily accept or donate electrons:

$$B_{Si}^x = B_{Si}^/ + h^{\bullet} \tag{2.23}$$

$$P_{Si}^x = P_{Si}^{\bullet} + e^/ \tag{2.24}$$

The neutral defects, the acceptor and donor, were here formed by dissolving the foreign elements in the same oxidation state as the host, namely 0. It is the desire of the crystal to fulfil the octet rule around each atom – the covalent nature of the bond – that causes the ionization in this case.

We may do something similar with ionic compounds such as oxides. For instance, hypothetical LiO, containing divalent lithium, Li^{2+}, may be dissolved in NiO to form effectively neutral Li_{Ni}^{2+} species; Li_{Ni}^{x}. Since Li is unstable in oxidation state +2, it easily accepts an electron from the valence band of the crystal:

$$Li_{Ni}^{x} = Li_{Ni}^{/} + h^{\bullet} \tag{2.25}$$

Thus, in this case, it is the formal oxidation state according to an ionic model that is the argument for why the acceptor is ionised. Similar arguments hold for donors in ionic compounds.

When dealing with aliovalent foreign elements it is common to assume directly the ionisation of the donor or acceptor into an effectively charged state. We often refer to the cases as donor- or acceptor-doped systems, but keep in mind that it is the unionised species that constitutes the donor or the acceptor.

Effects of dissolution of aliovalent oxides in the metal-deficient oxide $M_{1-y}O$.

Let us assume that the parent oxide is metal-deficient $M_{1-y}O$ and that the majority point defects in $M_{1-y}O$ are charged metal vacancies, e.g. $v_M^{//}$, compensated by electron holes, $h^{\bullet}$. The oxide is thus a p-type electronic conductor. Assume now that a higher valent oxide Mh_2O_3 is added. (Mh is used as an arbitrary chemical symbol for a higher valent metal.) Some Mh_2O_3 dissolves in the $M_{1-y}O$. The Mh-ions have in this case a valence of +3, and if it is assumed that the Mh-ions occupy normal M-sites in $M_{1-y}O$, the dissolved Mh-ions will have one positive effective charge, $Mh_M^{\bullet}$. These additional positive effective charges must for electroneutrality reasons be balanced by creation of an equivalent concentration of negative effective charges or annihilation of an equivalent concentration of positive effective charges. Let us consider these two alternatives.

The dissolution of Mh_2O_3 may in the first case be compensated by the formation of additional M-vacancies with two negative effective charges. In this case the equation for the dissolution of Mh_2O_3 is written

$$Mh_2O_3 = 2Mh_M^{\bullet} + v_M^{//} + 3O_O^{x} \tag{2.26}$$

and thus the dissolution of Mh_2O_3 in $M_{1-y}O$ increases the concentration of metal vacancies in the parent oxide.

But as will be discussed in detail in the next chapter an increased concentration of charged metal vacancies simultaneously results in a decrease in the concentration of electron holes. The dissolution of Mh_2O_3 may thus alternatively be written with a defect reaction involving annihilation of electronic holes:

$$Mh_2O_3 + 2h^{\bullet} = 2Mh_M^{\bullet} + 2O_O^x + \tfrac{1}{2}O_2(g) \qquad (2.27)$$

When the dissolution of Mh_2O_3 is written in this way, the equation emphasises that the concentration of electron holes is reduced. However, as the concentration of metal vacancies and electron holes are interrelated, it should be emphasised that both processes take place. Actually, the dissolution of the higher valent cation decreases the concentration of all effectively positive defects and increases the concentration of all negatively charged defects. Qualitatively, this may simply be seen as a consequence of the electroneutrality equation or, if one wishes, of Le Chatelier's principle. Details of these aspects will be further discussed using defect diagrams in Chapter 4.

If one has the choice of writing a doping reaction by creation or annihilation of defects, it is usually more meaningful to choose the creation of defects, since this can describe the doping reaction to levels beyond the defect concentrations that existed in the undoped oxide, and since this describes the defects that will dominate when the doping level gets high.

So far we have added a higher valent dopant oxide to $M_{1-y}O$. Let us alternatively consider what takes place when a lower valent dopant oxide, Ml_2O, is added to $M_{1-y}O$. If the Ml^+ ions dissolve substitutionally in $M_{1-y}O$, the dissolved Ml^+ ions have one negative effective charge, $Ml_M^/$. This will be compensated by the formation of positive effective charge or the annihilation of negative effective charge. In our example oxide this will mainly affect electron holes and metal vacancies, and the two defect reactions may thus be written

$$Ml_2O + \tfrac{1}{2}O_2(g) = 2Ml_M^/ + 2h^{\bullet} + 2O_O^x \qquad (2.28)$$

$$Ml_2O + v_M^{//} = 2Ml_M^/ + O_O^x \qquad (2.29)$$

Thus, the addition of Ml_2O to $M_{1-y}O$ has the opposite effect of addition of Mh_2O_3; it increases the concentration of electron holes and decreases the concentration of metal vacancies. One may note that the first of the two reactions is an oxidation and thus involves uptake of oxygen, while the latter is not. It is typical of such dissolutions of foreign aliovalent elements in oxides that they can be charge compensated by means of electronic defects – in which case we are dealing with reduction or oxidation and uptake or release of oxygen – or by means of point defects – in which case we have no change in oxidation states and no exchange of oxygen gas.

Effects of dissolution of aliovalent oxides on the oxygen-deficient oxide MO_{2-x}

The examples in the previous section show the effects of additions of higher and lower valent oxide to the p-conducting, metal-deficient $M_{1-y}O$. Let us also briefly consider the effects of doping an oxygen-deficient oxide MO_{2-x} with higher and lower valent oxides, respectively. The predominating defects in MO_{2-x} are oxygen vacancies compensated by defect electrons. The oxide it thus an n-type electronic conductor.

When the dopant oxide is higher valent, e.g. Mh_2O_5, and the dopant cation, Mh^{5+}, dissolves substitutionally, the dopant ions get one effective positive charge, $Mh_M^{\bullet}$. This charge must be compensated either by formation of negative effective charges in the form of electrons or by annihilation of positive effective charges, oxygen vacancies. These defect reactions may be written

$$Mh_2O_5 = 2Mh_M^{\bullet} + 2e^{/} + 4O_O^x + \tfrac{1}{2}O_2(g) \qquad (2.30)$$

$$Mh_2O_5 + v_O^{\bullet\bullet} = 2Mh_M^{\bullet} + 5O_O^x \qquad (2.31)$$

Thus when MO_{2-x} is doped with a higher valent oxides Mh_2O_5 the concentration of electrons is increased and the concentration of oxygen vacancies is decreased.

When the same oxide, MO_{2-x}, is doped with a lower valent oxide, Ml_2O_3, the negative effective charge of the dissolved atoms $Ml_M^{/}$ is compensated by annihilation of electrons or formation of oxygen vacancies:

$$Ml_2O_3 + 2e^{/} + \tfrac{1}{2}O_2(g) = 2Ml_M^{/} + 4O_O^x \qquad (2.32)$$

$$Ml_2O_3 = 2Ml_M^{/} + v_O^{\bullet\bullet} + 3O_O^x \qquad (2.33)$$

The latter is exemplified by yttria-stabilised zirconia where the acceptor doping is compensated by mobile oxygen vacancies that lay ground for this material as a solid electrolyte:

$$Y_2O_3 = 2Y_{Zr}^{/} + v_O^{\bullet\bullet} + 3O_O^x \qquad (2.34)$$

Numerous other examples may be formulated, depending on the defect structure of the parent oxide, the valence of the foreign ion and the site it occupies. It may be noted that a foreign cation dissolving interstitially always will have a positive effective charge and thus affect the defect structure in a similar manner as a higher valent cation dissolved substitutionally. We shall return to this and other examples in subsequent chapters and in connection with the review and discussion of a few individual oxides systems.

Foreign anions in oxides

In oxides, homovalent foreign anions comprise S^{2-}, while aliovalent foreign anions comprise F^- and N^{3-}. They can enter as impurities during synthesis, or dissolve from gaseous species under reducing atmospheres, e.g.,

$$O_O^x + H_2S(g) = S_O^x + H_2O(g) \tag{2.35}$$

$$3O_O^x + 2NH_3(g) = 2N_O^/ + v_O^{\bullet\bullet} + 3H_2O(g) \tag{2.36}$$

Dissolution of hydrogen in metal oxides

When metal oxides are exposed to gas atmospheres containing water vapour or other hydrogen containing gases, hydrogen will dissolve in the oxides. Under oxidizing or mildly reducing conditions, the hydrogen atoms ionise to protons and associate with oxygen atoms on normal lattice sites and thereby form hydroxide ions on normal oxygen sites, $OH_O^\bullet$. We may thus for instance write the dissolution reaction

$$H_2 + 2O_O^x = 2OH_O^\bullet + 2e^/ \tag{2.37}$$

or

$$H_2O(g) + 2O_O^x = 2OH_O^\bullet + 2e^/ + \tfrac{1}{2}O_2(g) \tag{2.38}$$

in which case the protons dissolved are charge compensated by the formation of defect electrons. Of course, the electrons may interact with other defects in the oxide so that the protons in effect are compensated by formation of other negative defects or by the annihilation of positive defects. From the dissolution reaction and through the interaction with native defects in the oxide it is clear that the dissolution of hydrogen in metal oxides is dependent both on the partial pressure of the hydrogen source (e.g. water vapour or hydrogen) and of oxygen. These aspects will be described in more detail in a later chapter.

The dissolution of hydrogen from its oxide, H_2O, is in principle similar to dissolution of other foreign cations. However, the possibility of a controlled water vapour pressure and the fast diffusion of protons makes it much easier to attain and vary (and more difficult to completely avoid) an equilibrium content of protons in the oxide.

In terms of defect chemistry the dissolved proton, located on a normal oxygen ion as hydroxide, may also be considered to constitute an interstitial hydrogen ion, and as such it is also in the literature alternatively written $H_i^{\bullet}$. One just has to bear in mind that the protons do not occupy regular interstitial positions (voids).

Under reducing conditions, where hydrogen is stable in oxidation state 0 (as H_2 in the gas phase) it might be possible to foresee neutral hydrogen atoms dissolved in oxides, probably interstitially, as H_i^x. Under even much more reducing conditions could also hydride ions be expected to become stable, e.g. as dissolved substitutionally for oxygen ions, as the defect $H_O^{\bullet}$.

Ternary and higher compounds

We have so far concentrated on binary compounds, mostly oxides, and only touched upon elemental solids. Ternary and higher compounds fall, however, under exactly the same rules of writing defect reactions.

A typical ternary compound is a ternary oxide such as perovskite, $CaTiO_3$. As an example of defect reactions for this case, we consider first the formation of Schottky defects. When we form new lattice sites in this reaction, we need to form vacancies on both Ca and Ti sites to maintain the ratio between them, in addition to the appropriate number of oxygen vacancies:

$$0 = v_{Ca}^{//} + v_{Ti}^{////} + 3v_O^{\bullet\bullet} \tag{2.39}$$

If we further consider the uptake of oxygen by formation of cation vacancies and electron holes, we again have to balance the cation sites:

$$\tfrac{3}{2}O_2(g) = v_{Ca}^{//} + v_{Ti}^{////} + 3O_O^x + 6h^{\bullet} \tag{2.40}$$

Similar principles should be applied also in cases where one and the same element is distributed on different crystallographic sites. For instance, Y_2O_3 has a structure where all oxygen ions are not strictly equal. Similarly, distorted perovskites may have unequal oxygen sites. In the pyrochlore structure, $A_2B_2O_7$, there are 6 oxygen sites of one type and 1 of slightly different coordination and energy (and one which is structurally empty and thus to be regarded as an interstitial site). In principle the fomation or annihilation of crystal units has to maintain the ratio between those different sites in all such cases. However, this is so far hardly ever practiced in defect chemistry.

Contrary to binary oxides, ternary and higher oxides can have non-stoichiometry not only in terms of the oxygen-to-metal ratio, but also internally between the various cations. This is in practice often a result of synthesis. For instance, it may be difficult to weigh in exactly equal numbers of moles of Ca and Ti precursors when synthesizing $CaTiO_3$, so that the synthesized material has a permanent number of vacancies on one of the cation sites. Such non-stoichiometry may also be a result of equilibria. For instance, if A-site deficiency is energetically favourable over B-site deficiency in the compound ABO_3, we may at very high temperatures (e.g., during sintering) see a preferential evaporation of the A component:

$$A_A^x + O_O^x = v_A^{//} + v_O^{\bullet\bullet} + AO(g) \qquad (2.41)$$

During oxidation we might similarly see a preferential incorporation of A-site vacancies, resulting in a precipitation of an A-rich phase:

$$A_A^x + \tfrac{1}{2}O_2(g) = v_A^{//} + 2h^{\bullet} + AO(s) \qquad (2.42)$$

It may be noted that these reaction equations do not violate the site ratio conservation requirement of the ternary oxide.

When we earlier doped elementary or binary compounds the reaction was fairly straightforward. When we dope a ternary or higher compound, however, the reaction may be less obvious – we have some choices. It is quite common, however, to do the synthesis and write the equation in such a way that one takes out a corresponding amount of the host element that is substituted. If we, for instance, want to dope $LaScO_3$ with Ca substituting for La, we go for a composition $La_{1-x}Ca_xScO_3$. In order to see how we write the doping reaction in this case we first just look at the trivial normal synthesis:

$$\tfrac{1}{2}La_2O_3 + \tfrac{1}{2}Sc_2O_2 = La_{La}^x + Sc_{Sc}^x + 3O_O^x \qquad (2.43)$$

Accordingly, we then write the defect reaction for the doping in the way that we let there be Sc_2O_3 reserved for the CaO:

$$CaO + \tfrac{1}{2}Sc_2O_3 = Ca_{La}^{/} + Sc_{Sc}^x + \tfrac{5}{2}O_O^x + \tfrac{1}{2}v_O^{\bullet\bullet} \qquad (2.44)$$

Summary

Defect reactions can be written in a similar manner as ordinary chemical reactions. They must confer with the requirements for conservation of mass, charge, and ratios of lattice sites, but are allowed to increase or decrease the total number of lattice units.

We have treated defect reactions mostly in cases for binary ionic compounds, notably oxides. Reactions for elemental crystalline solids follow the same basic rules and principles, but without the site ratio conservation requirement. Also higher compounds, e.g. ternary oxides, follow the same rules, only often with slightly more complex site conservation considerations and more defects.

Normally, one should seek to use reactions that create rather than annihilate defects, and one should avoid using reactions that form more defects than necessary (as such reactions are then a sum of simpler reactions).

We have formulated defect reactions which describe intrinsic ionic and electronic disorder, nonstoichiometry, variable ionisation of point defects, and substitutional dissolution of aliovalent cations and ions and anions. Aliovalent elements may be compensated by electronic defects or by point defects, of which the former involve red-ox-reactions.

We have treated hydrogen defects specially, as they arise from a special source and in its most stable form, the proton, take on a special size and type of defect.

Exercises

The excercises in this chapter mainly provide training in 1) formulating a reaction in terms of the appropriate reactants and products, and 2) balancing and checking them with respect to a) mass conservation, b) charge conservation, and c) site ratio conservation.

1. Write a reaction for the formation of Schottky defects in NaCl.
2. Write a reaction for the formation of Schottky defects in MO_2.
3. Write a reaction for the formation of Schottky defects in Cu_2O.
4. Write a reaction for the formation of anion Frenkel defects in MO.
5. Write a reaction for the formation of anion Frenkel defects in CaF_2.
6. Write a reaction for a charge transfer between the cation and anion in CeO_2, i.e. for reduction of the cerium ion and oxidation of the oxygen ion. Write the same process as an intrinsic ionisation assuming delocalised electronic defects.

7. Write a reaction for the formation of fully ionised oxygen vacancies and electrons when oxygen is lost in the reaction $M_2O_3 = M_2O_{3-x} + x/2\ O_2(g)$.

8. Write reactions for the direct formation of fully ionised metal interstitials in MO, M_2O_3, and MO_2.

9. Write reactions for the direct formation of fully ionised metal vacancies in MO, M_2O_3, and MO_2.

10. Write reactions for formation of sodium vacancies in NaCl by exchange with a) chlorine gas and b) sodium gas.

11. We have in the text claimed that charges can be associated directly with ideal point defects, even if the compound is not ideally ionic, and that this normally does not affect the results of the calculations. Under what conditions will it start to matter that the charge of the defect actually extends outside the defect itself?

12. Write reactions for dissolution of CaO into the oxide ZrO_{2-x}

13. Write reactions for dissolution of MfO into $M_{2-x}O_3$ when we assume that the foreign metal Mf is small and therefore dissolves interstitially.

14. Write reactions for dissolution of CaO substitutionally into the anion-Frenkel dominated Y_2O_3.

15. Write reactions for dissolution of ZrO_2 substitutionally into the anion-Frenkel dominated Y_2O_3.

16. Write reactions for dissolution of CaF_2 into CaO, assuming that F^- dissolves substitutionally and that CaO is dominated by Schottky defects in the pure, undoped state.

17. Write a reaction for dissolution of protons in an oxide with water vapour as source and with oxygen interstitials as compensating defects.

18. Write reactions for dissolution of protons in an oxide $M_{1-x}O$ with water vapour as source.

19. Consider an oxide M_2O_3 which is acceptor-doped by dissolving the oxide MlO. The dopants are compensated by oxygen vacancies. Write a defect reaction for the dissolution of MlO. Write a reaction for the dissolution of protons from water vapour into this acceptor doped oxide, by annihilation of the oxygen vacancies. Finally, write a reaction for the dissolution of MlO in the presence of water vapour and without forming oxygen vacancies. Can the latter reaction be constructed as a sum of the two previous reactions?

20. Hydride ions have been suggested to dissolve in oxides accompanied by protons by disproportionation of hydrogen gas. Write defect reactions for this in the case that the hydride dissolves a) interstitially, and b) substitutionally. (Assume in both cases that the oxide is perfectly stoichiometric and has no defects as a starting point).

21. Write defect reactions for formation of A-site Frenkel, B-site Frenkel, and anion Frenkel disorders in a perovskite ABO_3 (assuming both cations are trivalent).

22. Write a defect reaction for the formation of Schottky disorder in spinel $MgAl_2O_4$.
23. Suggest a defect reaction for formation of Schottky-type disorder in $LaMnO_3$ when it is assumed that cation vacancies are formed in the La sublattice only.
24. Write a defect reaction for dissolution of Sc_2O_3 in $CaTiO_3$ during synthesis, assuming that Sc substitutes Ti forming acceptors compensated by oxygen vacancies.

3. Defect equilibria

Introduction

After formulating the correct defect reactions in a compound, we turn our interest towards the equilibrium relations for those reactions. These correlate the equilibrium concentrations of the different defects with temperature, activities (partial pressures) of the components in the compound, and other parameters which affect the defect structure. The defect equilibria will here generally be described by the law of mass action.

The change in the Gibbs (free) energy of a chemical reaction in a closed system is given by[3]

$$\Delta G = \Delta H - T\Delta S \qquad (3.1)$$

where ΔH is the change in enthalpy and ΔS the change in entropy of the reaction and T is the absolute temperature.

The reaction will proceed spontaneously as long as $\Delta G < 0$, until G has reached its minimum. Here, $\Delta G = 0$, and this relation represents the condition for chemical equilibrium.

The creation of single, unassociated point defects in an elemental, crystalline solid increases the internal energy of the system and the enthalpy of the defect formation is positive. But the configurational entropy of the system also increases, and the equilibrium concentration of the defects will be reached when the Gibbs energy of the system is at minimum. Thermodynamically, point defects will thus always be present in a crystal above 0 K.

All types of defects will in principle be formed. However, the free energies of formation of the different types and systems of defects can have widely different values, and correspondingly it is often found that certain defects predominate in a particular solid. The relative concentrations of the different types of defects will be a function a temperature and other variables. Thus, defect equilibria with large positive enthalpy of formation, for instance, which are not favoured at low temperatures, may become important at high temperatures.

From a thermodynamic point of view a solid containing point defects constitutes a solid solution where the point defects are dissolved in the solid. In analogy with liquid solutions the solid may be considered to be the solvent and the point defects the solute. Similarly the defect equilibria may be treated in terms of the thermodynamics of chemical reactions and solutions.

---

[3] A closed system is one where the mass (chemical entities) of the system is confined, but where energy (heat) can be exchanged with the surroundings. The expression for the Gibbs energy arises as a result of the definition of entropy and of the first law of thermodynamics which preserves the total energy of the system and the surroundings.

Point defects and equilibrium in an elemental solid - a simple statistical thermodynamic approach

Before considering the thermodynamics of defect formation and of defect reactions in metal oxides and the corresponding use of the law of mass action, the simpler case of formation of one type of defect in an elemental solid will be treated by means of statistical thermodynamics.

Consider a perfect crystal of an elemental solid E with N atoms as starting material. Let n_v vacancies be formed according to the defect-chemical reaction

$$E_E = v_E + E_E \qquad (3.2)$$

where E_E is an E-atom on an E-site and v_E is a vacant E-site. The vacancies are formed by moving an E-atom from an E-site in the bulk to the surface of the crystal. The total number of sites then becomes $N+n_v$. The change in Gibbs energy can be expressed in terms of the enthalpy of formation of each vacancy, ΔH, and the entropy associated with the formation of the defects, ΔS.

The enthalpy change on forming n_v vacancies is thus $n_v\Delta H$. It expresses the energy cost associated with creating a new, empty site. Since we are dealing with small concentrations of defects, the enthalpy should be one that is valid for ideal dilute solutions of vacancies in the lattice.

The entropy change can be divided into two parts: i) a vibrational entropy change, ΔS_{vib}, reflecting the entropy created by vibrations associated with each new vacancy, and ii) a configurational entropy, ΔS_{conf}, arising from the distribution of n vacancies among $N+n_V$ sites.

The vibrational entropy change is associated with the vibrations in the neighbourhood of each vacancy, and the total vibrational entropy change is proportional to the number of vacancies, $n_v\Delta S_{vib}$. This entropy change is, as was the enthalpy change, related to a dilute solution of vacancies.

The enthalpy and vibrational entropy changes are, as we have discussed, to a first approximation proportional to the number of vacancies formed, while the configurational entropy is a more complex function. The change in Gibbs free energy associated with the formation of n_v vacancies may accordingly first be written

$$\Delta G = n_v(\Delta H - T\Delta S_{vib}) - T\Delta S_{conf} \qquad (3.3)$$

According to statistical thermodynamics the configurational entropy (entropy of mixing) can be expressed in terms of the thermodynamic probability, W, by the relation

$$S_{conf} = k\ln W \qquad (3.4)$$

where k is Boltzmann's constant. The thermodynamic probability W represents in this case the number of different distinguishable ways whereby n_V vacancies may be distributed on $N+n_V$ lattice sites and is given as

$$W = \frac{(N+n_v)!}{N!n_v!} \tag{3.5}$$

For large numbers of N and n_V, which is typical for real crystals, Stirling's approximation ($\ln x! = x\ln x - x$ for x>>1) may be applied, whereby Eq. 3.4 takes the form

$$S_{conf} = k(N\ln\frac{N+n_v}{N} + n_v\ln\frac{N+n_v}{n_v}) \tag{3.6}$$

This entropy scales with the size of the system (the number N) but is not proportional with the number of defects n_v.

Intuitively, and by insertion in the equation above, we have $S_{conf} = 0$ at $n_v = 0$. Thus, $\Delta S_{conf} = S_{conf}$ and by insertion into the expression for the change in free energy we get

$$\Delta G = n_v(\Delta H - T\Delta S_{vib}) - kT(N\ln\frac{N+n_v}{N} + n_v\ln\frac{N+n_v}{n_v}) \tag{3.7}$$

The changes in enthalpy, entropy and Gibbs energy are illustrated schematically in Fig.3.1.

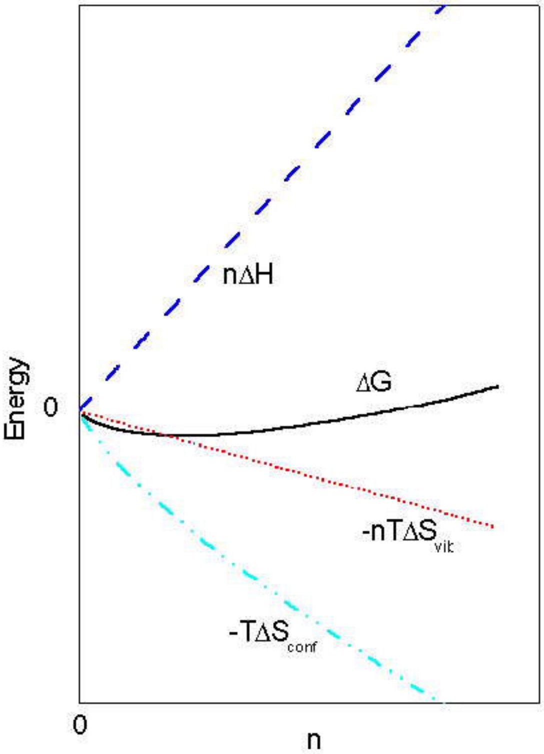

Figure 3-1. Example of the variation in enthalpy, vibrational and configurational entropy, and Gibbs energy with the concentration of vacancies in an elemental solid.

At equilibrium *ΔG* will be a minimum with respect to n_v , i.e. $\partial \Delta G / \partial n_V = 0$. By taking the derivative, and letting it be zero, one obtains

$$\frac{\partial \Delta \mathrm{G}}{\partial \mathrm{n_V}} = \Delta \mathrm{H} - T\Delta \mathrm{S_{vib}} + kT\ln\frac{n_v}{N+n_v} = 0 \tag{3.8}$$

The term $\frac{\mathrm{n_v}}{\mathrm{N+n_v}}$ represents the fraction of the total number of sites which are vacant and is through rearrangement of Eq.3.8 given by

$$\frac{\mathrm{n_v}}{\mathrm{N+n_v}} = \exp(\frac{\Delta \mathrm{S_{vib}}}{\mathrm{k}})\exp(-\frac{\Delta \mathrm{H}}{\mathrm{kT}}) \tag{3.9}$$

Since the formation of the vacancies is written $E_E = v_E + E_E$, the equilibrium constant may, following the law of mass action, be written

$$\mathrm{K_v} = a_{v_E} = \mathrm{X}_{v_E} = \frac{\mathrm{n_v}}{\mathrm{N+n_v}} = \exp(\frac{\Delta \mathrm{S_{vib}}}{\mathrm{k}})\exp(-\frac{\Delta \mathrm{H}}{\mathrm{kT}}) \tag{3.10}$$

a_{v_E} is the activity and $X_{v_E} = \frac{n_v}{N + n_v}$ represent the site fraction of vacancies in the crystal of the element E.

ΔS_{vib} in the expressions above represents the vibrational entropy change and ΔH the enthalpy change per vacancy. If one wants to express the same properties per mole of vacancies, k must be substituted by the gas constant R = N_Ak where N_A is Avogadro's number. Relations corresponding to Eq. 3.10 may be derived for other types of point defects.

At any temperature T>0 all solids will following Eq. 3.10 contain point defects. In the example above we have shown that the equilibrium constant, containing defect activities as site fractions, is given by an enthalpy change and an entropy term which constitutes the vibrational entropy change and not the (configurational) entropy of mixing. This applies to chemical mass action law expressions in general. Furthermore, it illustrates that the vibrational entropy enters without other factors only if the equilibrium constant is expressed properly with concentrations given as site fractions in the case of defects.

It is common to combine the enthalpy and entropy changes into a Gibbs energy change, in this case the standard Gibbs energy change:

$$K_v = \exp(\frac{\Delta S_{vib}}{R}) \exp(-\frac{\Delta H}{RT}) = \exp(-\frac{\Delta G^0}{RT}) \qquad (3.11)$$

We will see later that this standard Gibbs energy change is the one we have when the reaction proceeds with all reactants and products in their standard reference state (activity of unity). It is not the same as the general Gibbs energy change calculated for the progression of the defect formation reaction earlier.

Thermodynamics of chemical reactions

As a basis for the subsequent considerations of thermodynamics of defect equilibria the following section provides a brief review of some general aspects of thermodynamics of chemical reactions.

In treating chemical reactions and systems consisting of two or more constituents, it is convenient to introduce the concept of partial molar thermodynamic quantities. As regards equilibria, the partial molar Gibbs energy is particularly important. The partial molar Gibbs energy is commonly termed the chemical potential and written μ_i. The physical significance of a partial molar property is the increase in that property resulting from the addition of 1 mole of the constituent to such a large quantity of the system that its composition remains essentially unchanged. If a system consists of $n_1 + n_2 + ... + n_i$ moles of constituents 1, 2, …and i, the partial molar free energy for the "i"th constituent is given by

$$\mu_i = (\frac{\partial G}{\partial n_i})_{T,p,n_1,n_2,...} \qquad (3.12)$$

The Gibbs energy of a system is in terms of chemical potentials of the constituents given by

$$G = n_1\mu_1 + n_2\mu_2 + ...n_i\mu_\mathrm{i} \tag{3.13}$$

When a chemical reaction takes place, the numbers of various constituents change, and dG changes correspondingly. For an open system (i.e. where we may add or remove heat and volume) at constant temperature and pressure, dG is given by

$$dG_{T,p} = \mu_1 dn_1 + \mu_2 dn_2 + ...\mu_\mathrm{i} dn_i \tag{3.14}$$

At equilibrium we have

$$dG_{T,p} = \sum_i \mu_\mathrm{i} dn_i = 0 \tag{3.15}$$

It may also be shown that under equilibrium conditions

$$n_1 d\mu_1 + n_2 d\mu_2 + ...n_i d\mu_\mathrm{i} = 0 \tag{3.16}$$

This equation is one form of the Gibbs-Duhem equation.

In a chemical reaction , e.g.

$$aA + bB = cC + dD \tag{3.17}$$

the change in Gibbs energy is given by the difference in the total free energy in the final and initial state

$$\Delta G = c\mu_C + d\mu_D - (a\mu_A + b\mu_B) \tag{3.18}$$

The chemical potential of the constituent "i" in the mixture can be written

$$\mu_i = \mu_i^0 + RT\ln a_i \tag{3.19}$$

where a_i is the activity of constituent "i" in the mixture and μ_i^0 is the chemical potential of constituent "i" at a chosen standard state of unit activity. The activity of pure solids or liquids is usually taken as unity at atmospheric pressure (1 bar or 10^5 Pa), while for gases the activity is unity when the partial pressure of the gas constituent is 1 bar.

In solutions the activity is not unity, but may be equated with the mole fraction, X_i, in ideal solutions:

$$a_i = X_i \tag{3.20}$$

In non-ideal solutions the activity is related to the mole fraction through the activity coefficient, γ_i,

$$a_i = \gamma_i X_i \tag{3.21}$$

γ_i is thus equal to unity in ideal solutions. It may be noted that in a system consisting of several phases, the condition of equilibrium implies that the chemical potential of constituent "i" is the same in all phases.

If one introduces the expression for the chemical potential in Eq. 3.19 into Eq. 3.18 it follows that ΔG is given by

$$\Delta G = \Delta G^0 + RT \ln \frac{a_C^c a_D^d}{a_A^a a_B^b} \tag{3.22}$$

where $\Delta G^0 = c\mu_C^0 + d\mu_D^0 - (a\mu_A^0 + b\mu_B^0)$ represents the change in Gibbs energy in the standard state, i.e. at unit activities. ΔG° is a constant under specified standard states. At equilibrium $\Delta G = 0$ such that

$$\Delta G^0 = -RT \ln K \tag{3.23}$$

where K is a constant given by

$$K = \left(\frac{a_C^c a_D^d}{a_A^a a_B^b} \right)_{equilibrium} \tag{3.24}$$

K is termed the equilibrium constant and it relates the activities (or concentrations under ideal conditions) of products and reactants when equilibrium has been attained at a given temperature. ΔG° may as in Eq. 3.1 be expressed in terms of the standard enthalpy change, ΔH°, and entropy change, ΔS°:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT \ln K \tag{3.25}$$

This may be rewritten in the form

$$K = K^0 \exp(-\frac{\Delta H^0}{RT}) = \exp(\frac{\Delta S^0}{R}) \exp(-\frac{\Delta H^0}{RT}) \tag{3.26}$$

Thus, we have arrived at the same type of expression for the equilibrium constant as we did in our intitial example from statistical thermodynamics. From that, we know that the entropy term contained in K° = exp(ΔS°/R) contains vibrational terms only, and not the configurational entropy.

From the equation above we see that the temperature dependence of K is given by

$$\frac{d\ln K}{d(1/T)} = -\frac{\Delta H^0}{R} \tag{3.27}$$

and that a plot of lnK vs 1/T (van't Hoff plot) will give $\Delta S°/R$ as the intercept and $-\Delta H°/R$ as the slope.

This general treatment is applicable to any ideal solutions whether these are gaseous, liquid, or solid, and it applies also to defect chemistry as we have examplified initially.

Thermodynamics and point defects

Equilibrium thermodynamics may as stated above be applied to defect reactions. However, before so doing it is of interest to consider the thermodynamic properties of the individual point defects.

Virtual chemical potentials of point defects

The regular atoms on their normal sites and the point defects occupy particular sites in the crystal structure and these have been termed structural elements by Kröger, Stieltjes, and Vink (1959) (see also Kröger (1964)). As discussed in Chapter 2, the rules for writing defect reactions require that a definite ratio of sites is maintained due to the restraint of the crystal structure of the compounds. Thus if a normal site of one of the constituents in a binary compound MO is created or annihilated, a normal site of the other constituent must simultaneously be created or annihilated.

As discussed above the chemical potential of a constituent represents the differential of the Gibbs energy with respect to the amount of the constituent when the pressure, temperature and the amounts of all other constituents are kept constant. When a structure element is created, the number of complementary structure elements can not be kept constant due to the requirement of a definite site ratio, and it is therefore not possible to assign a true chemical potential to a structure element. However, Kröger et al. have shown that one may get around this difficulty by assigning a virtual chemical potential ζ (zeta) to each separate structure element and that the virtual chemical potential behaves like a true chemical potential and may be related in a in a similar way to activity:

$$\zeta_i = \zeta_i^0 + RT\ln a_i \tag{3.28}$$

The virtual potential differs from the true chemical potential by an undefined constant which is incorporated in $\zeta°$. For this reason the absolute value of the virtual chemical potential can not be determined experimentally.

In real changes in crystals where the site ratio is maintained and complementary structure elements (or building units) are created or annihilated, the undefined constants in the virtual chemical potential of each separate structure element cancel out, and the overall change is characterised by real changes in

Gibbs free energies. Real changes may thereby be described in terms of virtual potentials. Kröger et al. pointed out that a similar situation occurs in aqueous electrolytes. In this case, arbitrary thermodynamic potentials are assigned to separate ions, although it is impossible to determine experimentally the separate potentials.

Ideal and non-ideal models

As described in the preceding section the defect reactions and equilibria can be described as chemical reactions and treated in terms of the law of mass action. It is emphasised that the equilibrium constants relate activities of the structure elements involved in the defect reaction, but under ideal conditions and when the structure elements can be assumed to be randomly distributed over the available sites, the activities equal the concentrations of the structure elements. This is a valid assumption for very dilute solutions. However, for larger defect concentrations interactions take place between the defects, etc. and in principle activities should be used instead of concentrations.

The interactions between charged defects may be accounted for by using the Debye-Hückel theory in analogy with the interactions of ions in aqueous solutions. This requires knowledge of the relative dielectric constant, the smallest distance between charged defects, and other parameters for the solid. Debye-Hückel corrections have, for instance, been worked out and tested for cation vacancy defects in metal-deficient $Co_{1-y}O$ and $Ni_{1-y}O$. At infinite dilution the metal vacancies can be considered to have two negative charges, $v_{Co}^{//}$, and the equivalent number of electron holes is assumed to be randomly distributed in the oxide. With an increasing concentration of point defects and electron holes, there will be an increasing association of the metal vacancies (with two effective negative charges) and the positively charged electron holes. In these terms, the metal vacancies can be considered surrounded by a cloud of electron holes. This results in a deviation from a random distribution of electron holes and a corresponding deviation from the ideal solution model. For the case of $Co_{1-y}O$ the Debye-Hückel correction can explain the increasing deviation from ideality with increasing defect concentration. However, this model in its present form does not satisfactorily explain various electrical conductivity results for $Co_{1-y}O$ and so far apparently fails to give a full description of the defect structure and defect-dependent properties of the oxide.

An alternative approach is to consider that the association of defects leads to the formation of new "associated" defect species. These "new" defects are, in turn, treated in terms of the ideal solution models. Specifically, in the case of metal vacancies in, for instance, $Co_{1-y}O$ and $Ni_{1-y}O$ it is assumed that the metal vacancies with two effective charges become associated with one electron hole to give a metal vacancy with one negative effective charge, $v_{Co}^{/}$, as has already been described in Chapter 2. At this stage this simple approach appears to give an equally consistent description of the defect structure and defect-dependent properties of $Co_{1-y}O$ as the much more complicated Debye-Hückel model. However, it is to be expected that future developments will provide a more detailed understanding of defect interactions and how they influence defect-dependent properties of inorganic compounds.

Despite its shortcomings the ideal solution approach will be used in this book in treating defect equilibria. Defect interactions will generally be treated by assuming the formation of associated defects as is, for instance, exemplified above for the formation of singly charged metal vacancies through the association of an electron hole with a doubly charged metal vacancy. Furthermore, the defect equilibria will be treated using the law of mass action.

Concentrations of the structure elements may be expressed in many units. In semiconductor physics it is common to express the concentrations in number per cm^3. This has the disadvantage, however, that when two compounds with different molecular (packing) densities contain the same number of defects per cm^3, the number of defects per molecular unit is different in the two compounds, and *vice versa*. Furthermore, the molecular unit size (packing density, or unit cell parameters) of a compound changes as a function of its nonstoichiometry and temperature, and in this case the unit of defects per cm^3 does not unequivocally reflect the relative defect concentration per molecular unit or atom site under different conditions.

In the following, concentration will in most cases be expressed as the number of defects or atoms per molecule or site, i.e. in molar or site fractions, while we for some purposes need to use the number of defects per cm^3. As mentioned before, the value of the equilibrium constant depends on the units of concentration that are employed, but it is a simple matter to convert the values of the equilibrium constant from one system to another.

Examples of defect equilibria in pure, stoichiometric metal oxides

In this section we will see how we write defect reactions for selected simple cases of combinations of defects, express the equilibrium constants, and combine this with an electronutreality condition to express the concentration of dominating defects as a function of temperature and activities. We will stay strictly with statistical thermodynamics, such that entropy changes in the defect reaction can be interpreted strictly as a change in the vibrational entropy. Approximations will be implemented in order to visualise the simplicity f the method.

Schottky defects

As illustration let us consider the defect equilibrium for Schottky defects in the oxide MO and where the metal and oxygen vacancies are doubly charged. The defect equation for their formation is given in Chapter 2 as

$$0 = v_O^{\bullet\bullet} + v_M^{//} \tag{3.29}$$

The corresponding defect equilibrium may at low defect concentrations be written

$$K_S = X_{v_O^{\bullet\bullet}} X_{v_M^{//}} = \frac{[v_O^{\bullet\bullet}]}{[O]} \frac{[v_M^{//}]}{[M]} \tag{3.30}$$

where K_S is the equilibrium constant for a Schottky defect pair, X represent site fractions, and square brackets express concentrations of species. [O] and [M] represent the concentration of oxygen and metal sites, respectively.

If we let square brackets express concentration in mole fraction (i.e. the number of the species per mole of MO), then [O] = [M] = 1, and then

$$K_S = [v_O^{\bullet\bullet}][v_M^{//}] \tag{3.31}$$

When the concentrations of the defects are expressed in site fractions, or properly converted to mole fractions as here, K_S is related to the Gibbs molar energy, ΔG_S, enthalpy change, ΔH_S, and entropy change, ΔS_S, of formation of pairs of doubly charged Schottky defects through the relation above:

$$K_S = [v_O^{\bullet\bullet}][v_M^{//}] = \exp(-\frac{\Delta G_S}{RT}) = \exp(\frac{\Delta S_S}{R})\exp(-\frac{\Delta H_S}{RT}) \tag{3.32}$$

It may be useful to repeat that the entropy change , ΔS_S, does not include the configurational entropy changes associated with the formation of the defects, but mainly involves the change in vibrational entropy.

When the Schottky pair dominates the defect structure, we may see from the reaction equation or from the electroneutrality condition that in MO the concentrations of the metal and oxygen vacancies are equal:

$$[v_O^{\bullet\bullet}] = [v_M^{//}] \tag{3.33}$$

Insertion into the preceding equation yields

$$[v_O^{\bullet\bullet}] = [v_M^{//}] = K_S^{1/2} = \exp(\frac{\Delta S_S}{2R})\exp(-\frac{\Delta H_S}{2RT}) \tag{3.34}$$

Under these conditions the defect concentrations are only dependent on the temperature; they are independent of the activities of the components M and O_2. The slope in a van't Hoff plot of lnK vs 1/T would be $-\Delta H_S/2R$. The factor 2 results from the square root of the equilibrium constant, in turn resulting from the fact that there are *two* defects created, in comparison with the one we got in the element used as example earlier.

It may be useful to note that the procedure we have applied is to write down the defect reaction and the mass action law expression for its equilibrium constant, and the prevailing electronuetrality condition. Thus, we have two independent equations for the two unknowns (provided that either K_S or all of ΔS_S, ΔH_S, and T are known).

Frenkel defect pairs

In our next example we consider Frenkel defect pairs in MO. If, for the sake of illustration, the Frenkel defects in MO are assumed to be doubly charged, the defect equation for their formation is given as

$$M_M^x + v_i^x = M_i^{\bullet\bullet} + v_M^{//} \tag{3.35}$$

It may be noted that a vacant interstitial site has been included to keep track of the sites where the interstitials are going. The corresponding defect equilibrium is written as

$$K_F = \frac{X_{M_i^{\bullet\bullet}} \mathrm{X}_{v_M^{//}}}{X_{M_M^x} X_{v_i^x}} = \frac{\dfrac{[M_i^{\bullet\bullet}]}{[\mathrm{i}]}\dfrac{[v_M^{//}]}{[M]}}{\dfrac{[M_M^x]}{[M]}\dfrac{[v_i^x]}{[i]}} = \frac{[M_i^{\bullet\bullet}][v_M^{//}]}{[M_M^x][v_i^x]} \tag{3.36}$$

If there is one interstitial site per MO, and if the defect concentrations are small such that $[M_M^x] = [v_i^x] = 1$, we obtain the simpler

$$\mathrm{K_F} = [M_i^{\bullet\bullet}][v_M^{//}] \tag{3.37}$$

If Frenkel defects predominate in the pure stoichiometric compound, the concentrations of interstitial cation and vacancies are then equal, and we obtain:

$$[M_i^{\bullet\bullet}] = [v_M^{//}] = K_\mathrm{F}^{1/2} = \exp(\frac{\Delta \mathrm{S_F}}{2R})\exp(-\frac{\Delta \mathrm{H_F}}{2RT}) \tag{3.38}$$

where $\Delta \mathrm{H_F}$ is the enthalpy of formation of Frenkel defect pairs. Under these conditions the concentration of the Frenkel defect pairs is independent of the activities of the metal and oxygen.

Intrinsic ionisation of electrons.

For many oxides, and particularly when considering defect structure situations close to stoichiometry, it is essential to take into account the intrinsic ionisation of electrons. This can include localised defects (valence defects) or delocalised defects (valence band and conduction band).

In the case of valence defects, the ionisation of a pure binary metal oxide $\mathrm{MO_a}$ may typically be assigned to mixed valency of the metal and thus be written as

$$2M_M^x = M_M^{/} + M_M^{\bullet} \tag{3.39}$$

The equilibrium constant may then be expressed as site fractions, which in our case immediately simplifies into an expression containing only volume or molar concentrations:

$$\mathrm{K_i} = \frac{[M_M^{\bullet}][M_M^{/}]}{[M_M^x]^2} \tag{3.40}$$

If the intrinsic ionisation dominates the defect structure, we have

$$[M_M^{\bullet}] = [M_M^{/}] \tag{3.41}$$

Insertion into the equilibrium constant yields

$$[M_M^{\bullet}] = [M_M^{/}] = [M_M^x]\mathrm{K_i^{1/2}} = [M_M^x]\exp(\frac{\Delta S_i}{2\mathrm{R}})\exp(-\frac{\Delta H_i}{2\mathrm{RT}}) \tag{3.42}$$

In the case of delocalised electrons we write the ionisation reaction

$$0 = \mathrm{e}^{/} + \mathrm{h}^{\bullet} \tag{3.43}$$

In this model, classical statistical thermodynamics cannot be applied, and we abandon the accounting of sites and configurational entropy. Instead we enter simply the volume concentrations of defects in the widespread definition of the intrinsic ionisation constant:

$$\mathrm{K_i^{/}} = [e^{/}][h^{\bullet}] = \mathrm{n\,p} \tag{3.44}$$

where $\mathrm{K_i^{/}}$ is the equilibrium constant. The temperature dependency of the product np is mainly given by the ionisation energy, most often referred to as the band gap, $\mathrm{E_g}$. This has a slight temperature dependency that may be expressed the same way as we usually do for Gibbs energies and equilibrium constants:

$$\mathrm{K_i^{/}} = \exp(-\frac{E_g}{RT}) = \mathrm{K_{0,i}^{/}}\exp(-\frac{\Delta H_i}{RT}) \tag{3.45}$$

If the concentrations of electrons and electron holes predominate, then the electroneutrality condition may be approximated by

$$\mathrm{n = p = (K_i^{/})^{1/2}} = \exp(-\frac{E_g}{2RT}) = (\mathrm{K_{0,i}^{/}})^{1/2}\exp(-\frac{\Delta H_i}{2RT}) \tag{3.46}$$

Intrinsic ionization of electrons and holes is, for instance, concluded to apply to Fe_2O_3 and Cr_2O_3 at high temperatures.

Defect equilibria in nonstoichiometric oxides

Here, we will employ mainly the same methodology as in the preceding section, but when delocalised defect electrons and holes are introduced we cannot stay strictly with statistical thermodynamics and the equilibrium thermodynamics cannot be related in the same simple manner to the change in vibrational entropy any longer.

Oxygen deficient oxides

In oxides with oxygen deficit, the predominant defects are oxygen vacancies. If these are fully ionised (doubly charged), and if the electronic defects are localised as valency defects, their formation reaction and the corresponding equilibrium can be written:

$$O_O^x + 2M_M^x = v_O^{\bullet\bullet} + 2M_M^/ + \tfrac{1}{2}O_2(g) \tag{3.47}$$

$$K_{vO} = \frac{X_{v_O^{\bullet\bullet}} X_{M_M^/}^2}{X_{O_O^x} X_{M_M^x}^2} p_{O_2}^{1/2} = \frac{[v_O^{\bullet\bullet}][M_M^/]^2}{[O_O^x][M_M^x]^2} p_{O_2}^{1/2} \tag{3.48}$$

If these oxygen vacancies and the compensating electronic valence defects are the predominating defects in the oxygen deficient oxide, the principle of electroneutrality requires that

$$[M_M^/] = 2[v_O^{\bullet\bullet}] \tag{3.49}$$

By insertion, and assuming the oxide is MO and that $[M_M^x] = [O_O^x] = 1$ (small defect concentrations) we then obtain:

$$[\mathrm{M}_M^/] = 2[\mathrm{v}_O^{\bullet\bullet}] = (2K_{vO})^{1/3} p_{O_2}^{-1/6} = 2^{1/3}\exp(\frac{\Delta \mathrm{S}_{vO}}{3R})\exp(-\frac{\Delta \mathrm{H}_{vO}}{3RT}) p_{O_2}^{-1/6} \tag{3.50}$$

This shows that the concentrations of the dominating defects (defect electrons and oxygen vacancies) increase with decreasing oxygen partial pressure. It furthermore shows the relation between defect concentrations and entropy and enthalpy changes of the defect reaction. The factor 3 that enters results from the formation of 3 defects in the defect reaction.

Now, let us do the same treatment, but for an oxide where defect electrons are delocalised in the conduction band. The formation of oxygen vacancies is then written:

$$O_O^x = v_O^{\bullet\bullet} + 2e^/ + \tfrac{1}{2}O_2(g) \tag{3.51}$$

Since delocalised electrons in a band cannot be treated by classical statistics, we abandon the site fraction principle, and express all activities simply as concentrations, and denote the new equilibrium constant with a prime[4]:

$$K_{vO}^{/} = \frac{[v_O^{\bullet\bullet}]}{[O_O^x]} \mathrm{n}^2 p_{O_2}^{1/2} \tag{3.52}$$

If these oxygen vacancies and the compensating electrons are the predominating defects in the oxygen deficient oxide, the principle of electroneutrality requires that

$$\mathrm{n} = 2[v_O^{\bullet\bullet}] \tag{3.53}$$

By insertion, and assuming that $[O_O^x] = 1$ (small defect concentrations) we then obtain:

$$\mathrm{n} = 2[\mathrm{v}_O^{\bullet\bullet}] = (2K_{vO}^{/})^{1/3} p_{O_2}^{-1/6} = (2\mathrm{K}_{0,\mathrm{vO}}^{/})^{1/3} \exp(-\frac{\Delta \mathrm{H}_{\mathrm{vO}}}{3RT}) p_{O_2}^{-1/6} \tag{3.54}$$

and deliberately use a pre-exponential K_0 instead of an entropy change. Otherwise, the solution is the same as in the case with localised valence defects for the electrons.

A plot of log *n* or $\log [v_O^{\bullet\bullet}]$ vs $\log \mathrm{p}_{\mathrm{O}_2}$ (at constant temperature) will give straight lines with a slope of –1/6. Such plots are called Brouwer diagrams, and they are commonly used to illustrate schematically the behaviour of defect concentrations under simplified limiting cases of dominating defects.

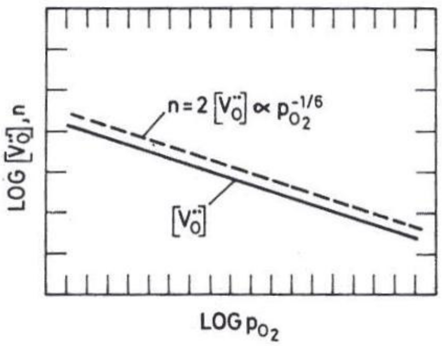

Figure 3-2. Brouwer diagram for $\mathrm{n} = 2[v_O^{\bullet\bullet}]$ as the electroneutrality condition.

---

[4] This text introduces the use of a prime in the equilibrium constant to denote cases where statistical thermodynamics has not been used throughout, and where the entropy of the reaction thus is not well defined. This notation is new and not in common use, as is neither the distinction it notifies. One may therefore expect also that this text relaxes its stringency in this respect in later chapters, where the distinction is less in focus.

If we want to investigate the more complex situation of involving neutral and partially ionised oxygen vacancies and the excitation of the trapped electrons (described in Chapter 2) we start by writing down the stepwise reactions and their corresponding equilibria:

$$O_O^x = v_O^x + \tfrac{1}{2}O_2(g) \tag{3.55}$$

$$K_{vOx} = \frac{[v_O^x]}{[O_O^x]} p_{O_2}^{1/2} \tag{3.56}$$

$$v_O^x = v_O^\bullet + e^/ \tag{3.57}$$

$$K_{vO1}^/ = \frac{[v_O^\bullet]}{[v_O^x]}\mathrm{n} \tag{3.58}$$

$$v_O^\bullet = v_O^{\bullet\bullet} + e^/ \tag{3.59}$$

$$K_{vO2}^/ = \frac{[v_O^{\bullet\bullet}]}{[v_O^\bullet]}\mathrm{n} \tag{3.60}$$

where $K_{vOx}, K_{vO1}^/, \text{and } K_{vO2}^/$ are the respective equilibrium constants. Now the principle of electroneutrality requires that

$$\mathrm{n} = [v_O^\bullet] + 2[v_O^{\bullet\bullet}] \tag{3.61}$$

The concentrations of the electrons and the neutral, singly, and doubly charged oxygen vacancies are related through the equations above, and by combination expressions for each of the defects may be obtained. The electron concentration is given by

$$n^3 = K_{vOx}K_{vO1}^/(2K_{vO2}^/ + n)p_{O_2}^{-1/2} \tag{3.62}$$

This equation has two limiting conditions.

If $n << 2K_{vO2}^/$, then

$$n = 2[v_O^{\bullet\bullet}] = (2K_{vOx}K_{vO1}^/K_{vO2}^/)^{1/3} p_{O_2}^{-1/6} \tag{3.63}$$

Under these conditions the concentrations of electrons and oxygen vacancies are relatively small, and under these conditions the doubly charged vacancies are the predominating oxygen vacancies, and the concentrations of electrons and oxygen vacancies are proportional to $p_{O_2}^{-1/6}$. Of course, this situation corresponds to the simple case presented initially for the doubly charged vacancies, and obviously $K_{vOx}K_{vO2}^/K_{vO2}^/ = K_{vO}^/$.

If, on the other hand, $n >> 2K'_{vO2}$, then

$$n = [v_O^{\bullet}] = (K_{vOx}K'_{vO1})^{1/2}\, p_{O_2}^{-1/4} \tag{3.64}$$

Under these conditions the concentration of electrons is relatively large, and the predominant oxygen vacancies are then singly charged. Furthermore, the concentrations of electrons and oxygen vacancies are then proportional to $p_{O_2}^{-1/4}$.

A general tendency similar to that of oxygen deficient oxides applies to metal deficient oxides; in the oxide $M_{1-y}O$ the metal vacancies are doubly charged at very small deviations from stoichiometry and tend to become singly charged with increasing nonstoichiometry.

It may be noted that the neutral oxygen vacancies are not affected by charged defects, nor do they affect the electroneutrality.

Oxide with excess metal

We will as an example consider the formation of fully ionised interstitial metal ions and complementary electrons in an oxide with excess metal $M_{2+y}O_3$. The equation for the formation of interstitial metal ions with three effective positive charges and three complementary electrons is given by

$$\tfrac{3}{2}O_O^x + M_M^x = M_i^{\bullet\bullet\bullet} + 3e' + \tfrac{3}{4}O_2(g) \tag{3.65}$$

The corresponding defect equilibrium is given by

$$K'_{Mi} = [M_i^{\bullet\bullet\bullet}]\mathrm{n}^3 p_{O_2}^{3/4}[O_O^x]^{-3/2}[M_M^x]^{-1} \tag{3.66}$$

where K_{Mi} is the equilibrium constant. If these defects are the predominating ones, and we as before assume small defect concentrations, the electroneutrality condition and the oxygen pressure dependence of the interstitial metal ions and the electrons is

$$3[M_i^{\bullet\bullet\bullet}] = \mathrm{n} = (3K'_{Mi})^{1/4}\, p_{O_2}^{-3/16} \tag{3.67}$$

Thus, also in this case the concentration of the defects increases with decreasing oxygen activity and is proportional to $p_{\mathrm{O_2}}^{-3/16}$. This oxygen pressure dependence is different from that for formation of singly as well as doubly charged oxygen vacancies. In such a case it would thus in principle be possible to decide from measurements of electron concentration (e.g. via electrical conductivity) or of nonstoichiometry as a function of oxygen activity whether the predominating defects are triply charged interstitial metal ions or oxygen vacancies with one or two charges.

However, had the metal interstitials predominantly been doubly charged in the oxide $M_{2+y}O_3$, a $p_{\mathrm{O_2}}^{-1/4}$ dependence would have been the result, and additional studies of defect-dependent properties (e.g. self-diffusion of metal or oxygen) would be needed to distinguish this situation from that of singly charged oxygen vacancies.

It may be noted that it is not only the absolute number of charges on a defect that determines the p_{O_2} dependence, but also the difference between the actual charge and the charge given by the nominal valence of the atoms involved. Thus, in the oxide MO (or $M_{1+y}O$) doubly charged metal interstitials will give a $p_{O_2}^{-1/6}$ dependence of the defect concentrations (as opposed to the case above) and in this case be indistinguishable from doubly charged oxygen vacancies.

Simultaneous presence of oxygen vacancies and interstitial metal ions

In an oxide M_aO_b where the ratio of metal to oxygen is larger than the stoichiometric ratio a:b it is *a priori* difficult to predict whether interstitial metal ions or oxygen vacancies predominate. In principle both types of defects may be important, at least in certain regions of nonstoichiometry. In the following, let us consider a case where it may be necessary to take into account the simultaneous presence of interstitial metal ions and oxygen vacancies.

Consider an oxide with a stoichiometric composition MO_2. Let us further assume for the sake of illustration that when the oxide is nonstoichiometric the important point defects are doubly charged oxygen vacancies and doubly charged interstitial metal ions. (It mat be noted that the metal interstitials are not fully ionised in this case.) The composition of the nonstoichiometric oxide may accordingly be written $M_{1+y}O_{2-x}$.

The defect equations for the formation of these two types of defects may be written

$$O_O^x = v_O^{\bullet\bullet} + 2e^{/} + \tfrac{1}{2}O_2(g) \tag{3.68}$$

$$2O_O^x + M_M^x = M_i^{\bullet\bullet} + 2e^{/} + O_2(g) \tag{3.69}$$

The corresponding defect equilibria (assuming $[M_M]$ and $[O_O]$ to be equal to unity) then become

$$K_{vO}^{/} = [v_O^{\bullet\bullet}]\mathrm{n}^2 p_{O_2}^{1/2}[O_O^x]^{-1} \tag{3.70}$$

$$K_{Mi2}^{/} = [M_i^{\bullet\bullet}]\mathrm{n}^2 p_{O_2}[O_O^x]^{-2}[M_M^x]^{-1} \tag{3.71}$$

The electroneutrality condition is given by

$$n = 2[v_O^{\bullet\bullet}] + 2[M_i^{\bullet\bullet}] \tag{3.72}$$

Two limiting conditions may be considered:

When $[v_O^{\bullet\bullet}] >> [M_i^{\bullet\bullet}]$ then

$$n = 2[v_O^{\bullet\bullet}] = (2K'_{vO})^{1/3} p_{O_2}^{-1/6} \qquad (3.73)$$

as obtained also earlier for the same conditions. By inserting this relationship for n into the equilibrium constant for formation of the minority defects, metal interstitials, we obtain

$$[M_i^{\bullet\bullet}] = K'_{Mi2}(2K'_{vO})^{-2/3} p_{O_2}^{-2/3} \qquad (3.74)$$

Under these conditions the minority concentration of metal interstitials, $[M_i^{\bullet\bullet}]$, increases rapidly with decreasing oxygen partial pressure, more rapidly than the two dominating defects, and may eventually catch up with them and become dominating at the expense of oxygen vacancies, as illustrated in the figure below.

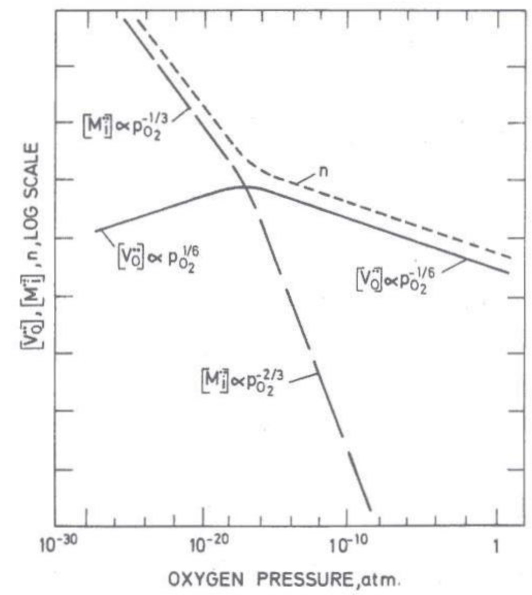

Figure 3-3. Schematic presentation defect concentrations as a function of oxygen pressure in an oxide MO_2 containing both doubly charged oxygen vacancies and doubly charged interstitial metal ions, compensated by electrons.

When, as the other limiting case, $[M_i^{\bullet\bullet}] >> [v_O^{\bullet\bullet}]$ then

$$n = 2[M_i^{\bullet\bullet}] = (2K'_{Mi})^{1/3} p_{O_2}^{-1/3} \qquad (3.75)$$

By inserting this relationship for n into the equilibrium constant for formation of the minority defect, oxygen vacancies, we obtain

$$[v_O^{\bullet\bullet}] = K'_{vO}(2K'_{Mi2})^{-2/3} p_{O_2}^{1/6} \qquad (3.76)$$

It may be noted that under these conditions the concentration of oxygen vacancies *increases* with the oxygen pressure, as shown in the figure.

The overall situation may be illustrated schematically as shown in Fig.3.3, in which the concentrations of the three defects are plotted as a function of the partial pressure of oxygen. In this diagram it is assumed that $[v_O^{\bullet\bullet}]$ is 10^5 times larger than $[M_i^{\bullet\bullet}]$ at 1 atm O_2, and with this assumption it is seen that the interstitial cations will predominate at a partial pressure below about 10^{-17} atm. Many of the high melting oxides are stable over large oxygen pressure ranges and it is not improbable that this hypothetical situation, with oxygen vacancies predominating at high partial pressures of oxygen (and small deviations from stoichiometry) and interstitial cations predominating at low partial pressures (relatively large deviations from stoichiometry) may generally apply to oxygen deficient oxides at elevated temperatures. In actual cases more than one ionisation state of the defects may of course also have to be considered.

In this section we have employed and illustrated a very important sequence of actions often used to solve ideal situations in defect chemistry, but also implicitly used in traditional aqueous and other chemistry. We write the list down in a numbered list for clarity:

1. Write down the full electroneutrality condition containing all charged defects.
2. Write down a number of independent chemical reactions and corresponding equilibrium constant expressions – if you have n defects you need normally n-1 such equilibria. (The n'th expression is the electroneutrality itself.)
3. Decide a pair of dominating charged defects, and simplify the electroneutrality condition to this limiting situation.
4. Insert this limiting condition into an appropriately chosen defect equilibrium to obtain the expression for the concentration of the two dominating defects. (Normally, simplifications such as assuming small defect concentrations are made here.)
5. Insert the obtained expression into another equilibrium expression and solve to obtain an expression of the concentration of a minority defect.
6. Extrapolate to see what minority defect comes up to become dominating as conditions change. Then repeat from step 3 until all defects are exhausted.

On the basis of such an exercise one may construct schematic diagrams of how defect pairs dominate defect structures and how minority defects behave and eventually take over. Transition zones are not solved explicitly in this way and are thus often drawn sharp and schematically. Logarithmic depictions are common, and such plots are then called Brouwer diagrams.

Metal-deficient oxide

The reactions and defect equilibria for the formation of single, unassociated neutral metal vacancies and the subsequent excitation of electron holes in MO may for small defect concentrations be written

$$\tfrac{1}{2}O_2(g) = v_M^x + O_O^x \tag{3.77}$$

$$K_{vMx} = [v_M^x][O_O^x]p_{O_2}^{-1/2} \tag{3.78}$$

$$v_M^x = v_M^/ + h^\bullet \tag{3.79}$$

$$K_{vM1}^/ = \frac{[v_M^/]}{[v_M^x]}\mathrm{p} \tag{3.80}$$

$$v_M^/ = v_M^{//} + h^\bullet \tag{3.81}$$

$$K_{vM2}^/ = \frac{[v_M^{//}]}{[v_M^/]}\mathrm{p} \tag{3.82}$$

where K_{vMx}, $K_{vM1}^/$, and $K_{vM2}^/$ are the equilibrium constants for the respective defect equilibria.

When the metal vacancies and their complementary electron holes are the predominating defects, the electroneutrality condition reads

$$\mathrm{p} = [v_M^/] + 2[v_M^{//}] \tag{3.83}$$

Through combinations of the equilibrium constant expressions and electroneutrality above, the concentration of the separate point defects and electronic defects can be evaluated. In an oxide $M_{1-y}O$ this leads to relationships similar to that for oxygen vacancies with the exception that the concentrations of electron holes and metal vacancies always increase with increasing oxygen pressure. The concentration of electron holes is, for instance, determined by the expreession

$$p^3 = K_{vMx}K_{vM1}^/(2K_{vM2}^/ + p)p_{O_2}^{1/2} \tag{3.84}$$

As regards the electron holes two limiting conditions may be considered:

At low concentrations of electron holes (small deviations from stoichiometry; $p << K_{vM2}^/$) we have

$$p = (2K_{vMx}K_{vM1}^/K_{vM2}^/)^{1/3}p_{O_2}^{1/6} \tag{3.85}$$

At high concentrations of electron holes ($p >> K_{vM2}^/$) we obtain

$$p = (K_{vMx} K'_{vM1})^{1/2} p_{O_2}^{1/4} \tag{3.86}$$

The oxygen pressure dependence of the concentration of electron holes and the total concentration of the charged metal vacancies (deviation from stoichiometry) correspondingly changes from $p_{O_2}^{1/6}$ to $p_{O_2}^{1/4}$ with increasing deviation from stoichiometry.

Metal oxides with excess oxygen.

The defect equilibria involving the formation of interstitial oxygen atoms or ions in oxides with excess oxygen may be set up following the same treatment as applied to the other defects which have been dealt with above. The concentration of interstitial oxygen species increases with p_{O_2}, and the formation of charged oxygen interstitials is accompanied by the formation of electron holes.

Examples of defect structures involving both oxygen deficiency and excess

In the preceding examples of defect structures in non-stoichiometric oxides it has been assumed that the oxides either have an oxygen deficit (or excess metal) or excess oxygen (or metal deficit). In many oxides the predominating defect may change from one type to another depending on the oxygen activity. As an illustration of such a defect structure situation a hypothetical case will be considered where an oxide predominantly contains oxygen vacancies at reduced oxygen activities and interstitial oxygen ions at high oxygen activities. In an intermediate region the oxide will be stoichiometric or close to stoichiometric. Some oxides with the fluorite structure exhibit such a defect structure, e.g. $UO_{2\pm x}$.

For the sake of simplicity in illustrating such a defect structure situation it will be assumed that both the interstitial oxygen ions and the oxygen vacancies are doubly charged. In this case it will then be necessary to consider the following equilibria, for the formation of oxygen vacancies, oxygen interstitials, electrons and holes, and for anion Frenkel pairs:

$$K'_{vO} = [v_O^{\bullet\bullet}]\mathrm{n}^2 p_{O_2}^{1/2} \tag{3.87}$$

$$K'_{Oi} = [O_i^{//}]\mathrm{p}^2 p_{O_2}^{-1/2} \tag{3.88}$$

$$K'_i = np \tag{3.89}$$

$$K_{AF} = [O_i^{//}][v_O^{\bullet\bullet}] \tag{3.90}$$

In these equilibria we have assumed small defect concentrations such that the concentrations of normal lattice sites and empty interstitial sites have been assumed constant and equal to unity and thus omitted from the expressions.

It should be noted that the defect equilibria are interrelated, and through a combination of the equations it may be shown that $K_i^{/} K_{AF}^2 = K_{vO}^{/} K_{Oi}^{/}$. Thus, only three out of the four equilibria are sufficient to describe the defect structure of the oxide.

The full electroneutrality condition is given by

$$2[v_O^{\bullet\bullet}] + \mathrm{p} = 2[O_i^{//}] + \mathrm{n} \tag{3.91}$$

We will explore the defect structure of the oxide by considering limiting conditions and follow the procedure listed before to construct Brouwer diagrams.

Oxygen deficit.

At large oxygen deficit the following approximation may be made

$$n = 2[v_O^{\bullet\bullet}] >> 2[O_i^{//}], \mathrm{p} \tag{3.92}$$

We first insert this into the appropriate equilibrium to find the concentrations of the dominating defects in the usual manner:

$$n = 2[v_O^{\bullet\bullet}] = (2K_{vO}^{/})^{1/3} p_{O_2}^{-1/6} \tag{3.93}$$

and then insert this into other equilibria to find the concentrations of the minority defects. By inserting the expression for the concentration of vacancies into the anion Frenkel equilibrium, we obtain for the concentration of interstitials:

$$[O_i^{\bullet\bullet}] = 2^{2/3} K_{AF} (K_{vO}^{/})^{-1/3} p_{O_2}^{1/6} \tag{3.94}$$

By inserting the expression for the concentration of electrons into the intrinsic electronic equilibrium, we obtain for the concentration of holes:

$$p = K_i^{/} (2K_{vO}^{/})^{-1/3} p_{O_2}^{1/6} \tag{3.95}$$

Oxygen excess

For relatively large excess oxygen, that is, when

$$p = 2[O_i^{//}] >> 2[v_O^{\bullet\bullet}], \mathrm{n} \tag{3.96}$$

we may in a manner analogous to the preceding case, derive the following relations:

$$p = 2[O_i^{//}] = (2K_{Oi}^{/})^{1/3} p_{O_2}^{1/6} \tag{3.97}$$

$$[v_O^{\bullet\bullet}] = 2^{2/3} K_{AF} (K_{Oi}^{/})^{-1/3} p_{O_2}^{-1/6} \tag{3.98}$$

$$n = K_i^{/} (2K_{Oi}^{/})^{-1/3} p_{O_2}^{-1/6} \tag{3.99}$$

Stoichiometric condition

At or close to stoichiometry two alternative limiting conditions must be considered, namely dominance by intrinsic electronic ionisation or by anion Frenkel disorder.

If intrinsic ionisation of electrons predominates, and thus

$$p = n = (K_i^{/})^{1/2} >> 2[O_i^{//}], 2[v_O^{\bullet\bullet}] \tag{3.100}$$

The concentrations of electrons and electron holes (n and p) are then independent of oxygen pressure. The point defect concentrations are obtained by insertion of the expression for the concentrations of the electronic defects into the appropriate equilibria, and we obtain:

$$[v_O^{\bullet\bullet}] = \frac{K_{vO}^{/}}{\mathrm{K}_\mathrm{i}^{/}} p_{O_2}^{-1/2} \tag{3.101}$$

$$[O_i^{//}] = K_{Oi}^{/} \mathrm{K}_\mathrm{i}^{/} p_{O_2}^{1/2} \tag{3.102}$$

A Brouwer diagram illustrating this case of stoichiometric condition, along with oxygen deficit and excess, is shown below.

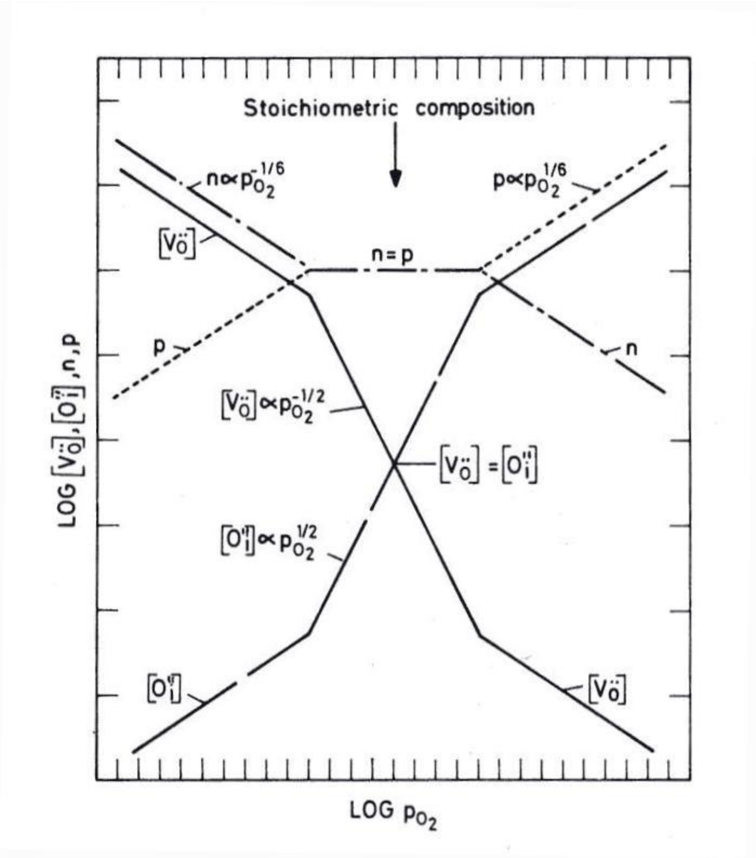

Figure 3-4. Schematic presentation of oxygen point defects and electronic defects as a function of oxygen pressure in an oxide which depending on the partial pressure of oxygen may have an excess or deficit of oxygen. Intrinsic electronic equilibrium is assumed to predominate at stoichiometric composition.

If, on the other hand, anion Frenkel disorder predominates under stoichiometric conditions, and thus

$$[O_i^{//}] = [v_O^{\bullet\bullet}] = \mathrm{K}_{\mathrm{AF}}^{1/2} >> \frac{p}{2}, \frac{n}{2} = \frac{(K_i^{/})^{1/2}}{2} \qquad (3.103)$$

then $[O_i^{//}]$ and $[v_O^{\bullet\bullet}]$ are independent of the partial pressure of oxygen, while the concentrations of electronic defects are given by

$$\mathrm{n} = (K_{vO}^{/})^{1/2} K_{AF}^{-1/4} p_{O_2}^{-1/4} \qquad (3.104)$$

$$\mathrm{p} = (K_{Oi}^{/})^{1/2} K_{AF}^{-1/4} p_{O_2}^{1/4} \qquad (3.105)$$

A Brouwer diagram illustrating this case of stoichiometric condition, along with oxygen deficit and excess, is shown below.

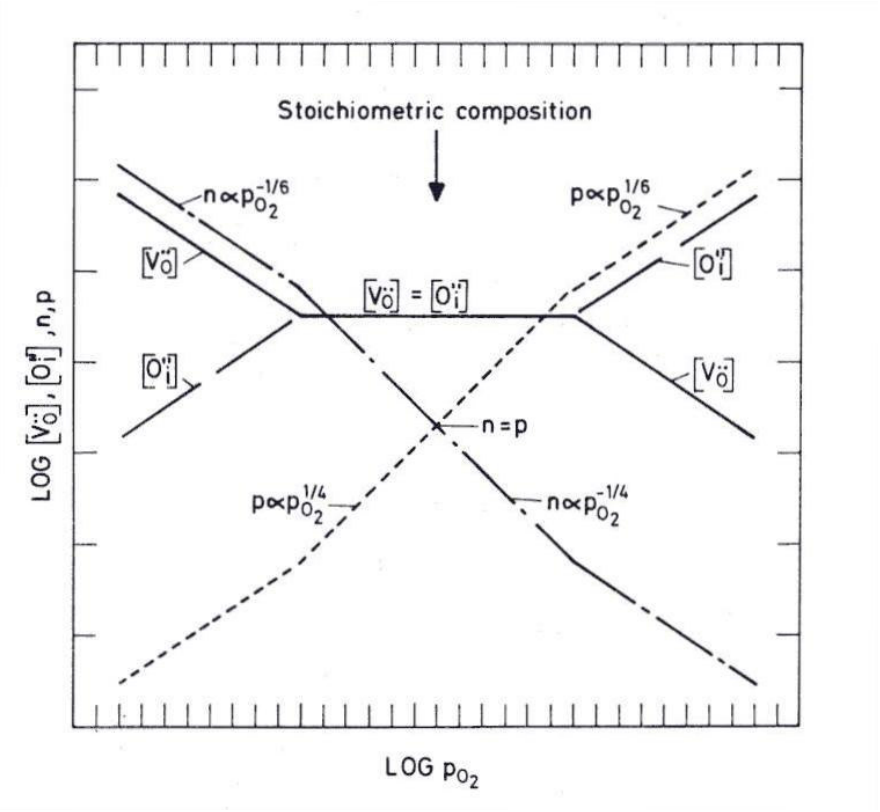

Figure 3-5. Schematic presentation of the concentration of oxygen point defects and electronic defects as a function of oxygen pressure. Oxygen defects are assumed to predominate at stoichiometric composition

From the diagrams we can conclude that at low oxygen activities the oxide has oxygen deficiency and will be an electronic n-type conductor, because the mobility of the electrons is always much higher than that of oxygen vacancies. At high oxygen partial pressures the oxide will correspondingly have oxygen excess and be a p-type electronic conductor. At intermediate oxygen activities it may be be a mixed n- and p-type electronic conductor in the case of intrinsic electronic disorder, while it may exhibit ionic or mixed ionic/electronic conduction in the case of anion Frenkel disorder.

Similar diagrams and analyses may be made for many other combinations of non-stoichiometric and stoichiometric defect situations of pure (undoped) oxides and other ionic compounds.

Summary

Chemical reactions for defects can be formulated and treated using the mass-action law. As for other chemical reactions, equilibrium constants can be defined in terms of the activities of the defects and other species. Under the normal constrictions we can approximate activities with concentrations of defects and partial pressures of gases. The equilibrium constants can also be expressed in terms of the Gibbs free energy change of the reaction, which in turn is expressed by the entropy and enthalpy changes.

We have seen that the standard entropy change of the standard Gibbs energy change contains only vibrational terms provided that the configurational terms are properly handled in expressing the equilibrium conditions. This is often possible in defect idealised defect chemistry since we can apply classical statistical thermodynamics. However, for non-classical defects such as delocalised electrons and holes this approach is less meaningful, and the entropy change may then be interpreted in a less straightforward manner. We have introduced the use of a prime (as in $K^{/}$) to denote equilibrium constants that are not expressed according to statistical thermodynamics and where the entropy change in the Gibbs energy change may have other than vibrationan contributions.

The mass action expressions can be combined with the full or limited cases of the electroneutrality condition to obtain exact or approximate (limiting case) expressions for the concentration of defects. Such concentrations are typically a function of the oxygen partial pressure, temperature, and doping. It is common to illustrate defect structures for oxides by plotting log defect concentrations vs log p_{O_2} (Brouwer diagrams) or ln or log defect concentrations vs 1/T. We have shown these principles and techniques through examples of intrinsic ionic and electronic disorder, various types of nonstoichiometry and variable ionisation of point defects.

We have restricted our treatment to simple cases and made simplifications where possible. Assumption of small defect concentrations have allowed the assumptions that the concentrations of normal lattice atoms and empty interstitials sites are constant, with activities equal to unity. Larger defect concentrations can to a first approximation be taken into account by including mass and site balances into the expressions used to solve the defect structure.)

Literature

Kröger, F.A. (1964) *The Chemistry of Imperfect Crystals*, North-Holland, Amsterdam, and Wiley, New York.

Kröger, F.A., Stieltjes, F.H. and Vink, H.J. (1959), *Philips Res. Rept.* **14**, 557.

Exercises

1. Do the insertion of Stirling's approximation and the derivation that leads up to the relation between equilibrium constant and entropy and enthalpy changes for the elemental reaction $E_E = v_E + E_E$ (Eqs. 3.3 and onwards).
2. Write the reaction for formation of Schottky defects in MO_2 and find an expression for the defect concentrations as a function of the equilibrium constant and thermodynamic parameters and temperature. Note in

particular how the solution deviates from the one obtained for the oxide MO treated in the text.

3. In the case of cation Frenkel defects in MO we assumed in the text that we had one interstitial site per MO. Derive the expression for the defect concentrations if the structure consists of fcc close-packed O ions, with M ions on each octahedral hole, and with interstitial sites on all tetrahedral holes (Hint: consult Chapter 1).
4. For a defect situation in an oxide dominated by doubly charged oxygen vacancies and electrons, sketch the van't Hoff plot (Logarithm of defect concentration vs 1/T) and a double-logarithmic plot of defect concentrations vs p_{O2} (Brouwer diagram).
5. Consider a metal oxide MO_{1-x} dominated by singly and doubly charged oxygen vacancies. Suggest a condition that expresses the changeover from dominance of one to dominance of the other. Next, find an expression for the pO_2 (as a function of equilibrium constants) of this transition point.
6. Consider a metal oxide $M_{2-x}O_3$ dominated by doubly and triply charged metal vacancies. Suggest a condition that expresses the changeover from dominance of the doubly charged to dominance of the triply charged. Next, find an expression for the pO_2 (as a function of equilibrium constants) of this transition point.
7. Find a general expression for m in the $[e^{/}] \propto p_{O2}{}^{1/m}$ relationship for oxides MO_b with non-stoichiometry dominated by fully ionised metal interstitials. (Hint: Find m as a function of b). Use this to find m in the case of M_2O, MO, M_2O_3, MO_2, M_2O_5 and MO_3.
8. Find a general expression for m (as a function of the valency z of the cation) in the $[h^{\cdot}] \propto p_{O2}{}^{1/m}$ relationship for binary oxides with non-stoichiometry dominated by fully ionised metal vacancies. Use this to find m in the case of M_2O_3.
9. Sketch a full Brouwer diagram (log defect concentrations vs log p_{O2}) for an oxide M_2O_3 dominated by fully ionised oxygen and metal vacancies at under- and overstoichiometry, respectively. Assume that Schottky defects predominate close to stoichiometric conditions. (Hint: the main goal is to obtain and illustrate the p_{O2}-dependencies. Use the rules we listed for such constructions of Brouwer diagrams.)
10. For the defect situation derived in the preceding exercise for M_2O_3 do the following analyses:
    i. find a condition and the expression for the p_{O2} when the oxide is fully stoichiometric.
    ii. Assume that you are able to determine this p_{O2} at various temperatures and plot ln $p_{O2,\ \text{stoichiometric}}$ vs 1/T. What information may you obtain from this plot?
    iii. Assume that you are able to determine n (the concentration of electrons) at the p_{O2} of exact stoichiometry at various temperatures and plot ln $n_{\text{stoichiometric}}$ vs 1/T. What information may you obtain from this plot?

iv. Assume that you are able to determine n (the concentration of electrons) at a fixed p_{O2} at various temperatures in the region of near stoichiometry and plot ln n vs 1/T. What information may you obtain from this plot?

11. Sketch a full Brouwer diagram (log defect concentrations vs log p_{O2}) for an oxide MO_2 dominated by fully ionised oxygen and metal vacancies at under- and overstoichiometry, respectively. Assume that intrinsic electronic equilibrium predominates close to stoichiometric conditions. (Hint: the main goal is to obtain and illustrate the p_{O2}-dependencies. Use the rules we listed for such constructions of Brouwer diagrams.)

12. For the defect situation derived in the preceding exercise for MO_2 do the following analyses:

    i. find a condition and the expression for the p_{O2} when the oxide is fully stoichiometric.

    ii. Assume that you are able to determine this p_{O2} at various temperatures and plot ln $p_{O2,\ \mathrm{stoichiometric}}$ vs 1/T. What information may you obtain from this plot?

    iii. Assume that you are able to determine n (the concentration of electrons) at various temperatures in the region of near stoichiometry and plot ln n vs 1/T. What information may you obtain from this plot?

13. Sketch a full Brouwer diagram (log defect concentrations vs log p_{O2}) for an oxide ABO_3 dominated by fully ionised oxygen and metal vacancies at under- and overstoichiometry, respectively. Assume that Schottky defects predominate close to stoichiometric conditions. You may assume that both cations are trivalent, but discuss also the effect it would have if A was divalent and B tetravalent. (Hint: the main goal is in any case to obtain and illustrate the p_{O2}-dependencies. Use the rules we listed for such constructions of Brouwer diagrams.)

14. Sketch a full Brouwer diagram (log defect concentrations vs log p_{O2}) for an oxide AB_2O_4 dominated by fully ionised metal interstitials and vacancies at under- and overstoichiometry, respectively. Assume that Frenkel defects predominate close to stoichiometric conditions. You may assume that A is divalent and B trivalent. (Hint: the main goal is to obtain and illustrate the p_{O2}-dependencies. Use the rules we listed for such constructions of Brouwer diagrams.)

15. In the text we considered an oxide MO_2 that had oxygen deficiency by both fully charged oxygen vacancies and doubly charged metal interstitials. The figure illustrating the case was drawn assuming that the ratio between the concentrations of the two defects was 10^5 at 1 atm. p_{O2}. Calculate the exact p_{O2} at which the concentrations are equal.

16. In the text we considered an oxide MO_2 that had oxygen deficiency by both fully charged oxygen vacancies and doubly charged metal interstitials. The figure illustrating the case is not strictly a Brouwer diagram, but rather a diagram showing the exact solution of the defect

structure, thus also correctly describing the transition region. The full solution can be reached in several ways:

i. By analytical solutions. Find and plot n (the concentration of electrons) and the concentrations of the other defects as a function of p_{O2} (double-logarithmically, over a reasonable range) using an analytical solution. Also, find and plot instead p_{O2} vs the concentration of one of the defects (again double-logarithmically, and over a reasonable and physical range). Were both approaches viable? Was one simpler?

ii. By a sequential calculation of an organised set of expressions. This may take any of the two analytical approaches mentioned above as starting point, but avoids finding the full analytical solution and instead lets a computer do all the calculations once the expressions are programmed[5].

iii. By a semi-analytical approach: In the full electroneutrality condition, replace each concentration by expressions such that only one defect is left. Then select p_{O2} and find the one defect concentration by minimising the error numerically[6].

iv. Numerically: Here one needs to adjust the parameters to fit a number of independent equations corresponding to the number of unknowns. Which equations would you include in the present case? Do you know a common method to fit them?[7]

---

[5] This approach can be implemented in Excel spreadsheets, as done for a number of complex defect structures of ternary oxides by F.W. Poulsen and others.

[6] This approach is implemented in software written by J. Abrantes.

[7] A program that solves the mixture of different types of equations numerically has been written by T. Norby. Compared to the other computerized approaches it executes very slowly.

4. Impurities and dopants

Introduction

In the preceding chapter we have only dealt with pure, crystalline metal oxides, and no account was taken of impurities and dopants and their effects on defect equilibria. Under real conditions it is impossible to produce truly pure crystals and for many oxides impurities may, depending on their concentration and the temperature, significantly affect or fully control defect concentrations in crystals. In the same terms dopants are purposely added to obtain or tailor defect-dependent properties of metal oxides and other materials.

The term dopant is from semiconductor science traditionally used for amounts of foreign atoms in the parts per million (ppm) range. However, many use it, as we shall here, also for foreign atoms in any concentration, up to several or even tens of percent. When one type of atoms is deliberately replaced to any degree by another type of atom, we may also use the term substituent instead of dopant.

Impurities and dopants may have various effects on crystals. Interstitially dissolved foreign atoms will normally affect the strain in the lattice. When ionised their charge affect the electroneutrality condition. Substitutionally dissolved impurities or dopants will also affect the properties of the host compound if they exhibit a difference in size or charge compared with the host atoms they replace.

In the following we will mainly consider cases where there is a difference in the valence between the foreign atom and the native atom it replaces. Such foreign atoms are termed *aliovalent*. Their effective charge will affect the electroneutrality condition and thereby the defect equilibria.

Larger dopant concentrations are generally expected to change also enthalpies of defect formation. This is, for instance, well demonstrated for oxides that have the same crystal structures and type of defects and that form solid solutions, e.g. solid solutions of NiO-CoO, MgO-NiO, a.o. However, we shall neglect such effects here and, as a first approximation, treat all cases as ideal dilute solutions, as we have done for defect equilibria earlier.

By the solubility of foreign atoms we mean the amount that can be held by the structure without precipitation of a second phase. This is thus the amount dissolved in equilibrium with that second phase. The solubility of foreign atoms obviously changes with temperature. Less recognised is the fact that the solubility under some defect structure situations also may vary with component (oxygen or metal) activities, as we shall see later.

On account of this it is important to distinguish between the following cases:

1) The concentration of impurity or dopant (solute) is **frozen in**. This is the case at relatively **low temperatures**; too low to allow transport to and from sinks and sources of the foreign atoms (a solute-rich phase). In this case the **concentration of solute is constant**. We may list some subcases of this situation:
    a) If the frozen concentration of solute is below the solubility limit, and no second phase is present, the system is then stable.
    b) If the concentration of solute is below the solubility limit, but a second solute-rich phase is present, the system is metastable.
    c) If the frozen concentration of solute exceeds the solubility limit, the system is metastable – a common case.
2) The concentration of impurity or dopant (solute) is in **equilibrium**. This is the case at relatively **high temperatures**; high enough to allow transport to and from sinks and sources of the foreign atoms. This case can be divided into two subcases:
    a) The concentration of solute is below the solubility limit. In this case there is no solute-rich phase present, and the concentration of solute is therefore again **constant**, as in all the cases 1).
    b) The concentration of solute is above the solubility limit. In this case there is a solute-rich second phase present, and the concentration of solute is **variable** by exchange of solute between the bulk and the sink/source of second phase.

In science and technology employing solid solutions (alloys or doped systems) it is often assumed that a constant concentration of solute is present. Furthermore, this is often assumed to be constant because it is fully soluble (case 2a), while it may in reality be constant because it is frozen in (cases 1).

During synthesis, sintering, equilibration, and use at high temperatures the concentration may change (case 2b) and this of course is also the basis for binary and higher phase diagrams. Defect chemistry offers a tool for understanding these changes, not only as a function of temperature, but especially as a function of component activities.

Despite this possibility, also defect-chemical treatments most commonly assume that the concentration of dopants and impurities is constant, often (tacitly) because they are assumed to be fully soluble (case 2a)). In reality, this may have been true at the high temperatures used during synthesis and fabrication, while during characterisation or use at lower temperatures the concentrations remain constant rather because they are frozen in (cases 1)). We shall start our treatment with cases of invariable concentrations of aliovalent foreign substituents, disregarding whether this is a result of full solubility or freezing-in.

Constant concentration of impurity/dopant

Schottky equilibria

As the predominance of Schottky defect situations have primarily been demonstrated in detail for ionic halides, e.g. alkali halides (NaCl, KBr, LiI a.o.), let us consider the effects of impurities/dopants on Schottky equilibria in a compound MX where the cations and anions have a valence of 1 and the cation and anion vacancies are singly charged, i.e. $v_X^\bullet$ and $v_M^/$. If divalent foreign cations, Mf^{2+}, are dissolved substitutionally in the lattice and occupy the normal M-sites, the Mf^{2+} ions will have one effective positive charge, $Mf_M^\bullet$. (It has donated an electron, and Mf is called a donor dopant). If we disregard other native defects than the Schottky defects, the electroneutrality condition becomes

$$[Mf_M^\bullet] + [v_X^\bullet] = [v_M^/] \tag{4.1}$$

The defect structure accordingly has two domains given by the limiting cases of this expression:

If the foreign cation is a minority species, i.e., $[Mf_M^\bullet] << [v_X^\bullet]$ then $[v_X^\bullet] = [v_M^/]$, and from the Schottky equilibrium

$$\mathrm{K_S} = [v_X^\bullet][v_M^/] \tag{4.2}$$

we obtain

$$[v_X^\bullet] = [v_M^/] = \mathrm{K_S^{1/2}} \tag{4.3}$$

If, on the other hand, $[Mf_M^\bullet] >> [v_X^\bullet]$ then

$$[v_M^/] = [Mf_M^\bullet] \equiv \text{constant} \tag{4.4}$$

,i.e., the metal vacancy concentration is fixed by the concentration of the foreign cation. The concentration of the minority species, $[v_X^\bullet]$, can be found by insertion into $\mathrm{K_S}$:

$$[v_X^\bullet] = \frac{\mathrm{K_S}}{[Mf_M^\bullet]} \tag{4.5}$$

We observe that this concentration decreases with increasing concentration of dopant, as expected also from direct inspection of the electroneutrality condition.

The two regions and their transition is illustrated in the figure below. We often call the region dominated by intrinsic disorder for intrinsic, while the one dominated by foreign species such as dopants or impurities is called extrinsic.

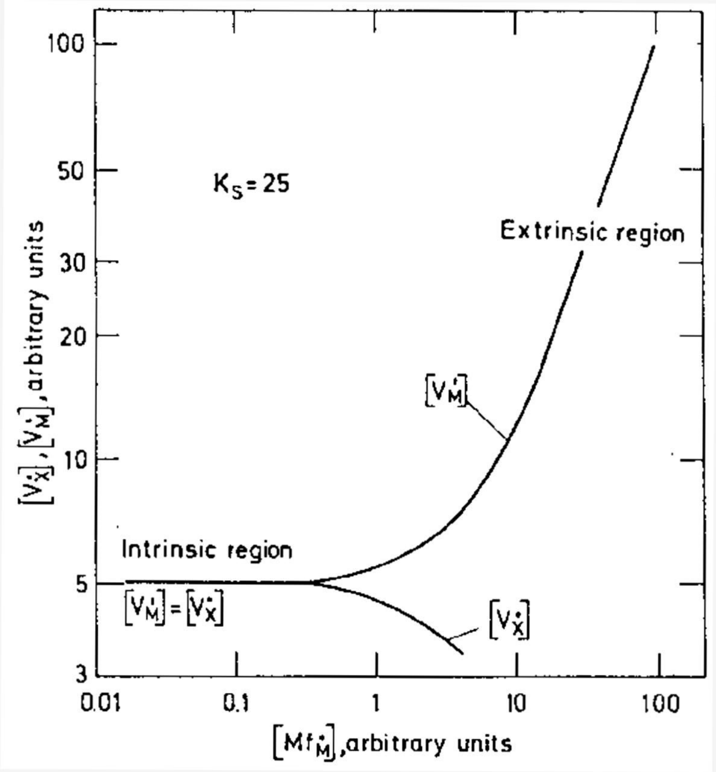

Figure 4-1. Effect of higher valent cation impurities on the defect concentrations in MX predominantly containing Schottky defect pairs, with K_S = 25.

Frenkel equilibria

The effects of higher valent cation impurities on Frenkel equilibria are analogous to those described for the Schottky equilibria. When divalent $Mf_M^\bullet$ impurities/dopants are dissolved in MX with predominating cation Frenkel defect pairs ($v_M^/$ and $M_i^\bullet$) the electroneutrality condition is given by

$$[Mf_M^\bullet]+[M_i^\bullet]=[v_M^/] \tag{4.6}$$

The addition of $Mf_M^{\bullet}$ decreases $[M_i^{\bullet}]$ and increases $[v_M^{/}]$ in a manner analogous to that described for the Schottky-dominated material.

Oxygen-deficient oxides doped with lower valent cations

A situation which is frequently encountered in applications of several oxide systems is that doubly charged oxygen vacancies are the predominating native point defects and that the cation impurities/dopants have a valence lower than that of the parent cation (impurities/dopants with negative effective charge are also called acceptors). An example is zirconia (ZrO_2) doped with lower valent cations such as for instance yttrium ($Y_{Zr}^{/}$) and calcium ($Ca_{Zr}^{//}$). For the sake of illustration let us assume that the impurity/dopant ions have one effective negative charge, $Mf_M^{/}$. The actual valence of the host cation will not matter. The defect equilibrium between vacancies and electrons is as before given by

$$O_O^x = v_O^{\bullet\bullet} + 2e^{/} + \tfrac{1}{2}O_2(g) \qquad (4.7)$$

$$K_{vO}^{/} = \frac{[v_O^{\bullet\bullet}]}{[O_O^x]} \mathrm{n}^2 p_{O_2}^{1/2} \qquad (4.8)$$

The electroneutrality condition will be given by

$$2[v_O^{\bullet\bullet}] = [Mf_M^{/}] + n \qquad (4.9)$$

For such a system two limiting conditions can be considered:

If $2[v_O^{\bullet\bullet}] \approx n >> [Mf_M^{/}]$, the foreign cations do not affect the native defect equilibrium, and the electron and oxygen vacancy concentrations are given by their own equilibrium, and they are proportional to $p_{O_2}^{-1/6}$ as we have shown in the preceding chapter. This will occur at relatively low oxygen activities, where these concentrations are relatively large.

If $2[v_O^{\bullet\bullet}] \approx [Mf_M^{/}] >> n$, the oxygen vacancy concentration is determined and fixed by the dopant content (extrinsic region). The concentration of minority electrons, n, is in this case given by

$$n = (2K_{vO}^{/})^{1/2}[Mf_M^{/}]^{-1/2} p_{O_2}^{-1/4} \qquad (4.10)$$

and thus decreases with a different dependency on p_{O_2} than in the former case. As the concentration of electrons and electron holes are related through the equilibrium K_i = np, the electron hole concentration in this extrinsic region correspondingly increases with increasing oxygen activity. Depending on the impurity content, oxygen activity and temperature, p may become larger than n.

The variations in $[v_O^{\bullet\bullet}]$, n and p as a function of the partial pressure of oxygen, are schematically illustrated for such a case in the figure below. In a situation like this the oxide will typically be an n-type electronic conductor at low oxygen partial pressure and an essentially ionic conductor at higher oxygen partial pressure. Zirconia doped with a lower valent cation such as Y, Ca, a.o., confers largely with the situation depicted in the figure.

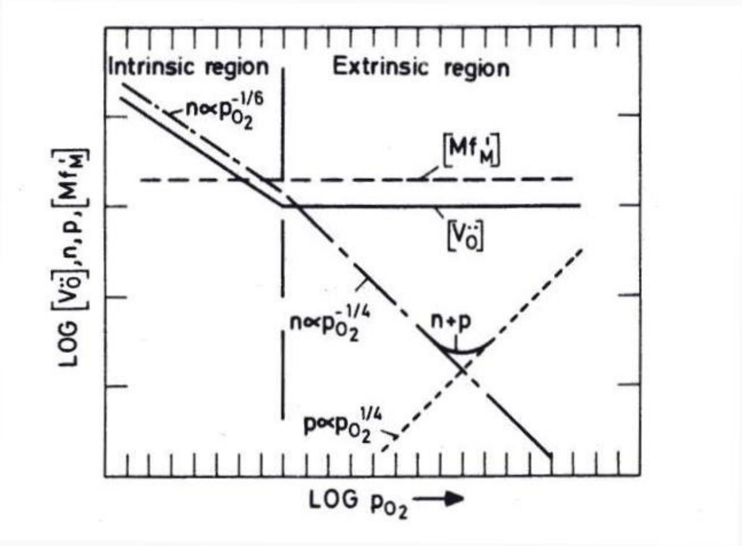

Figure 4-2. Brouwer plot of the concentrations of defects as a function of oxygen partial pressure in an oxygen deficient oxide predominantly containing doubly charged oxygen vacancies, showing the effects of a constant concentration of lower valent cation dopants, $[Mf_M^{/}]$.

Oxygen deficient oxides doped with higher valent cations

Addition of higher valent cations which dissolve substitutionally in the oxide will in the extrinsic region result in an increase in the concentration of electrons and a decrease in the concentration of oxygen vacancies. This may be shown by setting up the appropriate electroneutrality condition combined with the defect equilibria in the same manner as shown above.

Oxides with excess metal

Nonstoichiometric oxides with excess metal in the form of predominating interstitial metal ions exhibit the same qualitative effects of aliovalent foreign cations. By additions of lower valent cations, the concentration of interstitial cations will in the extrinsic region be determined by the concentration of the impurities/dopants. Addition of higher valent cations will bring about an increase in the electron concentration and a decrease in the concentration of metal interstitials.

Metal-deficient oxides doped with lower valent cations

When a metal-deficient oxide MO with singly charged metal vacancies as the prevalent native point defects is doped with singly valent foreign cations, $Mf_M^{/}$, the electroneutrality condition is given by

$$[v_M^{/}]+[Mf_M^{/}]=p \qquad (4.11)$$

The effects of the foreign cations are evaluated by combining this with the equations for the defect equilibria for the singly charged metal ion vacancies. As a result we find that the lower valent cations will increase p and decrease $[v_M^{/}]$. The variations in p and $[v_M^{/}]$ as a function of the partial pressure of oxygen in such a case are illustrated schematically in the Brouwer diagram below. At low oxygen activities interstitial metal ions or oxygen vacancies may become important, depending on the equilibrium constants for the Frenkel and Schottky defect equilibria. In the figure it is assumed that singly charged metal interstitials become the important native point defects at low oxygen activities.

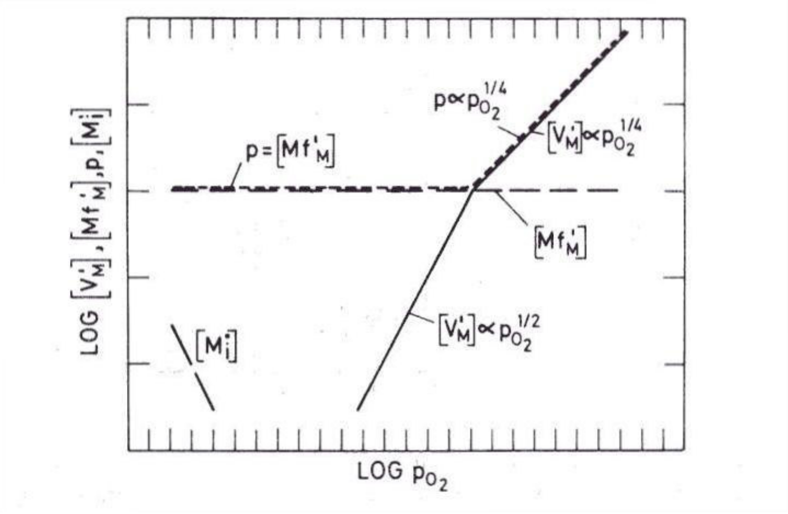

Figure 4-3. Effect of a constant concentration of foreign lower valent cations on the concentration of point and electronic defects as a function of oxygen pressure in a metal deficient oxide predominantly containing singly charged cation vacancies.

Metal deficient oxides doped with higher valent cations

If the foreign cation impurities/dopants are higher valent, e.g. $Mf_M^{\bullet}$, the electroneutrality condition of an oxide MO with predominantly singly charged metal vacancies becomes

$$[v_M^{/}] = [Mf_M^{\bullet}] + p \qquad (4.12)$$

In this case $[v_M^{/}]$ will increase and p will decrease with increasing concentration of $Mf_M^{\bullet}$. The variations in p and $[v_M^{/}]$ as a function of the oxygen partial pressure is illustrated schematically in the following figure, and the variation in the concentration of electrons (n) as minority defects is also included.

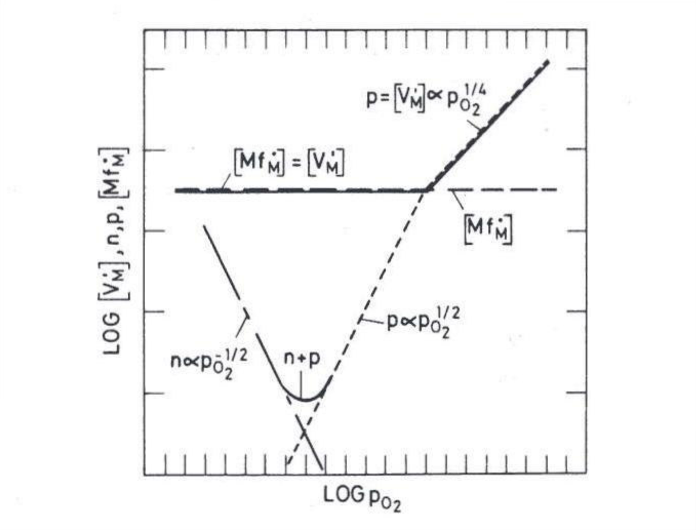

Figure 4-4. Brouwer diagram of the effects of a higher valent cation impurity/dopant on the concentration of point and electronic defects as a function of oxygen pressure in a metal deficient oxide predominantly containing singly charged metal vacancies.

Doping of oxides which may have regions with both oxygen and metal deficit

In the previous examples of the effects of lower valent oxides we have separately considered oxides which are either oxygen or metal deficient. We have seen that the doping is compensated by increased concentrations of either electronic defects or point defects, depending on the type of non-stoichiometry originally present. In principle it is straightforward to combine these treatments and consider the effects of aliovalent foreign atoms over the entire range of stoichiometry, from the oxygen-deficient to the metal-deficient side. Depending on oxygen partial pressure, the doping will be compensated by an electronic defect or a point defect of the opposite effective charge of the dopant. As we shall see it may be necessary to consider both possibilities when we deal with oxides near stoichiometry.

Let us therefore start with an example oxide M_2O_3 dominated by intrinsic electronic disorder when stoichiometric, and oxygen vacancies and metal vacancies giving oxygen and metal deficiency at, respectively, low and high oxygen activities. We acceptor-dope this oxide with the lower-valent Ml^{2+} cations, such that the electroneutrality condition reads:

$$\mathrm{n} + [Ml_M^{/}] + 3[v_M^{///}] = p + 2[v_O^{\bullet\bullet}] \tag{4.13}$$

We assume that the amount of MlO added is well below the solubility so that $[Ml_M^{/}]$ = constant. Fig. 4.5 shows a Brouwer diagram of the defect situation as a function of oxygen partial pressure when the level of aliovalent dopant is higher than the level of intrinsic disorder. At the lowest p_{O_2} the oxide is oxygen-deficient and oxygen vacancies and electrons predominate. As these defects decrease with increasing p_{O_2} we hit the level of the acceptor dopant. From here on the acceptor will be compensated by a constant concentration of the positive defect, i.e., the oxygen vacancies. This is analogous to the situation we obtained when we earlier considered this case in particular.

In this situation the concentration of defect electrons decreases and that of holes increases with increasing p□. The two terms cross at $n = p = K_i^{1/2}$, and assuming that K_i is independent of the doping, this is necessarily the same level as that of the intrinsic electronic disorder in the pure, undoped material. At this point also the doped material may be said to be stoichiometric; it contains oxygen vacancies exactly matching the presence of the lower-valent cation (acceptor): If we consider the doped material to consist of M_2O_3 and MlO, both these constituents are stoichiometric, and the additional presence of reduced and oxidised states (n and p) is effectively zero at this point. The doped oxide can be said to be stoichiometric with respect to the valence of its constituents.

By further increasing p_{O_2} the concentration of holes will eventually become dominating, taking over the dominance from the oxygen vacancies in terms of compensating the charge of the acceptors. We hereby create higher oxidation states (holes) and fill up the oxygen sublattice (removing oxygen vacancies). During the increase in p_{O_2} the concentration of metal vacancies increases and eventually balances the concentration of oxygen vacancies. This level is given by the Schottky equilibrium constant (K_S) and represents another stoichiometric point, this time with respect to the crystal structure of the host oxide (M_2O_3).

The situation encountered in this region is analogous to that depicted in Fig. 4.3, and with further increase in p_{O_2} the charge due to the increasing concentration of metal vacancies will become more important than that of the acceptor level, and the oxide enters into a situation of dominance by metal deficiency.

Fig. 4.5. is constructed by combination of limiting cases of the electroneutrality condition (Eq. 4.13) and defect equilibria in a similar manner as we have practised earlier.

It turns out by inspection of several cases that acceptor doping always leads to a region at relatively low p_{O_2} where we have compensation by a positively charged point defect and a region at relatively high p_{O_2} where we have compensation by electron holes. This is true independent of type of intrinsic disorder. However, the relative magnitudes of the equilibrium constants for intrinsic disorder (e.g. K_i vs K_S) determine the relative width of dominance for the two regions.

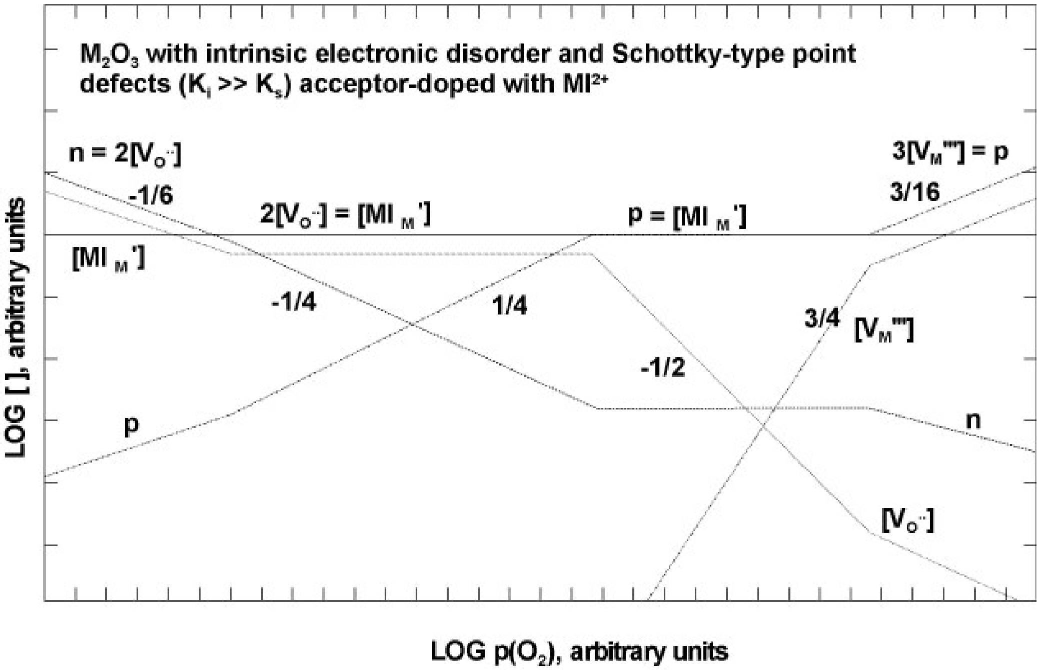

Figure 4-5. Brouwer diagram of the effects of lower valent dopant on the defect structure of $M_2O_{3\pm\delta}$.

In a manner equivalent to the above case we may construct a diagram showing the situation in the same host oxide M_2O_3 now donor-doped with a higher valent cation Mh^{4+} see Fig. 4.6. In this case and in all cases of donor-doping, there will be a region of charge compensating the dominating donors with defect electrons at relatively low p_{O_2} and with a negatively charged point defect at higher p_{O_2}

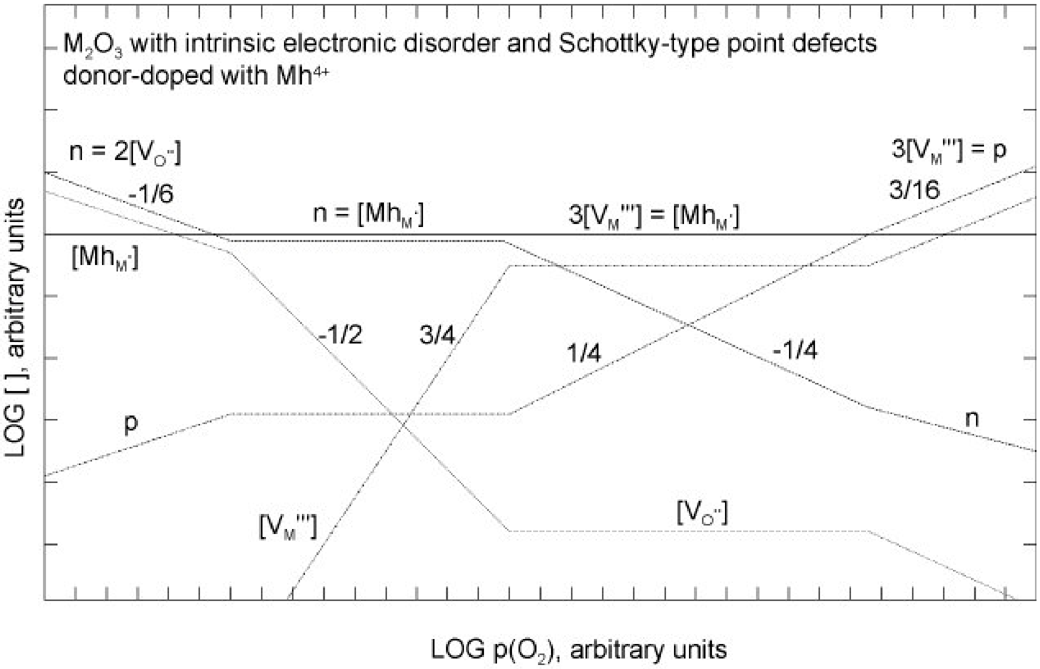

Figure 4-6. Brouwer diagram of the effects of higher valent dopant on the defect structure of $M_2O_{3\pm\delta}$.

We have here concentrated on changes in charge compensation as a function of oxygen activity, but the same may take place as a function of temperature and concentration of dopant (at constant p_{O_2}).

Aliovalent doping affects defects with higher number of charges more than those with fewer. Thus, with increasing doping, doubly and higher charged point defects may become dominating at the expense of the singly charged electronic defects, but not vice versa.

Variable concentration of impurity/dopant

In many cases an impurity or dopant may be present in amounts larger than the solubility limit. The excess of the oxide of the impurity or dopant is then presumably present in precipitates, often found at grain boundaries or surfaces of the parent oxide. When one considers the effects of impurities/dopants on defect equilibria in such cases, it is important to recognise that the solubility of the impurity or dopant will not only vary with temperature but in many cases also with the oxygen pressure. Let us illustrate these aspects by respectively considering the effects of the lower valent oxide MlO and the higher valent oxide MhO_2 on M_2O_3.

Doping M_2O_3 with excess MlO

When Ml dissolves substitutionally in M_2O_3, the dissolved Ml^{2+} has one negative effective charge, $Ml_M^/$. Let us further assume that these ions constitute the important negative effective charges in the oxide. These must then be compensated by positive defects, and let us assume that these are electron holes and doubly charged oxygen vacancies, depending on the properties of the oxide, the temperature, and the oxygen activity.

If the compensating defects are electron holes when MlO dissolves in M_2O_3, the defect equation for the dissolution may be written

$$2\mathrm{MlO(s)} + \tfrac{1}{2}\mathrm{O_2(g)} = 2Ml_M^/ + 2h^\bullet + 3O_O^x \qquad (4.14)$$

By taking the activity of the excess MlO(s) phase as well as that of oxygen ions as unity, the equilibrium expression can be written

$$\mathrm{K_p} = \frac{[Ml_M^/]^2\mathrm{p}^2[O_O^x]^3}{\mathrm{a}^2_{\mathrm{MlO(s)}}\mathrm{p}^{1/2}_{\mathrm{O_2}}} = [Ml_M^/]^2\mathrm{p}^2\mathrm{p}^{-1/2}_{\mathrm{O_2}} \qquad (4.15)$$

The important difference from the previous treatments is that now we treat this dissolution as an equilibrium. When the compensating defects are electron holes, the simplified electroneutrality condition can be expressed by $[Ml_M^/] = \mathrm{p}$, and by insertion into the equilibrium expression we obtain

$$[Ml_M^/] = \mathrm{p} = \mathrm{K_p^{1/4}}\mathrm{p}^{1/8}_{\mathrm{O_2}} \qquad (4.16)$$

Thus, the solubility, i.e., $[Ml_M^/]$, increases with p_{O_2} until all MlO is eventually dissolved. Thereafter the behaviour will become as with constant acceptor concentrations (all dissolved or frozen-in), described in Fig. 4.5.

When taking into account the defect equilibrium for the formation of oxygen vacancies and the intrinsic electronic equilibrium (np = K_i), the concentration of oxygen vacancies is under this situation seen to be proportional to $\mathrm{p}^{-1/4}_{\mathrm{O_2}}$. At lower oxygen activities the doubly charged oxygen vacancies will thus eventually become the charge compensating defects at the expense of electron holes. In this case it may be more relevant to describe the dissolution of MlO by the following defect reaction and equilibrium expression:

$$2\mathrm{MlO(s)} = 2Ml_M^/ + v_O^{\bullet\bullet} + 2O_O^x \qquad (4.17)$$

$$\mathrm{K}_{\mathrm{v}} = [Ml_M^{/}]^2[v_O^{\bullet\bullet}] \tag{4.18}$$

The electroneutrality condition and its combination with the equilibrium then yield:

$$[Ml_M^{/}] = 2[v_O^{\bullet\bullet}] = (2\mathrm{K}_{\mathrm{v}})^{1/3} \tag{4.19}$$

Thus under these conditions $[Ml_M^{/}]$ is independent of the oxygen activity at equilibrium with the excess MlO(s) phase. Thus, when the dopants are compensated by fully ionised point defects as in this case, the system would behave as if it had no excess or frozen-in dopant concentrations. However, the solubility would change with temperature (through the enthalpy involved in K_{v}) contrary to the cases without any excess or with frozen-in concentrations.

The overall defect structure situation is illustrated in the Brouwer diagram in Fig. 4.7; the important compensating defects are electron holes at comparatively high oxygen activities and oxygen vacancies at low oxygen activities. Both situations will eventually be overtaken by native non-stoichiometry at sufficiently low viz. high oxygen activities.

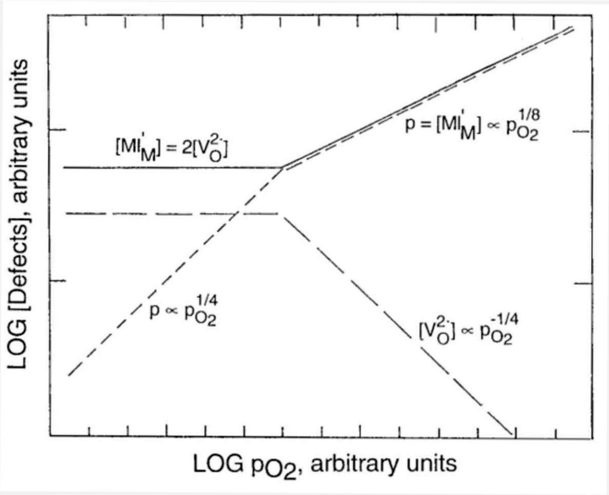

Figure 4-7. Brouwer plot of the effect of lower valent cations, $Ml_M^{/}$ *in* M_2O_3 *when the oxide MlO is present in amounts exceeding the solubility limit.*

Doping M_2O_3 with excess of MhO_2

Let us then dope M_2O_3 with an excess of MhO_2. The substitutionally dissolved Mh^{4+} ions have one positive effective charge, $Mh_M^{\bullet}$. They are compensated by defect electrons or point defects with negative effective charge, such as triply charged metal vacancies; $v_M^{///}$.

If the compensating defects are electrons, the reaction equation for the dissolution and the corresponding defect equilibrium are (under certain assumptions, as before) given by

$$2\mathrm{MhO_2(s)} = 2Mh_M^{\bullet} + 2e^{/} + 3O_O^x + \tfrac{1}{2}\mathrm{O_2(g)} \tag{4.20}$$

$$\mathrm{K_e} = \frac{[Mh_M^{\bullet}]^2\mathrm{n}^2[O_O^x]^3\mathrm{p}_{\mathrm{O_2}}^{1/2}}{\mathrm{a}_{\mathrm{MIO_2(s)}}} = [Mh_M^{\bullet}]^2\mathrm{n}^2\mathrm{p}_{\mathrm{O_2}}^{1/2} \tag{4.21}$$

Thus, when MhO_2 is dissolved by compensation of electrons, the solubility of MhO_2 increases with decreasing partial pressure of oxygen. If donor dopants and electrons dominate, the electroneutrality condition and resulting $\mathrm{p}_{\mathrm{O_2}}$ dependency become $\mathrm{n} = [Mh_M^{\bullet}] \propto \mathrm{p}_{\mathrm{O_2}}^{-1/8}$. Under these conditions it may be shown that $[v_M^{///}] \propto \mathrm{p}_{\mathrm{O_2}}^{3/8}$.

When the compensating defects at higher $\mathrm{p}_{\mathrm{O_2}}$ become $v_M^{///}$ the reaction equation for the dissolution of MhO_2 can be written

$$3\mathrm{MhO_2(s)} = 3Mh_M^{\bullet} + v_M^{///} + 6O_O^x \tag{4.22}$$

Thus, when the dissolution of MhO_2 is compensated by the formation of point defects the solubility of MhO_2 is independent of the oxygen activity. Under these conditions it may be shown that the concentration of electrons decreases with increasing oxygen activity and is proportional to $\mathrm{p}_{\mathrm{O_2}}^{-1/4}$. The overall defect structure situation is illustrated by the Brouwer diagram below.

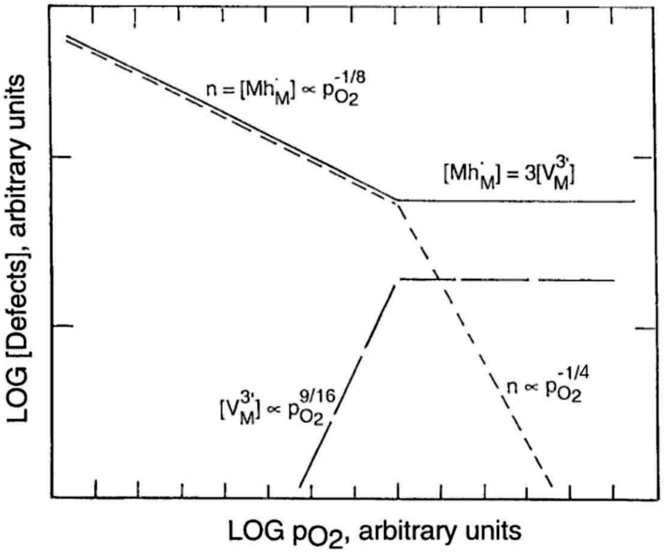

Figure 4-8. Brouwer plot of the effect of the dopant MhO_2 in M_2O_3 in amounts exceeding the solubility limit.

All in all, as a general rule the solubility of lower or higher valent oxides in a parent oxide will be dependent on the oxygen activity when the compensating defects are electronic defects, while they will be independent of the oxygen activity when the compensating defects are fully ionised point defects. It may also be noted that the temperature dependencies of the solubility will also be different in the two cases.

Partially ionised point defects act as combinations of point and electronic defects and generally gives oxygen activity dependent solubilities of aliovalent dopants.

In real cases it may be difficult to reach equilibrium if the diffusion of the cations to and from defect sinks (dislocations, grain boundaries a.o.) is slow. Another point worth mentioning is that the excess second phase may well be a compound of the parent and dopant oxide. For instance, MgO is normally present in excess of the solubility in so-called high-purity alumina (Al_2O_3) and the second phase in that case is spinel $MgAl_2O_4$. However, this does not alter the oxygen activity dependencies; the ternary compound serves as a source/sink for the dopant, with constant activity, and the defect chemical treatment remains essentially the same.

Hydrogen defects in metal oxides

When a metal oxide is equilibrated in gas mixtures with hydrogen-containing gases, e.g. H_2O, hydrogen will dissolve in the metal oxide. The extent of the dissolution of hydrogen will depend on the defect structure of the oxide and the ambient oxygen and hydrogen activities. In studies of the dissolved hydrogen it is then necessary to control both the oxygen and hydrogen (or water vapour) activities.

Hydrogen may in principle dissolve in the form of different species: as neutral atoms (H), hydride ions (H^-), and protons (H^+). In oxides the commonly observed dissolved hydrogen species are the protons, often termed interstitial protons, $H_i^\bullet$, but in reality always associated with oxygen ions as hydroxide groups. If that oxygen sits on a normal lattice site we thus get substitutional hydroxide defects, $OH_O^\bullet$.

A proper treatment of the dissolution of water and defect chemistry of protons in oxides was first given by Stotz and Wagner (1967). Later on, Norby (1987) and others have given extended treatments of hydrogen defect equilibria, and Kreuer (1996) has reviewed protons in materials in general.

The dissolution of protons from hydrogen gas (e.g. in a gas mixture of H_2+H_2O) may in these terms be written

$$H_2(g) + 2O_O^x = 2OH_O^\bullet + 2e^/ \tag{4.23}$$

Assuming that water and hydrogen are in equilibrium through $H_2 + 1/2\ O_2 = H_2O$, the defect reaction may equally well be written with water as the source of the protons:

$$H_2O(g) + 2O_O^x = 2OH_O^\bullet + 2e^/ + \tfrac{1}{2}O_2(g) \tag{4.24}$$

and the corresponding defect equilibrium by

$$K = \frac{[OH_O^\bullet]^2 n^2 p_{O_2}^{1/2}}{[O_O^x]^2 p_{H_2O}} \tag{4.25}$$

Thus the concentration of protons in metal oxides is dependent on the partial pressures of both the ambient oxygen and water vapour as well as the concentration of electronic defects.

It may be useful to remind ourselves that water in this context is an oxide present as a second phase in excess and, if we wish, at constant activity (controlled partial pressure). It is thus analogous to the case of an excess of metal oxide. The proton dissolves interstitially forming a positive defect ($H_i^\bullet$ or $OH_O^\bullet$) and thus behaves like a donor: When they are the dominating positive defects in an otherwise undoped oxide they may be compensated by electrons or negative point defects. In analogy with the higher-valent oxide dopant the first case would lead to a proton concentration dependent on the oxygen activity while it in the latter case would be independent.

As an additional possibility protons may also be compensated by a lower-valent dopant (acceptor). This may, in turn, be present at constant concentration (below the solubility or frozen-in) or itself in equilibrium with its own second phase. In the following examples we will consider some of these relations.

Effect of water vapour on oxygen-deficient M_2O_3.

Let us first examine the effects of water vapour on the properties of an undoped, oxygen-deficient oxide. We will use M_2O_3 in the example, but most of the treatment applies to any oxide. The predominant defects are electrons and oxygen vacancies and the electroneutrality condition in dry environments is then (from the preceding chapter)

$$n = 2[v_O^{\bullet\bullet}] = (2K_{vO}^{/})^{1/3} p_{O_2}^{-1/6} \qquad (4.26)$$

Let us keep the partial pressure of oxygen constant and vary the partial pressure of the water vapour. Protons are then dissolved in the oxide following Eq. 4.24. However, the concentration of electrons is given by Eq. 4.26 and is independent of p_{H_2O}. Insertion into Eq. 4.25 gives that the proton concentration is proportional to $p_{H_2O}^{1/2}$ (at constant p_{O_2}). This is illustrated in the Brouwer diagram in the figure below.

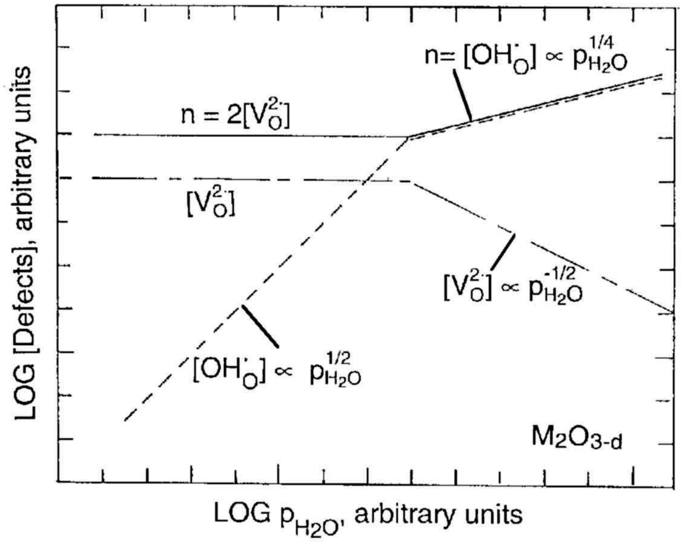

Figure 4-9. Brouwer plot of effects of water vapour on defect concentrations in oxygen deficient M_2O_{3-d}.

At sufficiently high partial pressure of the water vapour, the dissolved protons become the predominating point defects with positive effective charge and the electroneutrality condition becomes

$$n = [OH_O^\bullet] \tag{4.27}$$

By combination with Eq. 4.25 we obtain that the concentrations of protons and electrons are then proportional to $p_{H_2O}^{1/4}$:

$$n = [OH_O^\bullet] = K^{1/4} p_{H_2O}^{1/4} p_{O_2}^{-1/8} \tag{4.28}$$

This situation is part of the figure above. The figure also shows that the concentration of oxygen vacancies decreases with the water vapour pressure, illustrating that in general all charged native minority defects becoming dependent on water vapour pressure when protons are dominating defects.

Effect of water vapour on acceptor-doped M_2O_3

Let us next consider the effect of water vapour in the oxide M_2O_3 doped with the oxide MlO. Let us further assume that the concentration of the dopant ion $[Ml_M^/]$ is sufficiently large that this is the predominant defect with negative effective charge but that the amount of dissolved MlO is smaller than the solubility limit of MlO in M_2O_3.

In dry atmospheres the dopant ions are assumed to be predominantly compensated by the formation of doubly charged oxygen vacancies ($2[v_O^{\bullet\bullet}] = [Ml_M^/] = \text{constant}$) and under these conditions the concentration of electrons is proportional to $p_{O_2}^{-1/4}$ and the electron holes to $p_{O_2}^{+1/4}$, as derived earlier for this case.

Let us then consider the effect of water vapour on the defect concentrations. For the sake of simplicity let us keep the partial pressure of oxygen constant while the partial pressure of water vapour is increased. Protons then dissolve in M_2O_3 following Eq. 4.24, and from Eq. 4.25 it can be derived that the concentration of protons dissolved as hydroxide defects, $[OH_O^\bullet]$, is proportional to $p_{H_2O}^{1/2}$. This is illustrated in the following figure.

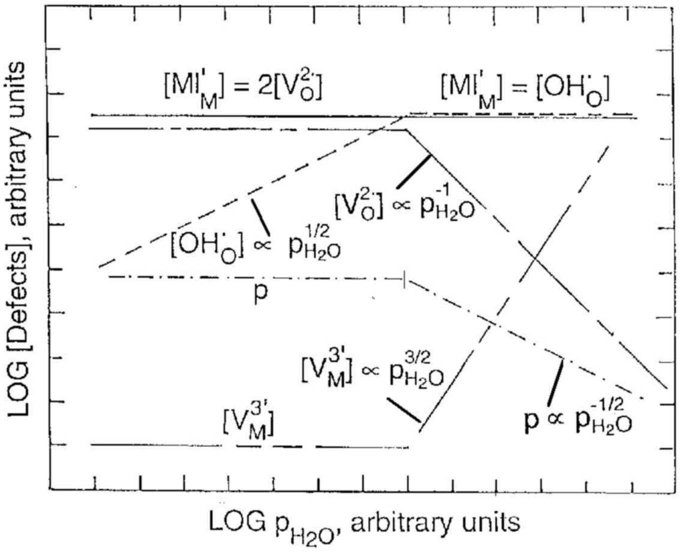

Figure 4-10. Brouwer plot of the effect of water vapour (at constant oxygen pressure) on defect concentrations in acceptor-doped, oxygen deficient M_2O_3.

Under these conditions the electroneutrality includes the concentration of dissolved protons:

$$2[v_O^{\bullet\bullet}] + [OH_O^{\bullet}] = [Ml_M^{/}] = \text{constant} \tag{4.29}$$

When we plot the concentration of the defects as a function of the partial pressure of water vapour (at constant oxygen pressure) as shown in the Brouwer diagram in Fig.4.10, it is seen that at sufficiently high water vapour pressures the dissolved protons become the important positively charged point defects and the electroneutrality condition is then approximated by

$$[OH_O^{\bullet}] = [Ml_M^{/}] = \text{constant} \tag{4.30}$$

Under these conditions the concentration of oxygen vacancies decreases with increasing partial pressure of water vapour; combining the equilibrium for formation of oxygen vacancies (Eq. 4.8) and the equlibrium for dissolution of protons, Eq. 4.25, and the dominating electroneutrality, Eq. 4.30, it is found that the concentration of oxygen vacancies is proportional to $p_{H_2O}^{-1}$. The concentration of electron holes is correspondingly proportional to $p_{H_2O}^{-1/2}$. These aspects are also illustrated in Fig. 4.10.

It is also of interest to consider any effects of the partial pressure of water vapour on the concentration of other minority defects. If so, the ambient partial pressure of water vapour may affect the sintering and creep of metal oxides at high temperatures. Let us continue to examine the acceptor-doped M_2O_3. For the sake of illustration let us assume that the important defects in the metal sub-lattice are metal vacancies and that these are triply charged, $v_M^{///}$. The concentration of oxygen vacancies and metal vacancies in M_2O_3 may be related through the Schottky equilibrium.

In the region where the electroneutrality condition is given by $2[v_O^{\bullet\bullet}] = [Ml_M^{/}] = \text{constant}$, the concentration of the metal vacancies is also independent of the partial pressure of water vapour, cf. Fig.4.10. However, in the region where the electroneutrality condition is $[OH_O^{\bullet}] = [Ml_M^{/}] = \text{constant}$, the concentration of the metal vacancies increases sharply with increasing partial pressure of water vapour and is proportional to $p_{H_2O}^{3/2}$, cf. Fig. 4.10. Thus, under these latter conditions the properties which are controlled by the concentration of the minority defects, i.e. sintering, creep, may be greatly affected by variations in the partial pressure of the ambient water vapour pressure. The direction of the change will depend on the effective charge of the rate-limiting defect; whether it is oxygen or metal defects, and whether it is vacancies or interstitials.

When, as in the last example, protons are dissolved at the cost of oxygen vacancies, the reaction may be written

$$H_2O(g) + v_O^{\bullet\bullet} + 2O_O^x = 2OH_O^{\bullet} \tag{4.31}$$

i.e. water vapour dissolves to fill the oxygen vacancies and replace their positive charge with protons. Reaction 4.31 has been studied for a number of oxides, and it is found that it always involves a negative enthalpy change such that oxygen vacancies and protons are dominant at high and low temperatures, respectively. For some oxides, the enthalpy is sufficiently negative that protons remain the dominant defects up above 1000°C in practically dry atmospheres and to even higher temperatures under wet conditions. Other oxides have only small negative enthalpies and have dominant protons up to only moderately high temperatures, sometimes sufficiently low to kinetically prevent the reaction with water (Eq. 4.31) to reach equilibrium

The relative dominance of protons vs oxygen vacancies also varies with the acceptor doping level: The higher charge of the vacancy relative to the proton makes the former relatively more dominant at higher acceptor levels, while protons are relatively more dominant at moderate acceptor levels.

Let us also briefly recollect the variable solubility of aliovalent dopants treated previously in this chapter. While the acceptors are compensated by oxygen vacancies, their solubility does not vary with anything else than the temperature, as shown before. However, the dissolution of acceptors by simultaneous dissolution of protons,

$$2MlO(s) + H_2O(g) = 2Ml_M^{/} + 2OH_O^{\bullet} + O_O^x \tag{4.32}$$

will, when in equilibrium (MlO(s) present and diffusion enabled) give solubility of acceptors which increases with the water vapour pressure:

$$[\mathrm{OH}_O^\bullet] = [\mathrm{Ml}_M^/] \propto \mathrm{p}_{\mathrm{H_2}O}^{1/4} \tag{4.33}$$

Thus, water vapour may in some cases play a role in the synthesis of solid solutions containing acceptors.

By inspection of Fig. 4.10 it is evident that at sufficiently high water vapour levels, the concentration of metal vacancies will become significant and eventually dominant; the oxide attains a metal deficiency (or, in other cases, oxygen excess), compensated by protons. Such defect structures can be considered to be the first step towards the phase limit of the oxyhydroxide or hydroxide.

Association between impurities/dopants and other point defects

As described in Chapter 1 impurities/dopants and native point defects may associate to form complex defects. An important driving force may be the coulombic attraction between impurities and point defects having opposite effective charges, but also relaxation of the lattice around the associate may play a role.

As an illustration, it will be assumed that the predominating defects are doubly charged oxygen vacancies and that the cation dopants have one effective negative charge, e.g. $\mathrm{Ml}_M^/$. The associated complex between $\mathrm{v}_O^{\bullet\bullet}$ and $\mathrm{Ml}_M^/$ can be written $(\mathrm{v}_O Ml_M)^\bullet$, and compared to the relatively mobile oxygen vacancy, this defect can be considered as immobile as the dopant ions themselves.

The complex defect will be in equilibrium with the single defects according to the reaction

$$\mathrm{v}_O^{\bullet\bullet} + \mathrm{Ml}_\mathrm{M}^/ = (\mathrm{v}_O Ml_M)^\bullet \tag{4.34}$$

If all the species are randomly distributed, the equilibrium can be written

$$\mathrm{K_a} = \frac{[(\mathrm{v}_O\mathrm{Ml}_M)^\bullet]}{[\mathrm{v}_O^{\bullet\bullet}][\mathrm{Ml}_M^/]} \tag{4.35}$$

where K_a is the equilibrium constant for the association reaction. If the acceptor is relatively immobile, the associated oxygen vacancy can be considered trapped; it is no longer able to contribute in charge or diffusional transport of oxygen, unless an extra activation energy is supplied to break up the association.

Many other alternative examples of the effects of impurities/dopants on defect equilibria may be cited, including association between acceptors and protons, acceptors and holes, donors and point defects, and donors and electrons.

Summary

Impurities or dopants may dissolve interstitially or substitutionally, and the defects formed are often easily ionised. Depending on whether they accept or donate electrons they become acceptors and donors. In sufficient quantities these may dominate the defect structure of the oxide. Thus, aliovalent cations substituting host cations are often introduced in large quantities in oxides in order to give high concentrations of ionic defects (for solid electrolytes) or electronic defects (for electrodes etc.) Foreign species may be treated like other defects, but are often assumed to be present in a fixed concentration due to the limited availability of the species (full solubility) or because the concentration is frozen-in. However, we have shown that the concentration of foreign species under certain conditions may change as a function of temperature and for some defect structures as a function also of p_{O_2}. Protons dissolved from water vapour is a special case which often has a large influence on the properties of oxides. They form hydroxide groups with the oxygen ions of the oxide. These are effectively positively charged and through them the defect structure may become a function of the water vapour partial pressure.

Literature (protonic defects only)

Kreuer, K.-D. (1996) *Chem. Mater.* **8**, 640.

Norby, T. (1987) *Advances in Ceramics* **23**, 107.

Stotz, S. and Wagner, C. (1967) *Ber. Bunsenges. Phys. Chem.* **70**, 781.

Exercises

1. Write the doping reaction for the doping of the Schottky-dominated metal halide MX by the higher valent metal halide MhX_2. Find the analytical solution expressing the concentration of metal vacancies as a funtion of doping concentration. Sketch also the temperature-dependency of all the defects involved. (Hint: This is the initial illustrative case in the text – use the equilibria from that treatment. Remember that the doping reaction may

not be in equilibrium – the dopant concentration is invariable as a function of outer parameters.)

2. Sketch a Brouwer diagram (double logarithmic diagram of defect concentrations vs p_{O2}) for an oxygen deficient oxide doped with a substitutional higher-valent dopant.

3. Sketch a Brouwer diagram (double logarithmic diagram of defect concentrations vs p_{O2}) for an oxide $M_{1+x}O$ doped with a substitutional lower-valent dopant.

4. Sketch a Brouwer diagram (double logarithmic diagram of defect concentrations vs p_{O2}) for an oxide $M_{1+x}O$ doped with a substitutional higher-valent dopant.

5. Sketch another Brouwer diagram, this time the double logarithmic diagram of defect concentrations vs *dopant concentration*, for the oxide $M_{1+x}O$ doped with a substitutional lower-valent dopant.

6. An oxide M_2O_3 is dominated by intrinsic electronic disorder in the stoichiometric state, while Schottky disorder is the next largest contributor to defects. Draw a Brouwer diagram showing log defect concentrations as a function of the concentration of lower-valent acceptor dopants. Do the same for the case of higher-valent donor dopants. In both cases cover the range from insignificant doping levels to the level where all possible defect domains are exhausted.

7. An oxide M_2O_3 is dominated by Schottky disorder in the stoichiometric state, while intrinsic electronic disorder is the next largest contributor to defects. Draw a Brouwer diagram showing log defect concentrations as a function of the concentration of lower-valent acceptor dopants. Do the same for the case of higher-valent donor dopants. In both cases cover the range from insignificant doping levels to the level where all possible defect domains are exhausted.

8. An oxide M_2O_{3-d} is doped substitutionally with MlO, and dominated by oxygen vacancies and protons as compensating defects. Derive the full expression for the concentration of protons (and oxygen vacancies) as a function of a fixed dopant concentration and the equilibrium constant for the equilibrium between oxygen vacanciens and protons. Make assumptions and simplifications as necessary, but assume that both vacancies and protons are significant compensating defects.

9. Derive equilibria, electroneutrality, and behaviour for an oxide M_2O_3 dominated by metal vacancies and protons dissolved from surrounding water vapour.

10. Choose an acceptor-doped oxide compensated by oxygen vacancies. Write the reaction equation for the association reaction and the equilibrium constant. Derive the temperature-dependency of the concentration of free, unassociated vacancies. Assume that the entropy change of the reaction is zero, and that the enthalpy change is 50 kJ/mol. What is the fraction of free vacancies at 723 K?

5. Diffusion

Introduction

Numerous chemical reactions or micro-structural changes in solids take place through solid state diffusion, i.e. the movement and transport of atoms in solid phases. In crystalline solids, the diffusion takes place because of the presence of defects. Point defects, e.g. vacancies and interstitial ions, are responsible for lattice diffusion. Diffusion also takes place along line and surface defects which include grain boundaries, dislocations, inner and outer surfaces, etc. As diffusion along linear, planar and surface defects is generally faster than in the lattice, they are also termed high diffusivity or easy diffusion paths. Another frequently used term is short circuit diffusion.

The relative contribution of the different types of diffusion in oxides and other inorganic compounds are functions of the temperature, partial pressures or activities of the constituents of the compounds, the microstructure, grain size, porosity etc. Grain boundary and dislocation diffusion generally have smaller activation energies than lattice diffusion and as a result become increasingly important the lower the temperature in solids with a given microstructure.

In the literature on diffusion and diffusion-controlled reactions or processes one encounters many different terms that describe the diffusional behaviour under different experimental conditions: tracer and self-diffusion of atoms and ions, diffusion of defects, chemical diffusion, ambipolar diffusion, a.o. Many of these are used for treating diffusion in compounds, and in the following chapters these phenomena and terms will be described in more detail. Here we will start out with a few simple phenomenological descriptions, and for simplicity we look only at diffusion of neutral, independent particles.

Models of diffusion

It is a well known phenomenon that heat flows from hot to cold regions. Such a flow of heat in a one-dimensional temperature gradient is described by Fourier's law

$$j_q = -\kappa \frac{dT}{dx} \tag{5.1}$$

where j_q is the heat flux density, i.e., the flow of heat per unit area of the plane through which the heat traverses per second, $\frac{dT}{dx}$ is the temperature gradient, and κ (kappa) is the thermal conductivity. It may be noted that the minus sign reflects that the heat flows from high to low temperatures (downhill).

Fick's first law

The expression for the flow of particles from high to low concentrations is analogous to that for the flow of heat and is given by Fick's first law:

$$j_{particles} = -D\frac{dc}{dx} \tag{5.2}$$

Here, $j_{particles}$ is the particle flux density, $\frac{dc}{dx}$ the concentration gradient of the particles and D the diffusion coefficient. As in the equation for the heat flux, the minus sign reflects that the particles flow from high to low concentration of particles (downhill). This relation is named after A. Fick who first formulated this relation. The particle flux and gradient is illustrated schematically in Fig.5.1.

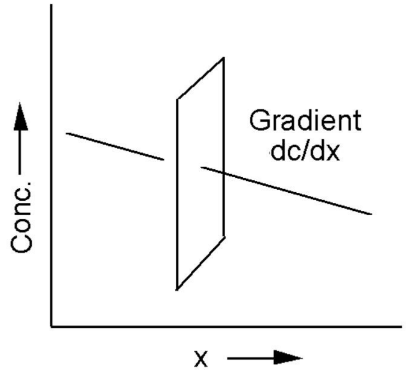

Figure 5-1. Schematic illustration of Fick's first law. The negative of the particle concentration gradient is the "driving force" of the diffusion.

$j_{particles}$ represents the number of particles (or moles) crossing a unit area (cm^2 or m^2) per unit time (seconds). If the concentration of particles is expressed in number of particles (or moles) per cm^3 and the distance x in cm, the diffusion coefficient has the dimension cm^{2}s^{-1}. In SI units the concentration is expressed in number per m^3 and the diffusion coefficient has the dimension m^{2}s^{-1}.

In Fick's first law in Eq. (5.2) the negative particle gradient, $-\frac{dc}{dx}$, may be considered to be an expression of the "driving force" for the particle flux (Eq.5.2). The larger the concentration gradient, the larger the particle flux. When Fick's first law is applied to uncharged (neutral) and independently diffusing particles, it is valid in the sense that the coefficient is a constant (independent of concentration and gradient). This may apply e.g. to dilute solutions of neutral defects in solids. Such defects may be vacancies, interstitials, and impurities in metals. In ionic solids they may comprise neutral interstitial impurities, homovalent substituents, or isotopic species. The applicability or inapplicability of Fick's first law will become clearer later in this chapter and in the forthcoming chapters.

Potential gradients as the driving force

More generally, the driving force for the diffusion constitutes the chemical potential gradient of the particles that diffuse (provided that no other forces act on the particles). Correspondingly, the driving force for the transport of electrical charges is the electrical potential gradient. In the following is given a brief derivation of Fick's first law using a potential gradient as the driving force, in detail in the case of a chemical potential gradient.

Let us consider the transport of particles of type "i" across a plane under a driving force ***F***. The particle flux density through a plane is given by the product of the volume concentration c_i of the particles at the plane and the average migration or drift velocity v_i of the particles

$$j_i \text{ (moles or particles/(cm}^2\text{sec))} = c_i \text{ (moles or particles/cm}^3) \cdot v_i \text{ (cm/sec)} \quad (5.3)$$

For uncorrelated movements the drift velocity v of a particle is proportional to the driving force F_i exerted on the particle:

$$v_i = B_i F_i \quad (5.4)$$

The proportionality factor B_i is termed the mobility ("Beweglichkeit") of the particles and is defined as the average drift velocity per unit driving force. It is often referred to as mechanical mobility to differentiate it from other types of mobility.

The driving force is, in turn, given by the negative value of the potential gradient normal to the cross-sectional area in the plane

$$F_i = -\frac{dP_i}{dx} \quad (5.5)$$

where P is a potential. The negative sign is, as above, due to the fact that the transport takes place from higher to lower values of P, see Fig.5.2.

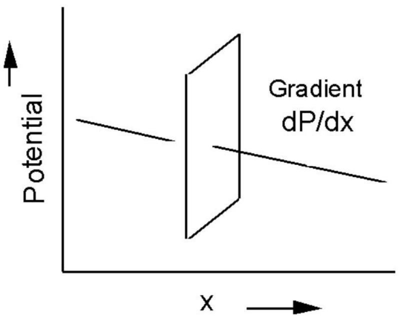

Figure 5-2. Schematic illustration of Fick's first law for transport in a potential gradient.

When one combines Eqs. 5.3 - 5.5, the particle flux becomes

$$j_i = c_i v_i = -c_i B_i \frac{dP_i}{dx} \tag{5.6}$$

Example of chemical potential gradient acting on neutral particles

If the particle moves in a chemical potential gradient, the potential P equals the chemical potential of particles of type i: $P_i = \mu_i$. Equation 5.6 then takes the form

$$j_i = c_i v_i = -c_i B_i \frac{d\mu_i}{dx} \tag{5.7}$$

The chemical potential μ_i is related to the chemical activity a_i of species i through

$$\mu_i = \mu_i^o + kT \ln a_i \tag{5.8}$$

If ideal conditions can be assumed, the activity can be expressed by the concentration divided by the concentration in the reference (standard) state:

$$a_i = \frac{c_i}{c_i^0} \tag{5.9}$$

and the chemical potential gradient is then given by

$$\frac{d\mu_i}{dx} = kT\frac{d\ln c_i}{dx} = \frac{kT}{c_i}\frac{dc_i}{dx} \qquad (5.10)$$

When one substitutes the expression for the chemical potential gradient in Eq.5.7, the particle flux becomes

$$j_i = -c_i B_i \frac{d\mu_i}{dx} = -B_i kT\frac{dc_i}{dx} \qquad (5.11)$$

B_ikT is termed the diffusion coefficient D_i of particles i:

$$D_i = B_ikT \qquad (5.12)$$

By combination with Eq.5.12, Eq.5.11 takes the form of Fick's first law as given in Eq.5.2:

$$j_i = -D_i\frac{dc_i}{dx} \qquad (5.2)$$

It is thus important to realise that Eq. 5.2 is fully valid only for ideal cases of diffusion of neutral particles in chemical potential gradients only. If the particles are charged, we need to take into account both the electrical potential gradent and requirements to the combination of fluxes to maintain a given total current and electroneutrality. These cases will be treated later on.

In the following, ideal conditions will be assumed and concentrations will be used for activities of atoms, ions and various types of defects. But it should be recalled that this always represents an approximation.

Simplified model for one-dimensional diffusion

Following Eq.5.2 the diffusion coefficient represents the proportionality constant between the particle flux and the concentration gradient. In order to describe the process of diffusion of particles or atoms in solids we need to realise that the diffusion represents the sum of a large number of particles or atoms that each makes a large number of jumps. We will therefore attempt to describe the diffusion coefficient in terms of the number of jumps per unit time (the jump frequency) and the distance that each particle or atom moves in each jump.

For this purpose let us consider a simplified one-dimensional model where particles jump between parallel planes separated by a distance s as illustrated in

Fig.5.3. The two neighbouring planes under consideration are termed plane 1 and 2. The number of particles per unit area in plane 1 and 2 is termed n_1 and n_2, respectively, and let us further assume that $n_1>n_2$. Consider further that the particles in plane 1 and 2 may jump from one plane to a neighbouring plane at a jump frequency Γ. The particles in plane 1 have an equal probability of jumping to plane 2 and to the neighbouring plane in the opposite direction. The total number of particles jumping out of a unit area of plane 1 per unit time is equal to the product of number of particles per unit area times the jump frequency: $n_1\Gamma$. As the particles may jump in opposite directions, the number of particles jumping from plane 1 to 2 is given by ½ $n_1\Gamma$. Correspondingly, the number of particles jumping from unit area of plane 2 to plane 1 is given by ½ $n_2\Gamma$. The difference in the jump rates is equal to the net flux density of particles:

$$j_{particles} = \tfrac{1}{2}(n_1 - n_2)\Gamma \qquad (5.13)$$

As all particles have the same jump frequency, there is a net flow of particles from plane 1 to 2 because there are more atoms per unit area in plane 1 than in plane 2.

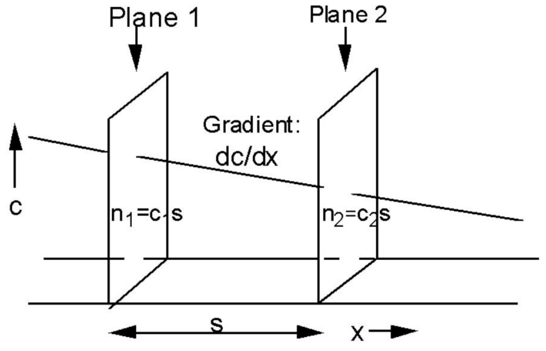

Figure 5-3. ***Schematic illustration of simplified model for one-dimensional diffusion***

The number of particles n_1 belonging to unit area of plane 1 is given by the volume concentration at plane 1, c_1, times the extension in the x direction, i.e., the plane separation s:

$$n_1 = c_1 s \quad \text{and} \quad n_2 = c_2 s \qquad (5.14)$$

By combining Eqs.5.13 and 5.14 the particle flux density becomes

$$j_{particles} = \tfrac{1}{2}(c_1 - c_2)s\Gamma \qquad (5.15)$$

The concentration gradient normal to the planes is termed $\frac{dc}{dx}$. The relation between c_1 and c_2 can then be expressed by seeing that a difference is equal to the gradient times the length;

$$c_2 - c_1 = s\frac{dc}{dx} \tag{5.16}$$

By insertion of this into Eq. 5.15 we obtain

$$j_{particles} = -\tfrac{1}{2}s^2\Gamma\frac{dc}{dx} \tag{5.17}$$

This expression applies to one-dimensional diffusion of particles, and the factor ½ reflects that only ½ of all jumps occur in the direction we consider as the flux direction.

If the diffusion can take place in the three orthogonal directions, only one third of the particles jump in one direction, and in the three-dimensional case the net flux in one direction is 1/3 of the flux of that when all atoms jump in one direction only:

$$j_{particles} = -\tfrac{1}{6}s^2\Gamma\frac{dc}{dx} \tag{5.18}$$

By comparing with Fick's first law, Eq. 5.2, it is seen that the diffusion coefficient D in the three-dimensional case is given by

$$D = \tfrac{1}{6}s^2\Gamma \tag{5.19}$$

If we consider a large number of jumps, n, which occurs during the time t, then

$$\Gamma = \frac{n}{t} \tag{5.20}$$

and inserting this in Eq.5.19, one obtains

$$ns^2 = 6Dt \tag{5.21}$$

The expression for the diffusion coefficient in Eqs.5.19 and 5.21 provides an interesting qualitative description of what is going on during the diffusion. By way of example, the diffusion coefficient for interstitial diffusion of oxygen atoms in niobium metal at 800 °C is approximately $D = 7 \cdot 10^{-8}$ cm^2s^{-1}. The jump distance can be assumed to be 1.65 Å ($1.65 \cdot 10^{-8}$ cm), and then from Eq. (5.19) the jump frequency Γ is about $1.54 \cdot 10^{9}$ s^{-1}. Thus, each oxygen atom makes a tremendously large number of jumps per second. But it should then also be recalled that the atoms vibrate with a (Debye) frequency of 10^{12}-10^{13} s^{-1}, and thus only a small fraction - about 1 in 10^3 or 10^4 - of the vibrations leads to a jump of the atoms. We shall consider this in more detail when the different atomistic mechanisms are treated below.

Although the number of jumps is tremendously large, the mean displacement of each atom is relatively small – most of the time it moves back and forth. In the diffusion process it is not possible to observe the individual jumps of the atoms, and it is necessary to find a relation between the individual atom jumps for large number of atoms and the diffusion phenomena which may be observed on a macroscopic scale. The problem is to find how far a large number of atoms will move from their original sites after having made a large number of jumps. Such relations may be derived statistically by means of the so-called random walk method.

Random diffusion

Let us consider that the jumps of the atoms are random, i.e. that the jumps of the atoms are independent of all the previous jumps and can occur in all directions. In that case the displacement of a diffusing atom from the starting point after n number of jumps, R_n, is given by the algebraic sum of the individual jump vectors:

$$\overrightarrow{R_n} = \overrightarrow{s_1} + \overrightarrow{s_2} + \ldots \overrightarrow{s_n} = \sum_{j=1}^{n} \overrightarrow{s_j} \tag{5.22}$$

If the individual jumps take place with equal probability in all directions and the individual jump distances are equal, this algebraic sum equals zero. This does not mean that the diffusing atom remains at its starting point after n jumps, but only that jumps in "positive " and "negative" directions are equally probable. In fact the total displacement may have any value between zero and ±ns.

In order to obtain a value for the magnitude (length) of the sum vector, one squares Eq. 5.22:

$$R_n^2 = \overrightarrow{R_n}^2 = \sum_{j=1}^{n} \overrightarrow{s_j}^2 + 2\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \overrightarrow{s_j}\overrightarrow{s_k} \qquad (5.23)$$

If, as above, the individual jump vectors are equal, i.e., $|s_1| = |s_2| \ldots |s_j| = s$, as, for instance, in crystals with cubic symmetry, and if they are random and uncorrelated, then the second term on the right hand side of the equality sign in Eq. 5.23 will approach zero for large numbers of jumps, as on an average s_j and s_k have an equal chance of being positive and negative. The first term on the righ hand side, on the other hand, is always non-zero and positive, and thus represents the average displacement length for a large number of jumps:

$$R_n^2 = \sum_{j=1}^{n} \overrightarrow{s_j}^2 = ns^2 \qquad (5.24)$$

The mean displacement is given by the square root of R_n^2 and is termed the root mean square displacement:

$$\overline{R_n} = \sqrt{R_n^2} = \sqrt{n}\, s \qquad (5.25)$$

From this it is seen that the mean displacement is proportional to the square root of the number of jumps times the individual jump distance.

By combining Eqs. 5.21 and 5.24 we may express the random diffusion in terms of the diffusion coefficient we dealt with in diffusion down a concentration gradient in the 3-dimensional cubic case:

$$R_n^2 = ns^2 = 6D_r t \qquad (5.26a)$$

$$R_n = \sqrt{6D_r t} \qquad (5.26b)$$

where t is the time during which the mean-square-displacement takes place. One may note that we have now started terming the diffusion coefficient D_r (r for random walk).

R_n is the radius of the sphere that a diffusing atom on average will distance itself by from the starting point after time t.

Let us now consider the displacement in a single dimension (e.g. the x-direction) as a result of this three-dimensional displacement in the cubic case. From simple geometry we have $R_n^2 = 3x^2$ where x^2 is the mean square displacement in each orthogonal direction. Thus, the mean diffusion length in one direction in a three-dimensional cubic crystal is:

$$x = \sqrt{2D_r t} \tag{5.27}$$

This length is thus shorter than the displacement radius, since displacements in the y and z directions are “wasted” for x dimension displacement.

As described above and using oxygen diffusion in niobium as an example, the oxygen atoms on an average exchange positions approximately $1.54 \cdot 10^9$ times per second at 800 °C. From the same considerations one may also estimate that an oxygen atom has randomly covered total jump distances of 25.4 cm and 914.7 m after 1 second and 1 hour, respectively. But what is the mean displacement? From Eq 5.27 one may estimate that the one-dimensional root-mean-square displacement after 1 hour only amounts to 0.022 cm. Thus, the mean displacement is very small and on an average the oxygen atoms spend most of their time jumping "back and forth".

What we did above was to consider random jumps, and then we related jumps, times and distances to a term we recognised from earlier, namely the diffusion coefficient, and we named it the random diffusion coefficient D_r. The diffusion coefficient was in turn something we recognised – while considering diffusion down a one-dimensional concentration gradient – from Fick’s first law. One should note, however, that random diffusion and the random diffusion coefficient can be considered and expressed and quantified also in the absence of a concentration gradient and also for charged particles.

As will be described below, diffusion is often measured by using tracer atoms and one then obtains values of the diffusion coefficient of the tracer atoms. Depending on the diffusion mechanism the tracer diffusion is in most cases not completely random, but is to some extent correlated with previous jumps. This will be further discussed later on.

Fick's second law

As described above Fick's 1st law assumes a fixed concentration gradient across the plane through which the flux of particles take place. But in numerous practical cases the concentration and concentration gradient changes with time. Such cases are covered by Fick's 2nd law. This is shown schematically in Fig.5.4 which illustrates the change in the concentration gradient through the solid. As shown in the figure let us consider a region within a solid, enclosed between planes separated by the distance dx. The net particle flux from the region of higher concentration into dx is j_1 and the net particle flux out of dx towards lower concentration is j_2. When j_1 is greater than j_2 the particle concentration in dx increases with time. This requires that the concentration gradient in plane 1 is larger than in plane 2, since Fick's first law applies in both planes (and at any position) at any time.

The change in concentration per unit time at any position is proportional to the gradient in flux at that position

$$\frac{\partial c}{\partial t} = -\frac{\partial j}{\partial x} \tag{5.28}$$

Although the concentration and concentration gradient change with time, Fick's 1. law is valid at any one time and position and thus

$$\frac{\partial c}{\partial t} = -\frac{\partial j}{\partial x} = \frac{\partial}{\partial x}(D\frac{\partial c}{\partial x}) \tag{5.29}$$

If D can be considered to be independent of concentration then

$$\frac{\partial c}{\partial t} = -\frac{\partial j}{\partial x} = D\frac{\partial^2 c}{\partial x^2} \tag{5.30}$$

The intuition of the qualitative arguments as well as the double derivative of the concentration in Eq.5.30 tell us that particles will flow from convex to concave regions in terms of distribution of particles.

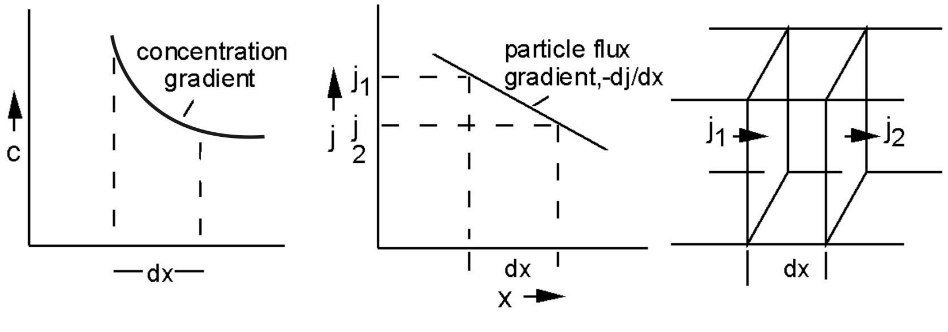

Figure 5-4. Schematic illustration of Fick's 2nd law; the concentration gradient changes with time.

Equations 5.29 and 5.30 are representations of Fick's 2nd law. It may be solved explicitly under certain boundary conditions that may be closely approximated experimentally (e.g. Crank (1975)). A couple of examples of this are given in the following.

Measurements of the diffusion coefficients by tracer techniques

The use of isotopes or tracers is a common means of studying diffusion. Tracer methods permit measurements of self-diffusion, that is, the diffusion of the crystal components in a crystal. Furthermore, they allow measurements of diffusion in homogenous materials, that is, without imposing chemical gradients (when one disregards the difference in atomic weight between atoms in the crystal and of the tracer).

A common technique is to deposit a very thin film of radioactive isotopes on a plane surface of a sample, and, after subsequent diffusion anneal, determine the activity of diffusion species as a function of distance from the plane surface. If the thickness of the sample is very much larger than the penetration depth of the tracers, the solid can be considered semi-infinite. Furthermore, if the diffusion is homogenous (e.g. taking place by lattice diffusion), the concentration of the diffusing tracers normal to the plane is through solution of Fick's second law with appropriate boundary conditions given by

$$c = \frac{c_o}{2(\pi D_t t)^{1/2}} \exp\left(-\frac{x^2}{4D_t t}\right) \qquad (5.31)$$

c is the activity (or concentration) of the tracer at a distance x from the surface, c_o is the activity originally present on the surface, and t is time of the diffusion anneal. D_t is the tracer diffusion coefficient. Following Eq.5.31 it is determined by plotting lnc vs x^2, in which case the resultant straight line has the slope - $\frac{1}{4D_t t}$.

Plots of c vs x and of lnc vs x^2 according to Eq.5.31 is illustrated in Fig.5.5a and b, respectively. At the point where the activity is half of the activity at the surface, then x = $\sqrt{2.77 D_t t}$. This distance corresponds approximately to the root-mean-square penetration distance (Eq. 5.26b).

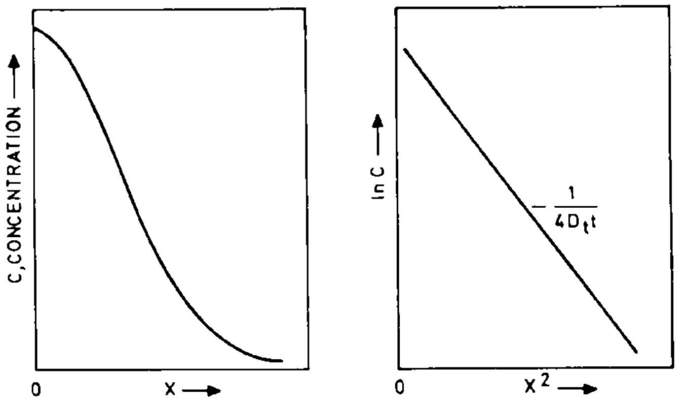

Figure 5-5. Graphical presentation of relationship between activity, c, and penetration distance, x, (Eq.5.31) for homogeneous diffusion of tracer initially deposited as a thin film on the surface the solid. a) c vs. x; b) lnc vs. x^2.

An alternative way of performing the experiment is to have a constant surface composition of the diffusing species. By solving Fick's second law under proper boundary conditions the penetration of the diffusing species is then described by the relationship

$$\frac{c-c_0}{c_s-c_0}=1-erf\frac{x}{2(D_t t)^{1/2}} \tag{5.32}$$

where c is the concentration of the diffusing species at penetration distance x at time t, c_S is the constant surface concentration, and c_O is the original concentration in the solid. D_S is the diffusion coefficient of the diffusing species. Figure 5.6 shows a graphical penetration of the diffusion profile according to Eq.5.32 in a case where c_0=0. It may be noted that at the point where $\frac{c-c_0}{c_s-c_0}$ = ½, that is, at the point where the concentration of the diffusing species is midway between the surface composition and the original composition, then x/$\sqrt{D_t t}$ = 0.954. Thus at this point x ~ $\sqrt{D_t t}$.

It may be noted that Eqs.5.31 and 5.32 involve the dimensionless parameter x/(2$\sqrt{D_t t}$), and accordingly the penetration and the amount of the diffusing species dissolving in the solid are proportional to the square root of time.

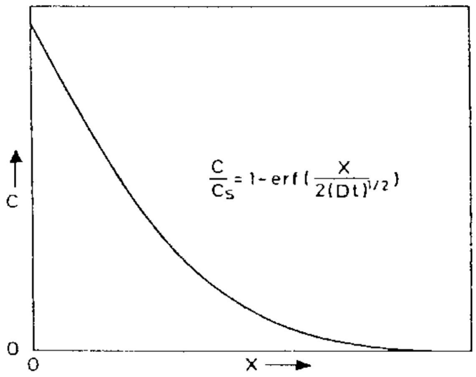

Figure 5-6. Graphical presentation of diffusion profile when the surface concentration of the diffusing species remains constant with time. It is assumed that c_O=0 (Eq.5.32).

The penetration by the diffusing species may be measured by means of the so-called sectioning method, that is cutting, grinding or etching off thin sections of layers of the sample parallel to the plane surface and subsequently determining the concentration of the diffusing species in each section. By cutting the specimen normal to or at an angle to the plane surface, the penetration of radioactive tracers may also be measured by means of so-called autoradiography.

Electron microprobe analysis and secondary ion mass spectroscopy (SIMS) combined with sputtering techniques also provide excellent tools for studying penetration and diffusion of foreign ions. As oxygen does not have a radioactive isotope suitable for tracer studies, SIMS is particularly useful for studying oxygen diffusion employing the stable ^{18}O isotope.

In the isotope exchange method the vapour of the diffusing component surrounding the sample is enriched with either a radioactive or an inactive isotope, and the diffusion is measured by following the exchange of the isotope with the sample. The diffusion coefficient may be evaluated if diffusion is the slower process and processes at the surface are rapid. Alternatively, the sample itself may be isotopically enriched, and the increase in concentration of the isotope in the vapour phase may be measured.

Diffusion rates may, in principle, also be determined from any property or reaction which depends on atomic mobility. By way of illustration, ionic conductivity of the anion is directly proportional to the anion diffusion coefficient (see Electrical conductivity). From high temperature solid state reactions, sintering, oxidation of metals etc. diffusion coefficients may be evaluated provided the detailed mechanism of the processes are known. Examples of this will be given in Chapter 7.

Diffusion mechanisms

Lattice diffusion takes place through the movement of point defects. The presence of different types of defects gives rise to different mechanisms of diffusion. These are illustrated schematically for elemental solids in the following descriptions. But they also apply to metal oxides and other inorganic compounds when the diffusion is considered to take place in the sublattices of the cations or anions.

Vacancy mechanism

The diffusion is said to take place by the vacancy mechanism if an atom on a normal site jumps into an adjacent unoccupied lattice site (vacancy). This is illustrated schematically in Fig.5.7. It should be noted that the atoms move in the direction opposite the vacancies.

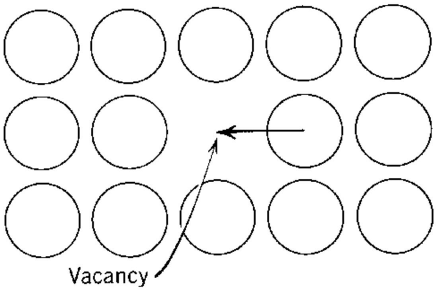

Figure 5-7. Schematic illustration of vacancy diffusion in solids.

Interstitial mechanism

If an atom on an interstitial site moves to one of the neighbouring interstitial sites, the diffusion occurs by an interstitial mechanism. This is schematically shown in Fig.5.8. Such a movement or jump of the interstitial atom involves a considerable distortion of the lattice, and this mechanism is probable when the interstitial atom is smaller than the atoms on the normal lattice positions. Diffusion of interstitially dissolved light atoms, e.g. H, C, N, and O in metals provides the best known examples of this mechanism.

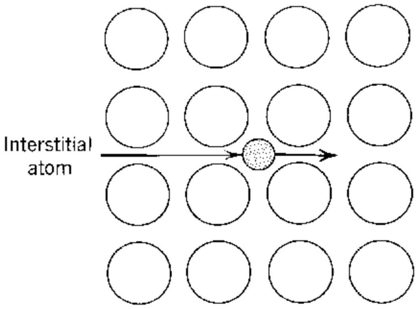

Figure 5-8. Schematic illustration of interstitial diffusion in solids.

Oxides with close-packed oxygen lattices and only partially filled tetrahedral and octahedral sites may also facilitate diffusion of metal ions in the unoccupied, interstitial positions. Finally, even large anions may diffuse interstitially if the anion sublattice contains structurally empty sites in lines or planes which may serve as pathways for interstitial defects. Examples are rare earth sesquioxides (e.g. Y_2O_3) and pyrochlore-type oxides (e.g. $La_2Zr_2O_7$) with fluorite-derived structures and brownmillerite-type oxides (e.g. $Ca_2Fe_2O_5$) with perovskite-derived structure.

Interstitialcy mechanism

If the distortion becomes too large to make the interstitial mechanism probable, interstitial atoms may move by another type of mechanism. In the interstitialcy mechanism an interstitial atom pushes one of its nearest neighbours on a normal lattice site into another interstitial position and itself occupies the lattice site of the displaced atom. This mechanism is illustrated schematically in Fig.5.9

In the interstitialcy mechanism one may distinguish between two types of movements. If the atom on the normal lattice site is pushed in the same direction as that of the interstitial atom, the jump is termed collinear (Fig.5.9). If the atom is pushed to one of the other neighbouring sites so that the jump direction is different from that of the interstitial atom, the jump is termed non-collinear.

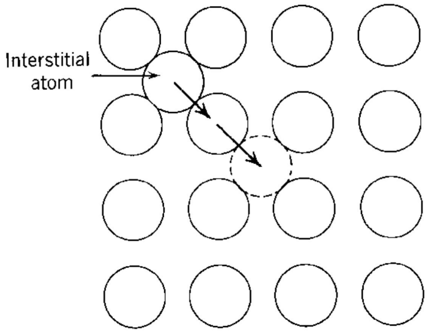

Figure 5-9. Schematic illustration of interstitialcy diffusion in solids. Collinear jump.

Other mechanisms.

In elemental solids also other mechanisms have been proposed. The crowdion is a variant of the interstitialcy mechanism. In this case it is assumed that an extra atom is crowded into a line of atoms, and that it thereby displaces several atoms along the line from their equilibrium positions. The energy to move such a defect may be small, but it can only move along the line or along equivalent directions.

For metals it has also been proposed that diffusion may take place through a so-called ring mechanism, but this mechanism is improbable in oxides or other inorganic compounds.

Diffusion of complex defects.

As defects in oxides with large deviations from stoichiometry constitute complex defects, there has been considerable discussion and speculation about the diffusion mechanism in such oxides. It has, for instance, been suggested that complex defects coexist in a dynamic equilibrium with single defects, and that the diffusion processes also under these conditions really involve diffusion of single defects. In the case of defect clusters it has alternatively been proposed that the smaller clusters may move as a unit. A translational mechanism that has been proposed for a 4:1 cluster in wustite is illustrated in Fig.5.10. The jump processes in the motion of a 4:1 complex is quite complex, and the mechanism requires two distinct, sequential jumps. Atom 1 jumps to fill a vacancy in the complex defect and thereby creates a new vacancy. In the process the interstitial ion in the cluster is displaced to the neighbouring vacancy. The movement of the cluster is completed when atom 2 is displaced from a normal site to the interstitial site of the new location of the cluster.

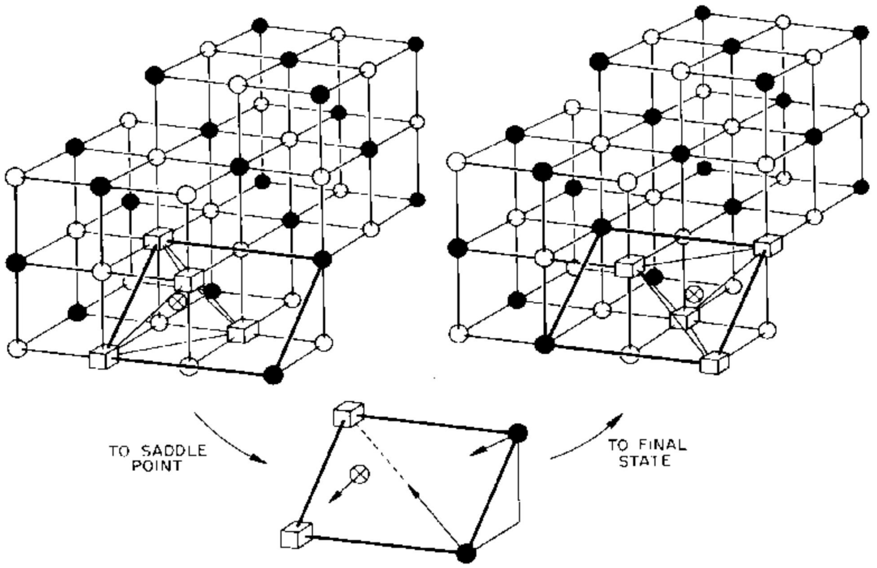

Figure 5-10. Proposed mechanism for the translational movement for a 4:1 cluster. Open circles represent oxygen ions, closed circles iron atoms, open circles with an "x" iron interstitials, and cubes iron vacancies.

Diffusion of protons in metal oxides.

Protons that dissolve in metal oxides associate with the oxide ions and form hydroxide ions. As the proton has no electron shell, it interacts strongly with the electron cloud of the oxide ion and in its equilibrium position in the hydroxide ion it is embedded in the valence electron cloud. The O-H bond that is formed has a bond length less than 100 pm; this may be compared with the ionic radius of 140 pm of the oxide ion.

In principle the protons may move by two different mechanisms: i) the free transport mechanism, which is also alternatively termed the Grotthuss mechanism or ii) the vehicle mechanism.

The free transport is the principal mode of transport of protons in oxides, and in this mechanism protons jump from one oxygen ion to a neighbouring one. After each jump the proton in the hydroxide rotates such that the proton reorients in the electron cloud and becomes aligned for the next jump. This is illustrated schematically in Fig.5.11. The rotation and reorientation is believed to involve a small activation energy and the jump itself is considered to be the rate-determining step.

In the vehicle mechanism the proton is transported as a passenger on an oxide ion. Thus this mechanism may be considered to constitute transport of hydroxide ions. The hydroxide ion may in principle move by an oxygen vacancy mechanism or as an interstitial hydroxide ion. It may be noted that the hydroxide ion has a smaller radius and charge than the oxide ion and may as such be expected to have a smaller activation energy for diffusion than the oxygen ion. Also other species such as water molecules and hydronium ions, H_3O^+, may serve as vehicles for protonic diffusion, notably in relatively open structures.

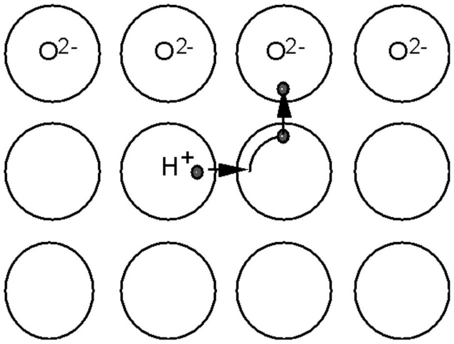

Figure 5-11. Schematic illustration of free transport of protons in metal oxides (Grotthuss mechanism).

Further considerations of factors that affect the diffusion coefficient in crystalline solids

In Eq. 5.19 the random diffusion coefficient has been expressed in terms of the jump distance and the number of jumps per unit time:

$$D_r = \tfrac{1}{6}s^2\Gamma = \tfrac{1}{6}s^2\frac{n}{t} \tag{5.19}$$

But it is necessary to further characterise D in different crystal structures (lattices or sub-lattices) and subsequently to derive expressions for the temperature and oxygen pressure dependence of diffusion in metal oxides.

Vacancy diffusion.

Let us consider vacancy diffusion in an elemental solid or a cation or an anion sub-lattice. The number of jumps per unit time, Γ, depends on several factors. First, it depends on the jump frequency ω towards an adjacent site. Furthermore, it is also proportional to the number of sites to which the atom may jump, i.e., the number of nearest neighbour positions of the atom, Z. Finally, the atom may only jump if a vacancy is located on an adjacent site, and this probability is given by the fraction (concentration) of vacancies in the crystal, N_d. Thus, Γ is in this case given by

$$\Gamma = \omega Z N_d \tag{5.33}$$

In crystalline solids the jump distance is a function of the crystal structure and may be expressed as a function of the lattice parameter.

In a bcc crystal of an elemental solid, for instance, each atom has 8 nearest-neighbour positions or atoms, and thus in this case Z = 8. From simple geometrical considerations of the crystal structure it may further be shown that the jump distance is given by $s = a_0 \frac{\sqrt{3}}{2}$, where a_o is the lattice parameter. When inserting these values of Z and s in Eq. 5.33 one obtains

$$D_r = a_0^2 \omega N_d \tag{5.34}$$

In general, D_r for a cubic structure is written

$$D_r = \alpha a_0^2 \omega N_d \tag{5.35}$$

where α is a geometrical factor involving the factor 1/6 (from Eq.5.21, the factor Z from Eq.5.33 and the relation between the jump distance and the lattice parameter. For vacancy diffusion in a bcc lattice α is thus equal to unity, From the same considerations it may also be shown that $\alpha = 1$ for vacancy diffusion in fcc lattices.

Interstitialcy diffusion

Consider an atom on a normal lattice site of a cation or anion sub-lattice. If this atom is to move by the interstitialcy mechanism, an atom on a nearest neighbour interstitial site has to push the atom on the normal site to a neighbouring interstitial site. Thus for this diffusion mechanism an atom may only diffuse when it has an interstitial atom on a neighbouring site, and as for vacancy diffusion the diffusion coefficient of the atoms is proportional to the fraction (concentration) of interstitial atoms or ions in the sub-lattice.

Interstitial diffusion

When one considers interstitial diffusion of an interstitially dissolved species in dilute solid solution, essentially all the nearest neighbour interstitial sites of the same type are unoccupied and available for occupancy by the diffusing interstitial atoms. Thus the interstitial atom may jump to any of the nearest neighbour interstitial sites and in this case N_d is equal to unity. The interstitial diffusion coefficient is then given by

$$D_r = \tfrac{1}{6} s^2 Z\omega \tag{5.36}$$

As mentioned above the best known examples of this mechanism is diffusion of O, N, C, and H atoms interstitially dissolved in metals. By way of example oxygen and nitrogen atoms in bcc metals, e.g. in the group 5 metals V, Nb, and Ta, occupy octahedral sites, and in this case each interstitial atom has 4 nearest neighbour octahedral sites to which they may jump, thus

Z= 4. Furthermore, the jump distance s is equal to $\frac{a_o}{2}$. Inserting these values in Eq.5.36 the diffusion coefficient for interstitial diffusion between octahedral sites in a bcc lattice becomes

$$D_r = \tfrac{1}{6} s^2 \omega \tag{5.37}$$

Thus in this case α=1/6.

Free proton transport

Similar considerations may be applied to free transport of protons (cf. Fig.5.11). For dilute solutions of protons in an oxide essentially all nearest neighbour oxygen ions are available, and thus in this case N_d is unity. However, the specification of Z, s and ω is not straightforward in this case. The dynamics of free proton diffusion in oxides are complicated by 1) the multistep process (jump+rotation), 2) the dependency on the dynamics of the oxygen ion sublattice, and 3) the quantum mechanical behaviour of a light particle such as the proton.

The uniquely large ratios between the masses of the isotopes of hydrogen give rise to a number of strong isotope effects in the case of diffusion of protons. (These are also in principle operative for diffusion of hydrogen atoms or hydride ions, but they would be essentially negligible for diffusion of protons on a heavier vehicle, such as in OH^-.) The isotope effects for proton diffusion can be classified as follows: The attempt frequency ν (in the pre-exponential of ω) is given as the O-H stretching frequency, and it is given by the inverse of the square root of the reduced mass of the harmonic oscillator. The reduced mass equals $(m_O+m_H)/m_Om_H$ and is roughly inversely proportional to the mass of the hydrogen species. Therefore, the ratios of the pre-exponentials of the diffusion coefficients of protons, deuterons, and tritons are approximately related by $D_{0H} : D_{0D} : D_{0T} = 1 : 1/\sqrt{2} : 1/\sqrt{3}$. This is called the classical effect. Furthermore, the oscillators have different ground-state or zero-point energies, such that diffusion of lighter isotopes may be expected to have a slightly smaller activation energy of jumping. Accordingly, proton diffusion typically has 0.04 - 0.06 eV lower activation energy than deuteron diffusion. This is called the non-classical effect. However, there are more factors involved, connected to the fact that the light proton/deuteron/triton must be treated quantum mechanically and to their dynamics in a dynamic lattice of much heavier oxygen ions. For instance, the lighter isotope has a lower sticking probability after an otherwise successful jump. In effect, this reduces the effective diffusivity of all hydrogen isotopes and it can to a varying degree counteract or even seemingly reverse the classical effect. The neglectance of the latter have in general made many investigators ascribe higher diffusivities for protons compared to deuterons to the classical effect, while in reality the non-classical zero-point energy difference appears to be the main contributor to the observed effect.

The possibility of tunnelling as a major component of diffusion is not expected to apply to protons except at very low temperatures, and the orders-of-magnitude isotope effects that would be expected for proton vs deuteron or triton diffusion have not been reported for oxidic materials at elevated temperatures.

Temperature and oxygen pressure dependence of diffusion in metal oxides

We will now address how diffusion coefficients vary with temperature and activity of the components of the compound, mainly the oxygen partial pressure over oxides. In order to evaluate these dependencies one must consider the temperature and activity dependencies of N_d (e.g. cf. Eq.5.35) and the temperature dependence of ω. We start by analysing N_d for some cases.

Dependencies related to the concentration of defects

Vacancies in an elemental solid

For the sake of simplicity let us first consider the case of diffusion by a vacancy mechanism in a pure elemental solid. The diffusing atoms may only make a jump when a neighbouring site is vacant. Thus the jump frequency n/t is proportional to the vacancy concentration and as described in Chapter 3 the fraction of vacancies may be written

$$N_d = \exp(\frac{-\Delta G_d}{RT}) = \exp(\frac{\Delta S_d}{R})\exp(\frac{-\Delta H_d}{RT}) \qquad (5.38)$$

where ΔG_d, ΔS_d, and ΔH_d denote the Gibbs free energy, entropy, and enthalpy of formation of the vacancies (defects). In elemental solids ΔH_d is positive and the vacancy concentration increases with increasing temperature.

Vacancies in an oxygen deficient oxide.

In a nonstoichiometric oxide the concentration of the predominating point defects will be a function of temperature but also of the oxygen pressure. By way of illustration, let us consider an oxygen deficient oxide, $M_aO_{b-\delta}$, in which doubly charged oxygen vacancies are the predominating point defects. If intrinsic ionisation and effects of impurities can be neglected, the concentration of the oxygen vacancies is as described in previous chapters (Eq. 3.68) given by

$$N_d = [v_O^{\bullet\bullet}] = (\tfrac{1}{4}K_{v_O^{\bullet\bullet}})^{1/3}\, p_{O_2}^{-1/6} = (\tfrac{1}{4})^{1/3}\, p_{O_2}^{-1/6} \exp(\frac{\Delta S_{v_O^{\bullet\bullet}}}{3R}) \exp(\frac{-\Delta H_{v_O^{\bullet\bullet}}}{3RT}) \tag{5.39}$$

where $K_{v_O^{\bullet\bullet}}$ represents the equilibrium constant for the formation of doubly charged oxygen vacancies, and $\Delta S_{v_O^{\bullet\bullet}}$ and $\Delta H_{v_O^{\bullet\bullet}}$ their entropy and enthalpy of formation, respectively.

In such an oxide the oxygen vacancy concentration may also be determined by the presence of dopants or a sufficiently large level of impurities with negative effective charge. If the dopant is doubly negatively charged (as, for instance, Ca^{2+}, an acceptor dopant, in ZrO_2), then

$$[v_O^{\bullet\bullet}] = [\mathrm{A}_M^{//}] \tag{5.40}$$

In this case $[v_O^{\bullet\bullet}]$ will be independent of oxygen pressure and most often temperature (see Chapter 4).

In a similar manner one may obtain the temperature and oxygen pressure dependencies of the concentration of any defect (majority or minority) when the defect structure is known. Of course the concentrations of acceptors or donors will enter when they dominate the defect situation and the water vapour partial pressure may enter when protons are dominating defects.

Temperature dependence of the attempt frequency ω

When atoms jump or move between definite sites in the crystal, they have to surmount energy barriers. A large part of this energy barrier involves the strain energy required to displace neighbouring atoms to create a sufficiently large opening between the atoms to permit the atom jump. The potential energy of the atom diffusing from one site to another may be qualitatively illustrated as shown in Fig. 5.12. The potential barrier height is ΔH_m and represents the activation energy which the atom has to surmount during the jump. Each atom vibrates in its position and during a fraction of time, which is given by the Boltzmann distribution factor exp $(-\Delta H_m/RT)$, it possesses sufficient energy to overcome the energy barrier. The jump frequency is thus proportional to $\exp(-\Delta H_m/RT)$.

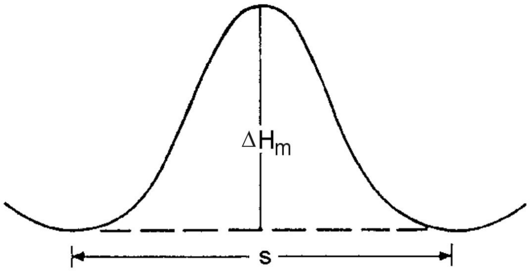

Figure 5-12. Potential energy of atom diffusing in a solid. ΔH_m is the activation energy for the diffusion.

A more complete analysis based on the theory of activated complexes and on statistical mechanics has been given by Zener (1951,1952). He considered the system or an atom in its initial equilibrium condition and in the activated state at the top of the potential barrier which separates the initial position from its neighbouring equilibrium position. The rate of transition from one equilibrium site to another is given by

$$\omega = \upsilon \exp\frac{-\Delta G_m}{RT} = \upsilon \exp\frac{\Delta S_m}{R}\exp\frac{-\Delta H_m}{RT} \qquad (5.41)$$

where ΔG_m, ΔS_m, and ΔH_m represent the free energy, entropy and enthalpy change, respectively, connected with the movement of the atom from the equilibrium position to the top of the potential barrier and υ (*"nu"*) represents the vibration frequency. υ is often assumed to equal the Debye frequency, i.e. about 10^{13} Hz, as an order of magnitude approximation. (For protons in oxides the O-H stretching frequency of about 10^{14} Hz is used.) ΔS_m is often assumed to be a small, positive term (below, say, 10 J/molK).

As a more detailed analysis of υ, Zener suggested that υ may be approximated as $\upsilon = \alpha/a \cdot \sqrt{\Delta H_m / M}$, where α is a structure- and mechanism-dependent factor, a is a lattice parameter, and M is the reduced mass of the oscillator. Intuitively, a particle vibrating in an energy valley will vibrate faster (υ increases) when the walls become steeper (ΔH_m increases). One may also view the particle as vibrating on a spring; the frequency becomes higher when she spring is shorter (α/a decreases), when the spring is stiffer (ΔH_m increases) or the particle becomes lighter (M decreases). Thus, the temperature-independent (pre-exponential) term in the diffusion coefficient increases when ΔH_m increases, so that the two tend to counteract each other. Experimental observations of this is sometimes referred to as the Meyer-Neldel effect (Meyer and Neldel (????)).

Resulting analysis of the diffusion coefficient D_r

Vacancy diffusion in an elemental solid. From Eq.5.35 it is seen that the temperature dependence of D_r for vacancy diffusion in an elemental solid is determined by that of N_d and ω. For an elemental solid with cubic structure, D_r is thus obtained by combining Eqs.5.35, 5.38 and 5.41:

$$D_r = \alpha a_0^2 \upsilon \exp\frac{\Delta S_d + \Delta S_m}{R} \exp\frac{-(\Delta H_d + \Delta H_m)}{RT} \tag{5.42}$$

Experimentally determined values of diffusion coefficients are usually obtained as

$$D = D_0 \exp\frac{-Q}{RT} \tag{5.43}$$

where Q is termed the activation energy and D_o the the pre-exponential factor.

By comparing Eqs.5.42 and 5.43 it is seen that the activation energy, Q, in this case comprises

$$Q = \Delta H_d + \Delta H_m \tag{5.44}$$

Correspondingly, D_o is given by

$$D_0 = \alpha a_0^2 \upsilon \exp\frac{\Delta S_d + \Delta S_m}{R} \tag{5.45}$$

If experiments are carried out under such conditions that the concentration of defects, N_d, is constant and independent of temperature, e.g. at sufficiently low temperatures that the defect concentration is frozen in, then D_r is given by

$$D_r = \alpha a_0^2 \upsilon N_{d,frozen} \exp\frac{\Delta S_m}{R} \exp\frac{-\Delta H_m}{RT} = D_0 \exp\frac{-\Delta H_m}{RT} \tag{5.46}$$

and correspondingly the activation energy is under these conditions given simply by

$$Q = \Delta H_m \tag{5.47}$$

Oxygen vacancy diffusion in oxygen-deficient oxides.

In an oxygen-deficient oxide in which oxygen vacancies predominate and for which effects of impurities can be neglected, the oxygen vacancy concentration is given by Eq. 5.39. Correspondingly, the oxygen diffusion coefficient for random oxygen vacancy diffusion in the oxide in equilibrium with the ambient oxygen gas at a partial pressure becomes

$$D_r = \alpha a_0^2 \upsilon (\tfrac{1}{4})^{1/3} p_{O_2}^{-1/6} \exp\frac{\frac{\Delta S_{v_O^{\bullet\bullet}}}{3} + \Delta S_m}{R} \exp\frac{-(\frac{\Delta H_{v_O^{\bullet\bullet}}}{3} + \Delta H_m)}{RT} \tag{5.48}$$

D_r thus increases with decreasing oxygen pressure. The activation energy for the diffusion is in this case given by

$$Q = \frac{\Delta H_{v_O^{\bullet\bullet}}}{3} + \Delta H_m \tag{5.49}$$

If the concentration of oxygen vacancies is determined by lower valent impurities or dopants, e.g. when Eq.5.40 applies, then D_r is given by

$$D_r = \alpha a_0^2 \upsilon [A_M^{//}] \exp\frac{\Delta S_m}{R} \exp\frac{-\Delta H_m}{RT} \tag{5.50}$$

Thus in this case the activation energy is equal to that of the mobility of the oxygen vacancies: $Q = \Delta H_m$. A similar situation would arise if the vacancy concentration was frozen in rather than determined by acceptor doping.

Such a transition from intrinsic to extrinsic diffusion may take place when the temperature is lowered from high temperatures, where the native point defects predominate, to low temperatures, where the point defect concentrations are determined by the impurity concentration or, in other cases, frozen in. The temperature dependencies and the corresponding change in activation energy of the random diffusion in such a case are illustrated in Fig. 5.13.

The above situations represent ideal cases. As the temperature is decreased defect interactions may become increasingly important. This may be treated as formation of associated defects. In the non-stoichiometric (intrinsic) case one may for instance have association between charged vacancies and electrons forming singly charged or neutral vacancies. This will change the oxygen pressure dependency of N_d and probably the temperature dependencies of N_d and ω, but probably not dramatically. Of larger effect, and more frequently observed, are the associations between the mobile vacancies and the relatively stationary acceptors in the extrinsic regime: The associated vacancies can be regarded as immobilised, and the concentration of mobile vacancies (N_d) starts to decrease with decreasing temperature. This is seen as an increasing activation energy of diffusion with decreasing temperature in many heavily doped oxides. One may note that instead of expressing the effect as a changing concentration of free vacancies one may express it as a changing mobility in that the activation energy for diffusion is increased by the trapping energy exerted by the acceptor. However, the simple model of free and stationary (associated) point defects appears capable of explaining most behaviours fairly well.

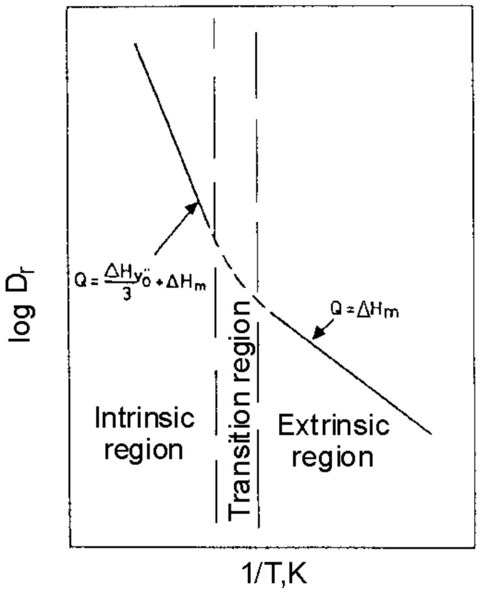

Figure 5-13. The diffusion coefficient for oxygen diffusion by the vacancy mechanism in an oxygen deficient oxide in which oxygen vacancies are the predominant native point defects. At high temperatures the oxide exhibits intrinsic behaviour and at reduced temperatures extrinsic behaviour (i.e. the oxygen vacancy concentration is determined by the concentration of lower valent cations).

Interstitial diffusion of solute.

From Eq.5.36 it is seen that N_d does not enter into the expression for the diffusion coefficient for interstitial diffusion in dilute solutions, thus in this case the activation energy, Q, represents that of the mobility of the diffusing interstitial atoms: $\Delta H_m = Q$.

For interstitial diffusion between octahedral sites in bcc metals D_o is by combination of Eqs. 5.37 and 5.41 given by

$$D_0 = \tfrac{1}{6} a_0^2 \upsilon \exp \frac{\Delta S_m}{R} \qquad (5.51)$$

Assuming that $\upsilon \sim 10^{13}$ s^{-1} and $a_o = 1.5 \cdot 10^{-8}$ cm, and as it is probable that $\Delta S_m > 0$, one may estimate a lower limit of D_o of

$$D_o > 5 \cdot 10^{-4}\ \text{cm}^2\text{s}^{-1} \tag{5.52}$$

The diffusion of protons by the free transport mechanism is another case of interstitial diffusion of a solute.

Interstitial diffusion of a constituent

The diffusivity of a constituent such as the host metal or oxygen ions by an interstitial mechanism is not only proportional to the probability that the interstitial defect jumps, but also to the probability that a constituent ion is interstitial, i.e., the fractional concentration of interstitials. Thus the diffusion coefficient of the constituent contains the temperature and oxygen pressure dependencies of the concentration of interstitials in addition to the temperature dependency of the mobility of these defects. As in the case of vacancy diffusion, the fixation of the defect concentration by doping or freezing as well as association and trapping of defects apply also to interstitial diffusion.

Diffusion coefficients of point defects

In the above treatment of vacancy diffusion, only the diffusion coefficients of the atoms have been considered. For many purposes it may be convenient to consider the diffusion coefficients of the vacancies themselves.

When an atom diffuses by the vacancy mechanism, it can only jump if a vacancy is located on an adjacent site, and the number of jumps per unit time is thus proportional to N_d (Eq. 5.33). However, the vacancy itself can jump to any one of the occupied nearest neighbour positions, provided it is occupied by an atom. Accordingly the vacancy diffusion coefficient D_V for a cubic system is given by (cf. Eq.5.35)

$$D_V = \alpha\, a_o^2\, \omega\, N \tag{5.53}$$

where N is the fraction of occupied atom positions. In dilute solutions of vacancies, N~1, and the diffusion coefficient of the vacancies is then not dependent on N_d.

From Eqs. 5.35 and 5.53 D_V for vacancies is related to D_r for the atoms through the relation

$$D_r N = D_V N_d \qquad (5.54)$$

where, as stated above, N denotes the fraction of sites occupied by atoms, often approximated as ~1. This relation can be generalised to be a very important and useful approximation for any point defect:

$$D_r N = D_d N_d \qquad (5.55)$$

where D_d is the defect diffusion coefficient. It proves to be a good approximation for component diffusion by the interstitialcy mechanism. It is also a useful approximation for component diffusion by interstitial diffusion in the case of small defect concentrations (N~1) but as the concentration of defects increases the term N must reflect the number of unoccupied interstitial sites.

Eq. 5.55 still holds for interstitial diffusion of a dilute solution of an interstitially dissolved solute, such as light elements (H, C etc. in metals or protons in oxides), but in this case the number of solute atoms or ions and the number of defects is of course the same, so that in these cases the diffusion coefficient for the solute and for the (interstitial) defects is the same.

Excercises

1. Random (self) diffusion

a) The self-diffusion coefficient of a metal with cubic structure can be expressed as

$$D = \frac{1}{6}\frac{n}{t}s^2$$

where n/t represents the jump frequency (i.e. number of jumps n over a time t). Close to the melting point most fcc and bcc metals have $D \approx 10^{-8}$ cm^2/s.

i) If the jump distance is 3 Å, what is the jump frequency near the melting point?

ii) What is the relation between this frequency and the vibrational frequency?

iii) How far has one atom traveled after 1 hour?

iv) What is the root mean square displacement after one hour?

v) What is the root mean square displacement in one dimension after one hour?

b) For a metal with cubic structure the diffusion coefficient can also be expressed as

$$D = \alpha a_0^2 \omega N_d$$

where α is a geometric factor, a_0 the lattice constant, ω is the jump frequency, and N_d is the defect concentration. Derive the value for α for vacancy diffusion in a metal with fcc structure.

6. Electrical conductivity

Introduction

In the preceding chapter we have described and discussed diffusion of particles in solids and particularly of ions and defects in metal oxides. The driving force for the diffusion has been taken to be the negative value of the particle gradient or more precisely the negative value of the chemical potential gradient. When using isotopes as tracers one may study self-diffusion, i.e. diffusion of the components in the oxide (metal and oxygen ions) in a homogeneous oxide; in this case the isotopic tracer gradient is the driving force for the diffusion.

In this chapter the transport of electrical charges will be described and discussed. In metal oxides the electrically charged particles comprise ions and electrons. The ionic charge carriers comprise the cations, anions, and foreign ions (e.g. impurity ions, dopant ions and protons) and the electronic charge carriers are the electrons and electron holes. The concentrations of the charge carriers are directly related to the defect structure of the oxide and in this chapter we will derive expressions for the temperature and oxygen pressure dependence of the electrical conductivity. The discussion will be limited to transport of charges in chemically homogeneous metal oxides (no chemical potential gradient) but with an electrical potential gradient as the driving force. In the next chapter transport of ionic and electronic charge carriers in metal oxides which are simultaneously exposed to chemical and electrical potential gradients, i.e. electrochemical potential gradients, will be discussed.

As the mobilities of electrons and electrons holes are normally much higher than those of ions, most oxides are electronic conductors. One type of charge carrier often predominates in an oxide under particular conditions of temperature and oxygen pressure. An electronically conducting oxide is an n-conductor if transport of electrons predominate and a p-conductor if electron holes prevail. However, some oxides are or may become ionic conductors or mixed ionic/electronic conductors depending on the temperature and oxygen pressure often as a result of appropriate doping with aliovalent foreign ions. Some oxides may also exhibit proton conductivity in hydrogen- or water vapour-containing atmospheres; predominant proton conductivity in such oxides is in some cases observed at reduced temperatures (< 600-700 °C).

Transport in an electrical potential gradient

As described in the previous chapter on diffusion in metal oxides the driving force is given by the negative of the potential gradient. The force exerted on a charged particle of type i with charge z_ie is given by

$$F = -z_i e \frac{d\phi}{dx} = z_i eE \tag{6.1}$$

where ϕ is the electrical potential and $E = -\frac{d\phi}{dx}$ is the electric field. The flux of particles of type i is the product of the concentration c_i, the particle mobility B_i, and the force F:

$$j_i = c_i B_i F = z_i e\, c_i B_i E \tag{6.2}$$

The current density i_i is given by the product of flux and charge:

$$i_i = z_i e j_i = (z_i e)^2 B_i\, c_i\, E \tag{6.3}$$

While B_i is the particle mechanical mobility ("beweglichkeit"), the product of B_i and the charge on each particle, $z_i e$, is termed the charge mobility u_i:

$$u_i = z_i e B_i \tag{6.4}$$

Equation 6.3 can then be written

$$i_i = z_i e\, c_i u_i\, E = \sigma_i\, E \tag{6.5}$$

where $\sigma_i = z_i e\, c_i u_i$ is the electrical conductivity due to the charge carriers of type i. The electrical conductivity is determined by the product of the concentration c_i of the charged particles, the charge $z_i e$ on the particles and the charge carrier mobility, u_i. It should be noted that Eq.6.4 is an expression of Ohm's law. The unit for the electrical conductivity is Siemens per cm, Scm^{-1} (one Siemens is the reciprocal of one ohm and in older literature the electrical conductivity is expressed as $ohm^{-1}cm^{-1}$). The unit for the charge is coulomb, the concentration of charge carriers is expressed as the number of charge carriers of type i per cm^3, and charge carrier mobility in units of cm^2/Vs. (Although the SI unit for length is m, cm is being used in the following as it is still by far the one most commonly used in the literature).

It may be noted that in the above terminology, F, E, i_i, z_i, u_i and j_i may each be positive or negative. u_i and z_i always have the same sign, and as long as no other forces than the the electrical act, i_i and E_i always have the same sign, and j_i and F always have the same sign. B_i and σ_i are always positive, and it is common also to neglect the sign when specifying charge mobilities u_i.

The total electrical conductivity σ of a substance is the sum of the partial conductivities σ_i of the different charge carriers:

$$\sigma = \sum_i \sigma_i \tag{6.6}$$

The ratio of the partial conductivity σ_i to the total conductivity σ is termed the transport (or transference) number of species i:

$$t_i = \frac{\sigma_i}{\sigma} \tag{6.7}$$

Charge carriers in ionic compounds

The native charge carriers in a binary oxide are the cations, anions, electrons, and electron holes. The total conductivity is then given by

$$\sigma = \sigma_c + \sigma_a + \sigma_n + \sigma_p \tag{6.8}$$

where σ_c, σ_a, σ_n, and σ_p are the cation, anion, electron and electron hole conductivities, respectively.

Following Eq. 6.7 the individual conductivities may be written in terms of their transport numbers: $\sigma_c = t_c\,\sigma$, $\sigma_a = \sigma\, t_a$, $\sigma_n = \sigma\, t_n$ and $\sigma_p = \sigma\, t_p$. Using these values Eq.6.7 takes the form

$$\sigma = \sigma\,(t_c + t_a + t_n + t_p) \tag{6.9}$$

It may be noted that the sum of the transport numbers of all the charge carriers equals unity:

$$t_c + t_a + t_n + t_p = 1 \tag{6.10}$$

The total electrical conductivity is often given by the sum of the ionic conductivity, $\sigma_{ion} = \sigma_c + \sigma_a$, and the electronic conductivity, $\sigma_{el} = \sigma_n + \sigma_p$, and the total conductivity can then be written

$$\sigma = \sigma_{ion} + \sigma_{el} \tag{6.11}$$

Often only one type of charge carrier dominates the charge transport, and in many cases and as an approximation contributions from minority carriers are neglected. For oxides the mobilities of electrons and electron holes are usually several orders of magnitude (~10^4 - 10^8) larger than those of the ions, and even when the concentration of electron or electrons holes is smaller than that of the ionic charge carriers (or, more precisely, than that of ionic charge carrier defects) the oxide may still be a predominantly electronic conductor. The relative importance of ionic and electronic conductivity will often vary greatly with temperature and oxygen pressure. This will be illustrated in the following chapters.

The Nernst-Einstein relation between mobility and diffusion coefficient

In the previous chapter it was shown that the relation between the random diffusion coefficient of particles of type i and the particle's mechanical mobility is given by

$$D_i = kTB_i \tag{6.12}$$

By combining this relation with Eqs. 6.4 and 6.5 one obtains the following relation between the random diffusion coefficient and the charge carrier mobility and the electrical conductivity:

$$D_i = kTB_i = u_i \frac{kT}{z_i e} = \sigma_i \frac{kT}{c_i z_i^2 e^2} \tag{6.13}$$

This relation is called the Nernst-Einstein relation. This relation and also the effect of an applied electric field on migration of charged species in a homogeneous crystal may also be derived from the following model, in which we will understand also when and why conduction is termed a linear process.

Consider a one-dimensional system with a series of parallel planes separated by a distance s (cf. Fick's first law in Chapter 5). It is assumed that the system is homogeneous and that the volume concentration of the particles in the planes is c_i. The particles in neighbouring planes 1 and 2 have equal probability of jumping to the neighbouring planes. In the absence of any external kinetic force, the number of particles which jump from plane 1 to plane 2 and from 2 to 1 per unit time is equal and opposite and given by $\frac{1}{2}\,\omega c_i s$. In a homogeneous system there will be no net transport of particles.

When there is no applied electric field, the activation energy associated with the jumps is ΔH_m. When an electric field E is applied, the jump frequency in the positive direction will be increased and that in the negative direction decreased in that the activation energies are changed. In the forward direction the activation energy is reduced to $\Delta H_m - \frac{1}{2} z_i esE$ and in the reverse direction increased to $\Delta H_m + \frac{1}{2} z_i esE$. This is illustrated schematically in Fig.6.1.

The net particle flux is given by the difference in number of jumps in the forward and reverse directions:

$$j_i = \frac{1}{2} c_i s \,\{\omega_{forw} - \omega_{rev}\} \tag{6.14}$$

where $\omega_{forw} = \nu \exp(\frac{\Delta S_m}{k}) \exp\left(- \frac{\Delta H_m - \frac{z_i esE}{2}}{kT}\right)$

and $\omega_{rev} = \nu \exp(\frac{\Delta S_m}{k}) \exp\left(- \frac{\Delta H_m + \frac{z_i esE}{2}}{kT}\right)$.

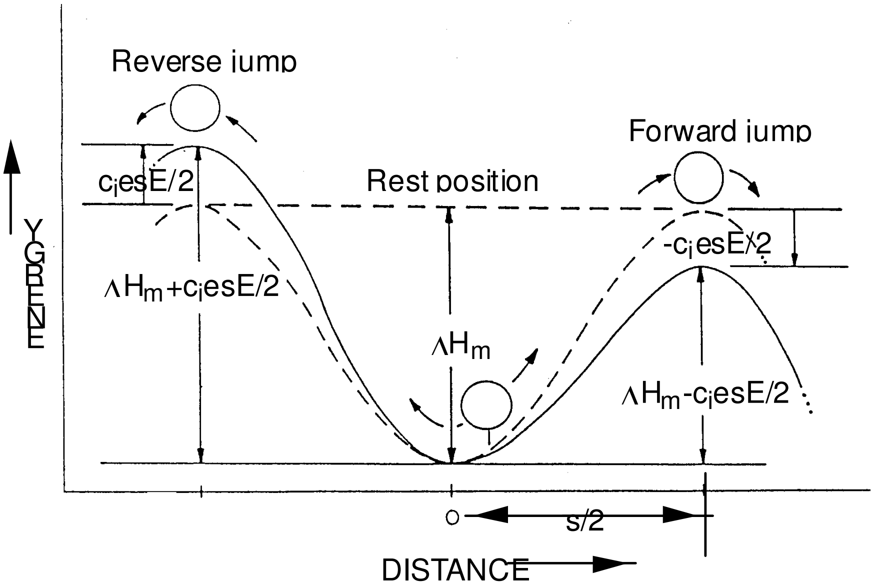

Figure 6-1. Schematic illustration of the effect of an electric field on the migration of charged species in a homogeneous crystal. E represents the electric field. ΔH_m is the activation energy in the absence of an electric field. In the forward direction the activation energy may be considered to be lowered by ½ $z_i esE$ and increased by the same amount in the reverse direction.

Equation 6.14 then becomes

$$j_i = \frac{1}{2}\, c_i s \omega \{\exp\left(\frac{z_i esE}{2kT}\right) - \exp\left(-\frac{z_i esE}{2kT}\right)\} \qquad (6.15)$$

where $\omega = \nu \exp\left(\frac{\Delta S_m}{k}\right) \exp\left(-\frac{\Delta H_m}{kT}\right)$.

When $z_i esE << 2kT$, which is valid for normal electrical measurements in bulk materials (and when Ohm's law is applicable), the difference in exponentials in Eq. 6.15 may be written $z_i esE/kT$ (since $e^x - e^{-x} = 2x$ for $x<<1$). Eq. 6.15 then takes the form

$$j_i = \frac{1}{2}\, \omega s^2 c_i \frac{z_i eE}{kT} \qquad (6.16)$$

The diffusion coefficient for one-dimensional random diffusion is given by $D_i = \frac{1}{2}\, \omega s^2$ and j_i then becomes

$$j_i = D_i \frac{c_i z_i eE}{kT} \qquad (6.17)$$

We have by this shown that net flux density of a hopping charge carrier in an electrical field in the small-signal (linear) range is proportional to the random diffusion coefficient. As we have shown before, the flux of particles with a charge $z_i e$ may also be expressed in terms of charge carrier mobility or conductivity:

$$j_i = z_i e\, c_i B_i E = c_i\, u_i\, E = \frac{\sigma_i E}{z_i e} \qquad (6.18)$$

and when one combines Eqs. 6.17 and 6.18 one obtains various forms of the Nernst-Einstein relation (Eq. 6.13):

$$D_i = B_i\, kT = u_i \frac{kT}{z_i e} = \sigma_i \frac{kT}{c_i z_i^2 e^2} \qquad (6.13)$$

or, rearranged,

$$\sigma_i = (z_i e)^2 c_i D_i / kT$$

It is emphasised that the relation is derived assuming random diffusion and that the mobilities and conductivity through this relation connects to the random (or self) diffusion coefficient D_r. It is thus meaningful for relating electrical and diffusional transport of atoms and ions. For electrons and holes this is only meaningful when they migrate by an activated hopping mechanism.

From the Nernst-Einstein relation it is also seen that the temperature dependence of the product $\sigma_i T$ is the same as that of D_r. Thus in evaluating the activation energy associated with the diffusion coefficient from conductivity measurements, it is necessary to plot ($\sigma_i T$) vs 1/T.

It is also important to note that in the derivation it is implicitly assumed that the ions and electrons move independently of each other, e.g. that there is no interference between ionic and electronic flows.

Electronic conductivity in oxides

Most metal oxides are electronic conductors at high temperatures. For many of these oxides the conductivity increases with increasing temperature and as the conductivity at the same time is much smaller than in metals, this type of conductivity is termed semiconductivity. The principal reason for the increasing conductivity is that the number of electronic defects increases with increasing temperature. A limited number of oxides - especially among transition metal monoxides - are metallic conductors and for which the conductivity decreases with increasing temperature. In this case this is attributed to a mobility of electronic defects decreasing with increasing temperature. Other oxides, e.g. p-conducting acceptor-doped perovskites to be discussed in a later chapter, also exhibit metallic-like conductivity in that the conductivity also here decreases with increasing temperature; however in these cases the decreasing conductivity is attributed to a decreasing number of electron holes with increasing temperature, and the conductivity is thus not to be classified as metallic.

The electronic conductivity, σ_{el}, of a semiconducting oxide is given by

$$\sigma_{el} = \sigma_n + \sigma_p = enu_n + epu_p \qquad (6.19)$$

where σ_n and σ_p are the electron and electron hole conductivities, n and p the charge carrier concentrations of electrons and electron holes, respectively, and u_n and u_p are the carrier (or drift) mobilities of electrons and electron holes. As mentioned above, one type of charge carrier will often dominate; however, in special cases where an oxide is close to stoichiometric both n- and p- conductivity may contribute significantly to the electronic conductivity.

In defect-chemical equations the concentrations of electrons and electron holes are often written in terms of the law of mass action without specifying where the electronic defects are located. The concentrations of electronic defects are often interpreted in terms of the band theory of solids and in the following is given a brief account of this theory and the relationship between this theory and the law of mass action.

The band theory

In single atoms the electrons may only possess discrete energies. These allowed energies are designated by quantum numbers which refer to the electron shell which the electron occupies (the principal quantum number), the orbital angular momentum of the electron (azimuthal quantum number), and the direction of the angular momentum vector (magnetic quantum number). In addition, and according to the Pauli exclusion principle, each energy state can only be accommodated by two electrons which have opposite spins.

When individual atoms are brought together in a solid, i.e. when interatomic spacing decreases and electronic levels overlap, a splitting of the energy levels begin to occur, and the energy levels may be considered to form energy bands in the solid. But the total number of levels within a band corresponds exactly to the total number of atoms present in the solid, and therefore the levels become more and more finely spaced the larger the number of atoms present. Because of the Pauli exclusion principle, each band can accommodate twice as many electrons as there are energy levels. Solids contain 10^{22} - 10^{23} atoms per cm^3, and the number of levels in each band is thus of the same order.

The energy bands may overlap or be separated by energy gaps. At absolute zero temperature (0 K) the electrons fill up the lowest possible energy levels. The highest filled band represents the orbitals of valence electrons and is termed the valence band. It is completely filled at 0 K while the next band, termed the conduction band is completely empty. In Fig.6.2, where the vertical axis represents the electron energy and horizontal axis the distance through the solid, the valence and conduction bands are separated by an energy gap, as is the case in semiconductors and insulators. In a pure, perfect and ideal solid the electrons may not possess energies within the energy gap, and this is therefore also often termed the forbidden energy gap. The size of the energy gap differs for inorganic compounds, and empirically it is found that the energy gap increases with the heat of atomisation of the compounds. For binary semiconductors and insulators it has specifically been shown that the band gap, E_g (in electron volts) can be expressed approximately as

$$E_g = 2(E_s - c) \tag{6.20}$$

where $E_s = E_{at}$/equivalent[*] and c is a constant approximately equal to 2.7.

---

[*] An equivalent in M_aO_b is equal to $az_{cat} = b|z_{an}|$

In materials where the valence band is only partly occupied or where it overlaps with the conduction band (no forbidden band gap) the electrons can move freely in the available energy levels and we have metallic conduction.

In an insulator or semiconductor at 0 K there is a band gap and the valence band is completely filled with electrons while the conduction band is completely empty. In such cases no electronic conduction takes place when an electric field is applied.

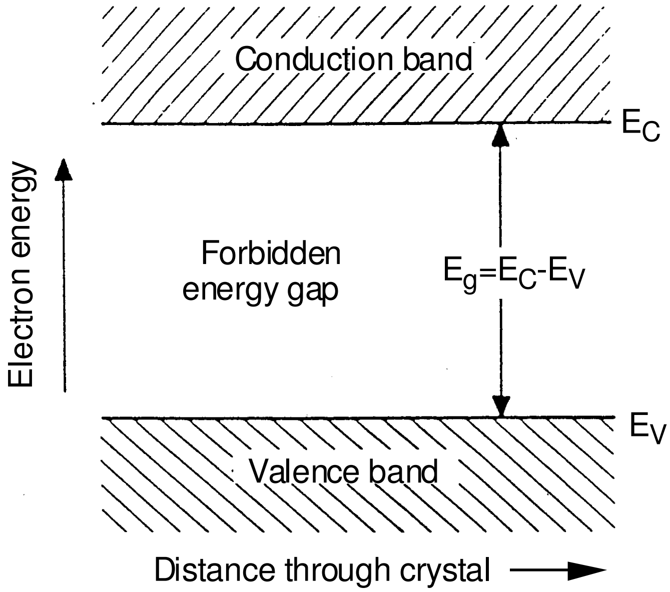

Figure 6-2. Schematic illustration of the energy band diagram for a pure semiconductor

Intrinsic ionisation

When the temperature is increased, electrons in the valence band are excited across the forbidden energy gap to the conduction band. This is the intrinsic ionisation. The electrons in the conduction band and the unoccupied electron sites in the valence band (electron holes) can move in an electric field. The electron holes behave as though they were positively charged and move in the opposite direction of the electrons. The intrinsic ionisation thus produces pairs of electron + electron hole charge carriers. When the electronic conductivity is due to intrinsic ionisation only, the semiconductor is called an intrinsic semiconductor. According to this model electron and electron hole conductivities increase with increasing concentrations of electrons in the conduction band and electron holes in the valence band.

In many oxides, and particularly those with large percentage ionic bonding, periodic fluctuations of the electric potential associated with each ion become too large (and energy bands too narrow) so that the band model provides an inadequate description or theory. In this case the electrons or holes may be considered to be localised at the lattice atoms. Such localised electronic defects are termed polarons and may from a chemical point of view be considered to constitute valence defects.

When classical statistics (the Boltzmann approximation) can be used it may be shown that the concentration of electrons and electron holes (expressed in number per cm^3) is given by

$$n = N_C \exp(-\frac{E_C-E_F}{kT}) \qquad (6.21)$$

$$p = N_V \exp(-\frac{E_F-E_V}{kT}) \qquad (6.22)$$

N_C and N_V represent the number of available states per unit volume - density of states - in the conduction and valence bands. E_C is the energy of the lowest level of the conduction band and E_V the highest level in the valence band. The parameter E_F is termed the Fermi level and we will return to this later on.

When the electrons occupy a narrow band of energies close to E_C, a parabolic relation between N_C and E_C may be assumed and it may be shown that N_C is then given by

$$N_C = (\frac{8\pi m_e^* kT}{h^2})^{3/2} \qquad (6.23)$$

where m_e^* is the effective mass of the electron. When the holes occupy a narrow region of energies close to E_V, N_V is correspondingly given by

$$N_V = (\frac{8\pi m_h^* kT}{h^2})^{3/2} \qquad (6.24)$$

where m_h^* is the effective mass of the electron hole. The effective masses of electrons and electron holes are seldom accurately known and as an approximation the free electron mass is then used in these relations.

The defect reaction and corresponding equilibrium between electrons in the conduction band and electron holes in the valence band can in terms of the law of mass action be expressed as

$$0 = e' + h^{\cdot} \qquad (6.25)$$

$$n \cdot p = K_i \qquad (6.26)$$

where K_i is the equilibrium constant for the intrinsic ionisation. Combination with Eqs. 6.21 and 6.22 gives

$$K_i = n \cdot p = N_C\, N_V \exp\left(-\frac{E_g}{kT}\right) \tag{6.27}$$

where E_g is the band gap, $E_g = E_C - E_V$. E_g may thus be considered the energy of the intrinsic ionisation. It is emphasised that Eq. 6.27 presupposes that classical statistics apply. It may also be noted that the Fermi level is eliminated in this expression of the law of mass action. It may be noted again that in Eq. 6.27 the concentrations of electrons and electron holes are expressed in terms of number per volume unit (e.g. per cm^3).

Intrinsic electronic semiconductor

In an intrinsic semiconductor the concentrations of electrons and electron holes are equal, and thus

$$n = p = K_i^{1/2} = (N_C\, N_V)^{1/2} \exp\left(-\frac{E_g}{2kT}\right) \tag{6.28}$$

and the electronic conductivity (Eq. 6.19) then becomes

$$\sigma_{el} = \sigma_n + \sigma_p = en\mu_n + ep\mu_p = e\,(N_C N_V)^{1/2}(u_n + u_p) \exp\left(-\frac{E_g}{2kT}\right) \tag{6.29}$$

It may be noted that N_C and N_V are temperature dependent and that also u_n and u_p may have various dependencies on temperature. If the latter are not exponential (as in diffusional hopping conduction processes) the exponential term of the energy gap tend to dominate the temperature dependence and as an approximation Eq. 6.29 is then often written

$$\sigma_{el} = \text{const.} \exp\left(-\frac{E_g}{2kT}\right) \tag{6.30}$$

From these relations it is evident that the intrinsic electronic conductivity increases with decreasing energy gap.

Effects of charged impurities or defects. Extrinsic semiconductors

As stated before, the valence and conduction bands may be considered to constitute the highest filled orbital and the lowest empty orbital, respectively. The valence band may in an oxide typically represent the 2p states of the O^{2-} ions, and removing an electron from here (creating a hole) may be regarded as the creation of an O^- ion, which is intuitively possible. The conduction band typically represents the reduced state of a metal constituent ion. For instance, in oxides with Ti^{4+} such as TiO_2 or $SrTiO_3$ the conduction band would typically represent the 4s state of the Ti^{3+} ion. The size of the band gap can be imagined to represent the difficulty of simultaneously oxidising O^{2-} to O^- and reducing Ti^{4+} to Ti^{3+}. In the chosen example we expect a moderate band gap. Similar oxides with Mn^{4+} would be more easily reducible and have smaller band gaps, while oxides with Si^{4+} or Zr^{4+} are less easily reducible and have larger band gaps.

Let us now consider the addition of small amounts of imperfections or impurities. These may be considered to contribute additional localised energy levels in the crystal. It is commonly assumed that these always fall within the forbidden gap. As we shall see later on this leads to situations where a certain temperature is needed to excite these defects to become effectively charged. This is not in accordance with experimental evidence and is not intuitive for imperfections with very stable aliovalent valence states. Thus, we will first consider cases where imperfections introduce levels outside the forbidden gap.

If these levels fall below the valence band edge E_V they will always be occupied (even at 0 K) and be in a reduced state. We refer to it as an acceptor imperfection. For instance, Al^{4+} substituting Ti^{4+} in the abovementioned oxides may always be considered to be reduced to Al^{3+} and as such be a charged defect even at 0 K, for instance taking the electron from the valence band. We may draw this conclusion from the higher stability of O^- as compared to Al^{4+}.

Similarly, an imperfection introducing a level above the conduction band edge E_C may always be considered to lose its electron and be oxidised even at 0 K. We refer to it as a donor. For instance, La^{2+} substituting Sr^{2+} and possibly Ta^{4+} substituting Ti^{4+} would be oxidised to La^{3+} and Ta^{5+} by giving off the electron to the easier formed Ti^{3+} of the conduction band.

The above examples lead to compensating electronic defects even at 0 K, but real cases may comprise simultaneous formation of point defects that may annihilate the electronic defects, (depending on oxygen activity). Thus, while Ca substituting La in $LaCrO_3$ is compensated by electron holes at high oxygen partial pressures, Al or Ta substituting Ti in $SrTiO_3$ will be compensated by oxygen vacancies or metal vacancies, respectively, under the same conditions. Of course, this brings up a question of whether native point defects such as vacancies at different stages of ionisation themselves introduce levels within or outside the band gap. If aliovalent dopants charged at 0 K are to be compensated by point defects at 0 K, these point defects must themselves be charged at 0 K and thus have levels outside the band gap.

In the following we consider imperfections introducing levels inside the forbidden energy gap as illustrated in Fig.6.3. When the imperfection introduces an energy level which is located below the lower edge of the conduction band (E_C), electrons can be thermally excited to the conduction band. As such the imperfection donates an electron, it is called a donor and the corresponding energy level a donor level. Under conditions where such imperfections dominate the defect structure the oxide becomes an n-conductor. Correspondingly, imperfections with energy levels just above the upper edge of the valence band (E_V) are termed acceptors; electrons in the valence band may be excited to the energy level of the imperfection (acceptor level) and the oxide may become a p-conductor.

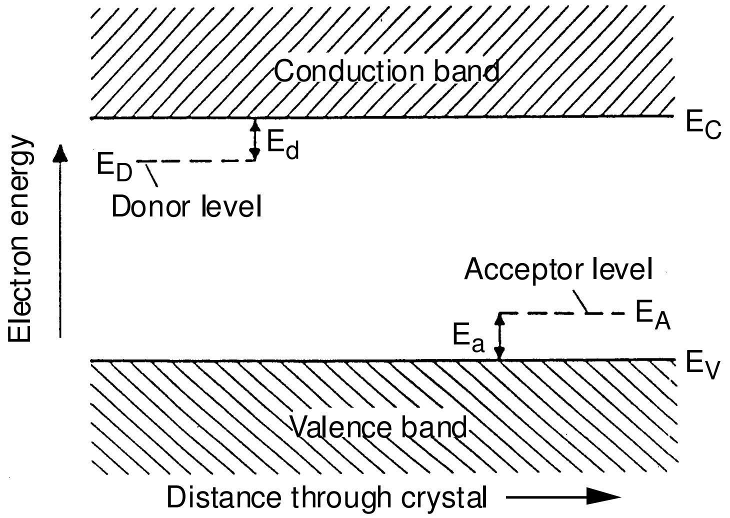

Figure 6-3. Schematic illustration of additionally localised energy levels due to donors and acceptors in the forbidden energy gap in the energy band diagram of a semiconductor.

In $SrTiO_3$ it is for instance found that Fe^{4+} substituting Ti^{4+} is an acceptor inside the gap; Fe^{3+} (the result after acceptance of an electron) is less favourable than Fe^{4+} (when the electron is taken from O^{2-}) and at a higher energy than the valence band of the O^{2-} electrons, but Fe^{3+} is clearly more favourable than Ti^{3+} and thus way below the conductance band. Similarly, oxygen vacancies are considered to be donors in the gap, and this can be viewed as a trapping of electrons at Ti^{4+} (as Ti^{3+}) near the vacancy.

While donors and acceptors inside the forbidden gap have positive energies of ionisation, the donors and acceptors outside the gap may be considered simply to have negative energies of ionisation.

Effects of donors

Let us again describe these processes in terms of the law of mass action. Thus the ionisation of a donor D^x may be written

$$D^x = D^{\cdot} + e' \qquad (6.31)$$

and the corresponding equilibrium by

$$\frac{[D^{\cdot}]\, n}{[D^x]} = K_D \qquad (6.32)$$

If the total number of donors is N_D, then

$$N_D = [D^{\cdot}] + [D^x] \qquad (6.33)$$

If the energy of the donor state is E_D, and the donor state is mainly empty ($[D^X] << N_D$) then the concentration of electrons in the donor state, $[D^X]$, may following Boltzmann statistics (be expressed by

$$[D^X] = N_D \exp\left(-\frac{(E_D - E_F)}{kT}\right) \qquad (6.34)$$

Since we assume $\frac{[D^X]}{N_D} << 1$, and that Boltzmann statistics also apply to conduction band electrons and by further combining Eqs. 6.21 and 6.32-6.34, we get an expression for K_D:

$$K_D = N_C \exp\left(-\frac{E_C - E_D}{kT}\right) = N_C \exp\left(-\frac{E_d}{kT}\right) \qquad (6.35)$$

where $E_d = E_C - E_D$ represents the ionisation energy of the donor (cf. Fig. 6.3). E_d may in Eq. 6.35 be considered to be the enthalpy of the ionisation of the donor.

When no other imperfections are present, the situation described here can be approximated by

$$n = [D^{\cdot}] \approx N_D \qquad (6.36)$$

and if the donors are present in an invariable amount as a fully soluble or frozen-in impurity, then $n = [D^{\cdot}] \approx N_D$ = constant. From Eq. 6.32, we then get the minority concentration of neutral unionised donors: $[D^x] = N_D^2 N_C^{-1} \exp(+E_d/kT)$.

If, on the other hand the donor level is mainly unionised, then $[D^{\cdot}] << [D^x] \approx N_D$ and Boltzmann statistics give

$$[D^{\cdot}] = N_D \exp(-\frac{(E_F - E_D)}{kT}) \tag{6.37}$$

and then, by combination with Eqs. 6.21, 6.32, and 6.33 we again obtain $K_D = N_C \exp(-E_d/kT)$ as in Eq. 6.35.

If the electroneutrality condition in the oxide is still given by

$$n = [D^{\cdot}] \tag{6.38}$$

(but now $[D^{\cdot}] << N_D$), then the concentration of electrons is, from insertion in Eq. 6.32:

$$n = [D^{\cdot}] = (K_D N_D)^{1/2} = (N_C N_D)^{1/2} \exp(-\frac{E_d}{2kT}) \tag{6.39}$$

The reader may find it useful to note the similarity between the form of this expression and Eq. 6.28. However, we also remind ourselves that N_D may be a constant concentration of impurities or it may be a constant or varying level of other point defects such as anion vacancies or cation interstitials.

In the above we have seen that the same equilibrium constant K_D for Eq. 6.31 applies whether the donor level is approximately empty (fully ionised) or approximately fully occupied (practically unionised, or, in this case, neutral). But our treatment does not hold for intermediate situations, i.e. what we may call partly filled donor levels. In the latter case Boltzmann statistics do not represent a sufficiently good approximation, and instead Fermi-Dirac statistics must be used.

Let us recapitulate briefly the treatment of the ionisation of a constant concentration of a donor in the forbidden gap of an otherwise pure, stoichiometric, ideal semiconductor: At low temperatures, the concentration of electrons is given by a minor degree of ionisation of the donors, as given by Eq. 6.39. If we neglect the temperature dependence of N_C, the situation can be illustrated as in the right-hand part of Fig. 6.4; the concentration of electrons increases with an apparent enthalpy of $E_d/2$. At a sufficiently high temperature (middle part of Fig. 6.4) all donors are ionised, and the concentration of electrons becomes constant (Eq. 6.36). At even higher temperatures intrinsic semiconduction may predominate, i.e., n = p, and in principle the temperature dependence becomes as illustrated in the left hand part of Fig. 6.4, with an apparent enthalpy close to $E_g/2$ (cf. Eq. 6.25). Thus at low temperatures this oxide is an n-conductor due to the ionisation of the donors and at high temperatures intrinsic ionisation predominates. The behaviour over the entire temperature range could in principle be solved from the full electroneutrality equation

$$n = [D^{\cdot}] + p \tag{6.40}$$

in combination with the constancy of the donor concentration and the expressions for the equilibrium constants K_i (Eq. 6.27) and K_D (Eq. 6.35). However, as noted above, the expression for K_D would not be valid between the intermediate and low temperature (right hand side) domains in Fig. 6.4.

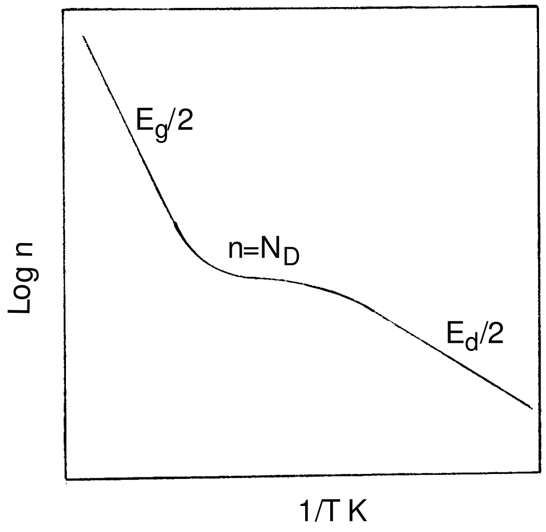

Figure 6-4. Schematic illustration of the logarithm of the concentration of defect electrons as a function of the reciprocal absolute temperature for a semiconductor with donors.

Effects of acceptors

A corresponding treatment may be made for ionisation of acceptors, which in terms of a defect reaction may be written

$$A^x = A' + h^{\cdot} \tag{6.41}$$

The equilibrium constant for the defect reaction is given by

$$\frac{[A'] p}{[A^x]} = K_A \tag{6.42}$$

Following a similar treatment as for the donor, K_A may under the assumption that Boltzmann statistics apply to the valence band (small concentration of holes) and the acceptor (nearly full or nearly empty acceptor levels) be expressed by

$$K_A = N_V \exp\left(-\frac{E_A - E_V}{kT}\right) = N_V \exp\left(-\frac{E_a}{kT}\right) \tag{6.43}$$

where E_A is the energy level of the acceptor and $E_a = E_A - E_V$ is called the ionisation energy of the acceptor (cf. Fig. 6.3).

For an acceptor doped oxide the temperature dependence of the concentration of holes will be qualitatively analogous to that of electrons in the donor doped case (cf. Fig. 6.4) with the same constrictions.

The Fermi level and chemical potential of electrons

In this chapter we have so far treated the defect equilibria for semiconductivity in terms of the band model. In this model one makes use of the parameter termed the Fermi level, E_F. As the defect equilibria are otherwise described by equilibrium thermodynamics, it is of interest to correlate the band model with the thermodynamic approach.

A chemical equilibrium implies that the chemical potential of a species is the same in all phases. As regards electrons in a system, this also means that their chemical potentials (or electrochemical potentials if the inner potential can not be neglected) must be equal, although they may have different energies. Thus the chemical potential of the electrons in general, μ_e, must be equal to the chemical potential of valence electrons, conduction electrons, etc.,

$$\mu_e = \mu(\text{cond. electrons}) = \mu(\text{valence electrons}) \tag{6.44}$$

The chemical potential of electrons, for instance, the conduction electrons, may be written in terms of the chemical potential in a standard state, $\mu°$(cond. electrons) and a term for the entropy of mixing:

$$\mu_e = \mu(\text{cond. electrons}) = \mu^\circ(\text{cond. electrons}) + kT \ln \frac{n}{N_C} \qquad (6.45)$$

From the relation between the concentration of conduction electrons, n, the conduction band energy level E_C and the Fermi level E_F (Eq. 6.18) we may write

$$E_F = E_C + kT \ln \frac{n}{N_C} \qquad (6.46)$$

By comparing Eqs. 6.45 and 6.46 it is seen that if E_C is considered to be the chemical potential in the standard state for conduction electrons, the Fermi level represents the chemical potential of the conduction electrons, and thus of all electrons in a substance.

Charge mobilities of electrons and electron holes

In the preceding chapters we have looked at temperature dependencies of concentrations of electronic defects and point defects, and we have looked at the conductivity and mobility of thermally activated diffusing species. In the following we consider the charge carrier mobilities of electrons and holes in some more detail. For instance for an intrinsic electronic semiconductor (where n=p) we can from Eq. 6.29 in combination with Eqs. 6.23 and 6.24 write an expression for σ_{el}:

$$\sigma_{el} = \{2e(\frac{2\pi k}{h^2})^{3/2} (m_e^* \cdot m_h^*)^{3/4} T^{3/2} \exp(-\frac{E_g}{2kT})\}(u_n + u_p) \qquad (6.47)$$

The effective mass of electrons and electron holes can be interpreted by a quantum mechanical treatment of electronic motion of electrons and electron holes in solids. The effective mass differs from the real mass of electrons due to the interaction of electrons with the periodic lattice of the atoms. Only for a completely free electron is the mass equal to the real mass, $m_e = m_e^*$. As mentioned above, the values of the effective masses are not accurately known, and the value of m_e is often used as an approximation.

In order to obtain an accurate description of the temperature dependence of the electronic conductivity it is necessary to consider the temperature dependencies of the charge carrier mobilities.

Non-polar solids – itinerant electron model

The temperature dependence of the charge carrier mobility is dependent on the electronic structure of the solid. For a pure non-polar semiconductor - as in an ideal and pure covalent semiconductor - the electrons in the conduction band and the electron holes in the valence band can be considered as quasi-free (itinerant) particles. Then the mobilities of electrons and electron holes, u_n and u_p, are determined by the thermal vibrations of the lattice in that the lattice vibrations result in electron and electron hole scattering (lattice scattering). Under these conditions the charge carrier mobilities of electrons and electron holes are both proportional to $T^{-3/2}$, e.g.

$$u_{n,latt} = \text{const. } T^{-3/2} \tag{6.48}$$

In this case the temperature dependence of σ_{el} (Eq. 6.47) becomes

$$\sigma_{el} = \text{const. } T^{3/2} \cdot T^{-3/2} \exp\left(-\frac{E_g}{2kT}\right) = \text{const. } \exp\left(-\frac{E_g}{2kT}\right) \tag{6.49}$$

If, on the other hand, the scattering is mainly due to irregularities caused by impurities or other imperfections, the charge carrier mobility is proportional to $T^{3/2}$, e.g.

$$u_{n,imp} = \text{const. } T^{3/2} \tag{6.50}$$

If both mechanisms are operative, the mobility is given by

$$u = \frac{1}{\frac{1}{u_{latt}} + \frac{1}{u_{imp}}} \tag{6.51}$$

and from the temperature dependencies given above it is evident that impurity scattering dominates at low temperature while lattice scattering takes over at higher temperature.

Polar (ionic) oxides

When electrons and electron holes move through polar compounds, such as ionic oxides, they polarise the neighbouring lattice and thereby cause a local deformation of the structure. Such an electron or electron hole with the local deformation is termed a polaron. The polaron is considered as a fictitious single particle.

When the interaction between the electron or electron hole and the lattice is relatively weak, the polaron is referred to as a large polaron. Large polarons behave much like free carriers except for an increased mass caused by the fact that polarons carry their associate deformations. Large polarons still move in bands, and the expressions for the effective density of states in the valence and conduction bands are valid. The temperature dependence of the mobilities of large polarons at high temperatures[*] is given by

$$u_{\text{large pol.}} = \text{const.}\ T^{-1/2} \qquad (6.52)$$

For an intrinsic semiconducting oxide where the electronic conductivity mechanism can be described in terms of a large polaron mechanism, the temperature dependence of σ_{el} can be written by combination of Eqs. 6.47 and 6.52

$$\sigma_{el} = \text{const.}\ T \cdot \exp\left(-\frac{E_g}{2kT}\right) \qquad (6.53)$$

Thus in this case the value of the band gap can be deduced by plotting $\log(\sigma_{el}/T)$ vs 1/T.

The large polaron mechanism has been suggested for highly ionic non-transition metal oxides. Such oxides have large energy gaps ($E_g \geq 6$ eV), where the band structure is characterised by large band widths. Due to the large band gap one expects electronic semiconductivity to be low in these oxides, and to predominate possibly only at high temperatures.

For other oxides it has been suggested that the interactions between the electronic defects and the surrounding lattice can be relatively strong and more localised. If the dimension of the polaron is smaller than the lattice parameter, it is called a small polaron or localised polaron, and the corresponding electronic conduction mechanism is called a small polaron mechanism.

The transport of small polarons in an ionic solid may take place by two different mechanisms. At low temperatures small polarons may tunnel between localised sites in what is referred to as a narrow band. The temperature dependence of the mobility is determined by lattice scattering and the polaron mobility decreases with increasing temperature in a manner analogous to a broad band semiconductor.

However, at high temperatures (for oxides above roughly 500 °C) the band theory provides an inadequate description of the electronic conduction mechanism. The energy levels of electrons and electron holes do not form bands, but are localised on specific atoms of the crystal structure (valence defects). It is

---

[*] "High temperatures" are temperatures above the optical Debye temperature, θ. For oxides $\theta \sim (h\omega)/2\pi k$, where h is the Planck constant, k the Boltzmann constant and ω the longitudinal optical frequency which for an oxide is $\sim 10^{14}\ s^{-1}$.

assumed that an electron or electron hole is self-trapped at a given lattice site, and that the electron (or electron hole) can only move to an adjacent site by an activated hopping process similar to that of ionic conduction. Consequently it has been suggested that the mobility of a small polaron can be described by a classical diffusion theory as described in a preceding chapter and that the Nernst -Einstein can be used to relate the activation energy of hopping, E_u, with the temperature dependence of the mobility, u, of an electron or electron hole:

$$u = \frac{e}{kT} D = \text{const.}\ T^{-1} \exp\left(- \frac{E_u}{kT} \right) \tag{6.54}$$

where E_u is the activation energy for the jump.

At high temperatures, the exponential temperature dependence of small polaron mobilities can thus in principle be used to distinguish it from the other mechanisms.

The different mechanisms can also be roughly classified according to the magnitude of the mobilities; the lattice and impurity scattering mobilities of metals and non-polar solids are higher than large-polaron mobilities which in turn are larger than small-polaron mobilities. Large polaron mobilities are generally of the order of 1-10 $cm^2/V^{-1}s^{-1}$, and it can be shown that a lower limit is approximately 0.5 $cm^2V^{-1}s^{-1}$. Small polaron mobilities generally have values in the range 10^{-4}-10^{-2} $cm^2V^{-1}s^{-1}$. For small polarons in the regime of activated hopping the mobility increases with increasing temperature and the upper limit is reported to be approximately 0.1 $cm^2V^{-1}s^{-1}$.

Nonstoichiometric semiconductors

Corresponding expressions for σ_{el} for nonstoichiometric electronic semiconductors readily follows by considering the temperature and oxygen pressure dependence of the concentration of the electronic defects.

For nonstoichiometric oxides the concentration of electronic defects is determined by the deviation from stoichiometry, the presence of native charged point defects, aliovalent impurities and/or dopants. The concentration of electronic defects can be evaluated from proper defect structure models and equilibria. Various defect structure situations have been described in previous chapters and at this stage only one example - dealing with oxygen deficient oxides with doubly charged oxygen vacancies as the prevalent point defects - will be described to illustrate the electrical conductivity in nonstoichiometric oxides.

Oxygen deficient oxides

Let us recapitulate the equations for formation of doubly charged oxygen vacancies. As described in Chapter 3 the defect equation may be written

$$O_O = V_O^{2\cdot} + 2e' + \frac{1}{2} O_2 \qquad (6.55)$$

The corresponding defect equilibrium is given by

$$[V_O^{2\cdot}]\, n^2 = K_{V_O^{2\cdot}}\, p_{O_2}^{-1/2} \qquad (6.56)$$

If we deal with a high-purity oxide where the concentration of impurities can be ignored compared to the concentration of oxygen vacancies and electrons, the electroneutrality condition becomes

$$n = 2[V_O^{2\cdot}] \qquad (6.57)$$

By combining Eqs. 6.57 and 6.58 the concentration of electrons is given by

$$n = 2[V_O^{2\cdot}] = (2K_{V_O^{2\cdot}})^{1/3}\, p_{O_2}^{-1/6} \qquad (6.58)$$

The total electrical conductivity is given by the sum of the conductivity of the electrons and of the oxygen vacancies:

$$\sigma_t = 2\, e\, [V_O^{2\cdot}]\, u_{V_O^{2\cdot}} + e\, n\, u_n \qquad (6.59)$$

$2\, e\, [V_O^{2\cdot}]\, u_{V_O^{2\cdot}}$ represents the ionic conductivity due to the oxygen vacancies and where $u_{V_O^{2\cdot}}$ is the mobility of the oxygen vacancies. However, if the electrons and oxygen vacancies are the prevalent charge carriers, the contribution due to oxygen vacancies can be ignored due to the much higher mobility of electrons than oxygen vacancies, and the oxide is an n-conductor where the conductivity can then be written

$$\sigma_t = \sigma_n = e\, n\, u_n = e\, u_n\, (2K_{V_O^{2\cdot}})^{1/3}\, p_{O_2}^{-1/6} \qquad (6.60)$$

As described in previous chapters the equilibrium constant for the formation of doubly charged oxygen vacancies and 2 electrons is given by

$$K_{V_O^{2\cdot}} = \exp\left(\frac{\Delta S_{V_O^{2\cdot}}}{k}\right) \exp\left(-\frac{\Delta H_{V_O^{2\cdot}}}{kT}\right) \tag{6.61}$$

When one combines Eqs. 6.61 and 6.62 the n-conductivity may be written:

$$\sigma_t = \sigma_n = e\, u_n \exp\left(\frac{\Delta S_{V_O^{2\cdot}}}{3k}\right) \exp\left(-\frac{\Delta H_{V_O^{2\cdot}}}{3kT}\right) p_{O_2}^{-1/6} \tag{6.62}$$

Let us further assume that the electrons are small polarons and thus that the mobility of the electrons are given by Eq. 6.54. The conductivity can then be expressed by

$$\sigma_n = \text{const.}\, \frac{1}{T} \exp\left(-\frac{\Delta H_{V_O^{2\cdot}}/3 + E_u}{kT}\right) p_{O_2}^{-1/6} \tag{6.63}$$

Thus following this equation the n-conductivity is proportional to $p_{O_2}^{-1/6}$, and if this defect structure situation prevails over a temperature range from T_1 to T_5, one will obtain a set of isotherms of the n-conductivity as shown in Fig.6.5.

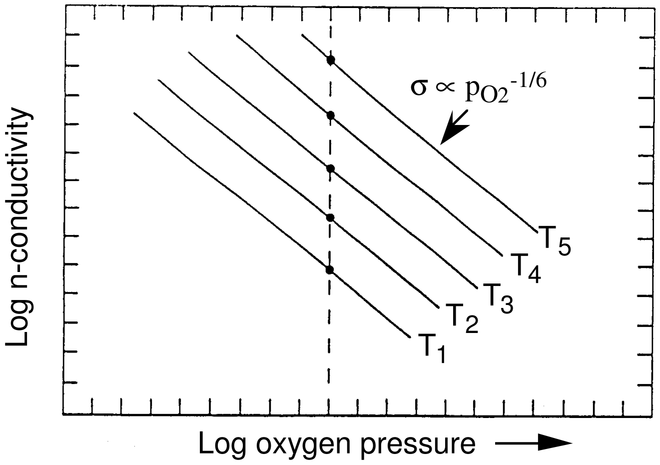

Figure 6-5. Schematic presentation of different isotherms of the n-conductivity at temperatures from T_1 to T_5 for an oxygen deficient oxide where the predominant defects are doubly charged oxygen vacancies and electrons.

Furthermore, if it can be assumed that mobility of the charge carriers (defect electrons) is independent of the defect concentration, then a plot of the values of $\log_{10}(\sigma T)$ at a constant oxygen pressure yields a straight-line relationship as illustrated in Fig. 6.6. The slope of the line is given by $-\frac{1}{2.303k}\frac{\Delta H_{V_O^{2.}}}{3} + E_u$, where the factor 2.303 is the conversion factor in changing from $\ln_e$ to $\log_{10}$. The activation energy is given by the term

$$E_\sigma = \frac{\Delta H_{V_O^{2.}}}{3} + E_u. \qquad (6.64)$$

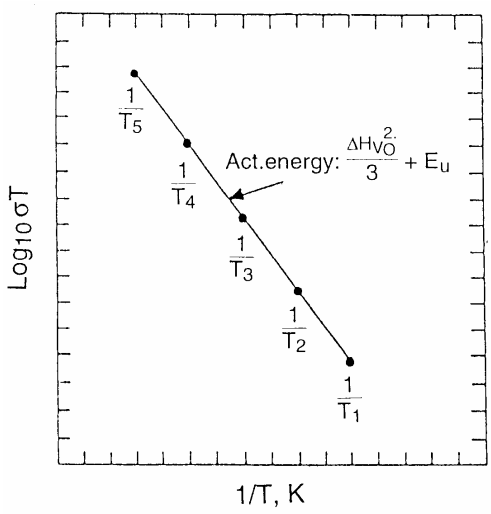

Figure 6-6. Schematic illustration of a plot of $\log_{10}(\sigma_n T)$ *vs. the reciprocal absolute temperature at constant oxygen pressure (cf. Fig.6.5). The slope of the line is given by* $-\frac{1}{2.303k}\frac{\Delta H_{V_O^{2.}}}{3} + E_u$ *where* $\frac{\Delta H_{V_O^{2.}}}{3} + E_u$ *is the activation energy for the n-conductivity.*

In general the temperature dependence of the charge carrier mobility of the electrons is much smaller than the enthalpy term associated with the formation of doubly charged oxygen vacancies.

The mobility of electronic charge carriers may be determined by measuring the electrical conductivity and combine these measurements with independent measurements of the concentration of the electronic charge carriers. The concentration of the charge carriers may be estimated from measurements of the Seebeck coefficient or by measurements of the nonstoichiometry combined with the proper description of the defect structure (cf. Ch. 7).

For mixed conductors that exhibit both ionic and electronic conductivities it is necessary to delineate the ionic and electronic contributions. A commonly used technique for this is the emf method originally derived by Wagner. This will be described in the next chapter (Ch. 7) dealing with electrochemical transport in metal oxides.

Ionic conductivity

Ionic conductivity follows the Nernst–Einstein relationship, for hopping species, derived early in this chapter:

$$\sigma_i = z_i e\, c_i\, u_i = (z_i e)^2 c_i D_i / kT \qquad (6.65)$$

It may be noted that the species considered may be a defect (e.g. oxygen vacancies) or a constituent (oxide ions). For defects the mobility and diffusivity are large and roughly constant, while the concentration is small and variable. For the constituent, the mobility and diffusivity are small and variable (with defect concentration) while the concentration is large and roughly constant. The conductivity (here oxide ion conductivity by the vacancy mechanism) is the same.

In many cases of utilizing ionic conduction, the concentration c_i of ionic defects is constant, given by a dopant. For instance, ionic conductivity in yttria-doped zirconia is determined by the concentration of oxide ion vacancies, in turn given as charge compensating the yttrium acceptors, $2[v_O^{\cdot\cdot}] = [Y_{Zr}^{/}]$. If the concentration of acceptors is given in mole-fraction, then it is necessary to multiply the resulting mole fraction of vacancies by the formula density or molar density of the compound in order to obtain the volume density required for insertion in Eq. 6.65:

$$\sigma_{vO..} = 2e\, [Y_{Zr}^{/}]/2 * c_{ZrO2}\, u_{vO..} = 2F\, [Y_{Zr}^{/}]/2 * C_{ZrO2}\, u_{vO..} \qquad (6.66)$$

where c_{ZrO2} and C_{ZrO2} are, respectively, the molecular and molar densities (number of formula units or moles per unit volume) of the oxide.

There are of course also cases – also of practical interest – where the concentration of ionic defects vary, e.g. with temperature in intrinsically disordered compounds, and with temperature and non-stoichiometry in non-stoichiometric compounds. In proton conducting oxides the proton conductivity varies with proton concentration, typically a function of water vapour partial pressure.

We leave further learning about ionic conduction to exercises and Chapter 7 (electrochemical transport), and here only briefly mention a couple of aspects of ionic transport that relates it in more detail to diffusion.

Correlation effects: tracer diffusion and ionic conduction

In the discussions of diffusion mechanisms in Chapter 5 it was pointed out that successive jumps of tracers atoms in a solid may for some mechanisms not be completely random, but are to some extent correlated. This is, for instance, the case for the vacancy and interstitialcy mechanisms. For a correlated diffusion of a tracer atom in a cubic crystal the tracer diffusion coefficient, D_t, is related to the random diffusion coefficient for the atoms, D_r, through the correlation coefficient f:

$$D_t = f\, D_r \qquad (6.67)$$

The value of f is governed by the crystal structure and the diffusion mechanism.

Ionic conductivity method

Values of the correlation coefficient may be determined by comparing the measured values of the ionic conductivity and the tracer diffusion coefficient. Thus the use of the Nernst-Einstein relation gives the following expression for the correlation coefficient:

$$f = \frac{D_t}{D_r} = \frac{D_t}{\sigma_i} \frac{c_i\,(z_i e)^2}{kT} \qquad (6.68)$$

This equation is applicable to any diffusion process for which the atom jump distance is equal to the displacement of the effective charge, e.g. for vacancy and interstitial diffusion.

However, in interstitialcy diffusion the charge displacement is larger than the atom jump distance, and a displacement factor S must be included in the Nernst-Einstein relation. In collinear interstitialcy diffusion (Fig. 5.9) the effective charge is, for instance, moved a distance twice that of the tracer atom and D_t/D_r is given by

$$\frac{D_t}{D_r} = \frac{D_t}{S} \frac{c_i\,(z_i e)^2}{\sigma_i kT} = \frac{f}{S} \qquad \text{(collinear)} \qquad (6.69)$$

where S = 2. For a collinear jump in an fcc structure the displacement factor is 4/3.

Studies on alkali and silver halides have provided illustrative, and by now classical examples of the applicability of the ionic conductivity method for determining the correlation factor and detailed aspects of the jumps in diffusion processes. NaCl, for instance, is essentially a pure cationic conductor. Measured ratios of D_t/D_r are in good agreement with the assumption that f = 0.78, i.e. that the Na-ions diffuse by a vacancy mechanism.

However, such a simple relationship was not found for AgBr. AgBr is also a cationic conductor and comparative values of D_t (diffusion of Ag in AgBr) and of values of D_r evaluated from conductivity measurements are shown in Fig. 6.7.

From studies of the effect of Cd-dopants on the ionic conductivity it could be concluded that cationic Frenkel defects predominate in AgBr. Thus the diffusion was therefore expected to involve both vacancy diffusion and transport of interstitial ions. The experimentally measured ratios of D_t/D_r varied from 0.46 at 150 °C to 0.67 at 350°C. For vacancy diffusion a constant ratio of 0.78 (=f) would have been expected, and the diffusion mechanism could thus be ruled out. For interstitial diffusion f=1, and this mechanism could also be excluded.

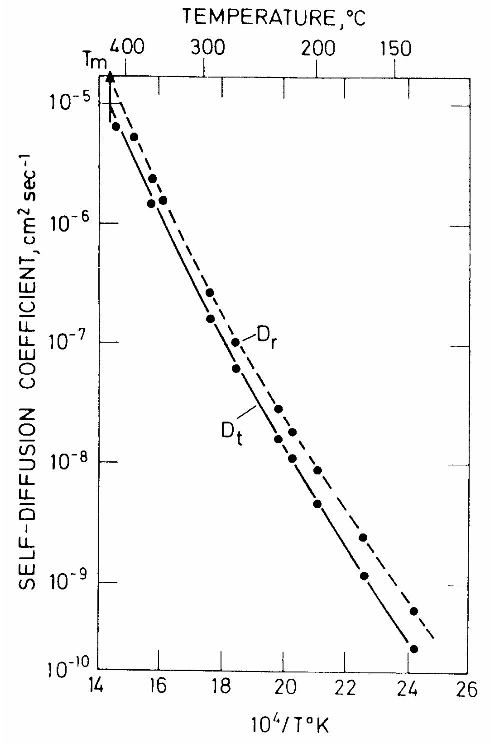

Figure 6-7. Values of D_t and of D_r evaluated from conductivity measurements for diffusion of Ag in AgBr. Results after Friauf(1957,1962).

For interstitialcy diffusion of Ag in AgBr the value of f equals 2/3 for a collinear jump and 0.97 for a non-collinear jump. Following Eq.6.64 one would thus expect that D_t/D_r would range from 0.33 for a collinear jumps to 0.728 for non-collinear jumps. On this basis Friauf (1957, 1962) concluded that the interstitialcy diffusion is the important mechanism in AgBr and that collinear jumps are most important at low temperatures while non-collinear jumps become increasingly important the higher the temperature.

Simultaneous diffusion and electric field

The ionic conductivity and D_t may in principle be studied in a single experiment, as described by Manning (1962) and others. If a thin layer of the isotopes is sandwiched between two crystals and the diffusion anneal is performed while applying the electric field, the tracer distribution profile is displaced a distance $\Delta x = u_iEt$ relative to the profile in the absence of the applied field (Eq. 5.31). The resultant tracer distribution is given by

$$c = \frac{c_o}{2(\pi D_t t)^{1/2}} \exp\left(-\frac{(x - \Delta x)^2}{4D_t t}\right) \qquad (6.70)$$

The maximum in the concentration profile is - as illustrated in Fig. 6.8 - displaced a distance Δx, and u_i and D_t may be determined from the same experiment. If the crystal is a mixed ionic/electronic conductor, the value of the ionic transport number under the experimental conditions must be known.

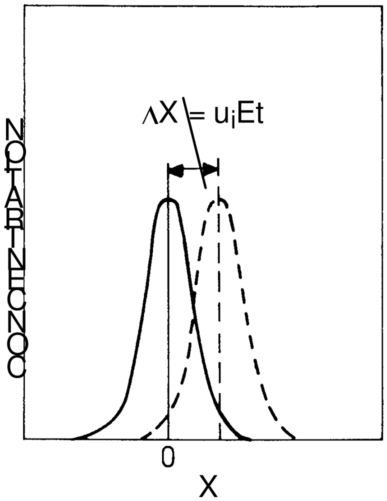

Figure 6-8. Schematic illustration of the concentration profile of a radioactive tracer when an electric field is applied during the diffusion anneal. The tracers are originally located at 0, but the concentration profile is displaced a distance $\Delta X = u_iEt$.

Literature

Friauf, R.T. (1957) *Phys. Rev.* **105**, 843; (1962) *J. Appl. Phys.* **33** suppl., 494.

Manning, J.R. (1962) *J. Appl. Phys* **33,** 2145; *Phys. Rev.* **125**, 103.

Exercises

1. Cobalt oxide: The electronic conductivity of $Co_{1-y}O$ at 1350°C and p_{O2} = 0.1 atm is 25 S/cm. Thermogravimetric measurements show that y = 0.008 under the same conditions. It is assumed that singly charged cobalt vacancies are the dominating point defects. Identify the charge carriers responsible for the conductivity and calculate their charge mobility. (Assume that the density of CoO at 1350°C equals that at room temperature, 6.4 g/cm^3. Atomic weights M_{Co} = 58.93, M_O = 16.00.)

2. Nickel oxide: Assume that doubly charged nickel vacancies and electron holes are the dominating defects in $Ni_{1-y}O$ under oxidising conditions. At 1245°C and p_{O2} = 1 atm we know the following for the compound:

The self diffusion coefficient for nickel: $D_{Ni} = 9*10^{-11}$ cm^2/s

Electrical conductivity: $\sigma = 1.4$ S/cm

(Data from M.L. Volpe and J. Reddy, *J. Chem. Phys.*, **53** (1970) 1117.)

Nickel vacancy concentration, in site or mole fraction: $[v_{Ni}'']=2.5*10^{-4}$

(Data from W.C. Tripp and N.M. Tallan, *J. Am. Ceram. Soc.,* **53** (1970) 531.)

i) Calculate the concentration of electron holes under the given conditions, given as site fraction and as volume concentration (e.g. number/cm^3). (Atomic weights M_{Ni} = 58.71, M_O = 16.00, density of NiO = 6.67 g/cm^3.)

ii) Calculate the charge mobility of the electron holes.

iii) Calculate the diffusion coefficient of nickel vacancies.

iv) Calculate the charge mobility of the nickel vacancies and the ionic conductivity under the conditions referred to above.

3. Ca-stabilised ZrO_2 (CSZ)

We shall here consider a densely sintered ZrO_2 doped with 15 mol% CaO ($Zr_{0.85}Ca_{0.15}O_{1.85}$).

i) Assume that the oxide contains doubly charged oxygen vacancies compensating the Ca dopant. What are the site-fractions of dopants and of oxygen vacancies?

ii) Derive equations showing how the minority concentrations of defect electrons and electron holes vary with p_{O2} in this oxide under the given conditions.

iii) The conductivity of this oxide is independent of p_{O2} from oxidising to very reducing conditions. What can we deduce from this?

iv) Simpson and Carter (*J. Am. Ceram. Soc.* **49** (1966) 139) measured the self diffusion coefficient for oxygen in $Zr_{0.85}Ca_{0.15}O_{1.85}$ and found it to be $D_O = 2.0*10^{-7}$ cm^2/s at 1100°C. Calculate the electrical conductivity based on this.

v) Find also the diffusion coeffeicient and charge mobility for the oxygen vacancies.

7. Electrochemical transport

Electrochemical potential

In the chapter on diffusion we learned that random diffusion is driven by thermal energy and that it in crystalline solids requires defects. Moreover, we learned that a net flux of a species can result from a gradient in its chemical potential. In the chapter on electrical conductivity we furthermore saw that a net flux of charged particles would result in a gradient of electrical potential.

In ionic media (materials, liquids, and solutions) the chemical and electrical potentials act simultaneously, and it is often convenient to combine them into an electrochemical potential. For the species *i* we have

$$\eta_i = \mu_i + z_i e\phi \tag{7.1}$$

This is true whether the species is real or a defect, but in general we shall onwards deal mainly with real species and real charges z_i.

The electrochemical potential gradient is accordingly, in the one-dimensional case,

$$\frac{d\eta_i}{dx} = \frac{d\mu_i}{dx} + z_i e\frac{d\phi}{dx} \tag{7.2}$$

The combination of chemical and electrical potentials and potential gradients forms the basis for the treatment of all mass transport processes involving charged species (ions) in ionic solids, and is the theme of this chapter. The theory is termed Wagner-type after Carl Wagner, who was the first to derive it, originally to describe oxidation of metals.

Flux equations – Wagner theory

General derivation of Wagner-type flux expressions

We have seen in a preceding paragraph that a force, expressed as gradient in a potential P_i acting on a species *i*, gives rise to a flux density of that species which is proportional to its self-diffusion coefficient D_i. By assuming that the potential acting is the electrochemical potential, we obtain

$$j_i = \frac{-D_i c_i}{kT}\frac{dP_i}{dx} = \frac{-D_i c_i}{kT}\frac{d\eta_i}{dx} = \frac{-D_i c_i}{kT}[\frac{d\mu_i}{dx} + z_i e\frac{d\phi}{dx}] \tag{7.3}$$

Via the Nernst-Einstein equation we may substitute conductivitiy for the random diffusivity and obtain the following alternative expression for the flux density:

$$j_i = \frac{-\sigma_i}{(z_i e)^2}[\frac{d\mu_i}{dx} + z_i e\frac{d\phi}{dx}] \tag{7.4}$$

If the species is charged, the flux density for *i* gives rise to a partial current density i_i:

$$i_i = z_i e\, j_i = -\frac{\sigma_i}{z_i e}[\frac{d\mu_i}{dx} + z_i e\frac{d\phi}{dx}] \tag{7.5}$$

The net current density in the sample is obtained by summing the partial current densities over all the species *k*:

$$i_{tot} = \sum_k z_k e\, j_k = -\sum_k \frac{\sigma_k}{z_k e}[\frac{d\mu_k}{dx} + z_k e\frac{d\phi}{dx}] \tag{7.6}$$

The sample is next assumed to be connected to an external electric circuit. This may be real, with electrodes and wires and an electrical supply or load, or it may be absent, in which case we say that we have an open circuit. Since the sample makes a series connection with the external circuit, the currect in the sample must be the same as in the external circuit.

For a bare sample, a gas permeation membrane, or for an open-circuit fuel cell or other electrochemical device, the total current is zero. For fuel cells or electrolysers in operation, on the other hand, the current is non-zero.

By using the definition of total conductivity, $\sigma_{tot} = \sum_k \sigma_k$, and the definition of transport number, $t_k = \frac{\sigma_k}{\sigma_{tot}} = \frac{\sigma_k}{\sum_k \sigma_k}$, we obtain the following very important expression:

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} - \sum_k \frac{t_k}{z_k e}\frac{d\mu_k}{dx} \tag{7.7}$$

This relates the electrical potential gradient to the total (net, external) current density, the total conductivity, and the transport number and chemical potential gradient of all charge carriers.

From charged to neutral species: the electrochemical reaction

The chemical potentials of charged species are not well-defined, and we need to represent them instead by chemical potentials of neutral species. For this purpose we may assume equilibria between neutral and charged species and electrons, i.e. in the electrochemical red-ox- reaction

$$S = S^{z} + ze^{-} \tag{7.8}$$

where S is a neutral chemical entity and z may be positive or negative. The equilibrium condition for this is expressed in terms of the chemical potentials of products and reactants:

$$d\mu_{S^z} + zd\mu_{e^-} - d\mu_S = 0 \tag{7.9}$$

and can be rearranged with respect to the ionic species:

$$d\mu_{S^z} = d\mu_S - zd\mu_{e^-} \tag{7.10}$$

We insert this for all ionic species *n* in the expression for the electrical potential gradient. The entry (among k) for electrons is left unsubstituted. By using $\sum_k t_k = 1$ we obtain

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} - \sum_n \frac{t_n}{z_n e}\frac{d\mu_n}{dx} + \frac{1}{e}\frac{d\mu_e}{dx} \tag{7.11}$$

for which the chemical potentials now refer to the neutral forms of each carrier.

The voltage over a sample

We now integrate the electrical potential gradient over the thickness of the sample, from side 1 to side 2, in order to obtain the voltage over the sample:

$$\int_I^{II} d\phi = -\int_I^{II} \frac{i_{tot}}{\sigma_{tot}} dx - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n + \int_I^{II} \frac{1}{e} d\mu_e \qquad (7.12)$$

$$\phi_{II} - \phi_I = -\int_I^{II} \frac{i_{tot}}{\sigma_{tot}} dx - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n + \frac{1}{e}(\mu_{e,II} - \mu_{e,I}) \qquad (7.13)$$

We further assume that the voltage is measured on each side using the same inert metal, e.g. Pt. This eliminates the difference between chemical potentials of electrons on the two sides, and the voltage measured between the two sides is

$$U_{II-I} = \phi_{II} - \phi_I = -\int_I^{II} \frac{i_{tot}}{\sigma_{tot}} dx - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n \qquad (7.14)$$

Under open circuit conditions, $i_{tot}=0$, and we obtain

$$U_{II-I} = -\sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n \qquad (7.15)$$

We shall later see how this is used to calculate transport numbers based on open circuit voltage measurements of cells exposed to a well-defined gradient in chemical activities. Alternatively, if t_n is known to be unity for ionic charge carriers this expression will yield the open circuit voltage of a fuel cell or a galvanic sensor.

If current is drawn from the sample, as in a fuel cell or battery under load, we need to know how i_{tot} varies with x. If we assume that it is constant, we get

$$U_{II-I} = -\frac{i_{tot} X}{\sigma_{tot}} - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n = -i_{tot} r_{tot} - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n \qquad (7.16)$$

where X is the thickness of the sample and r_{tot} is the area-specific resistance of the sample. Alternatively, $i_{tot}r_{tot}$ (current density and area specific resistance) may be replaced by RI, (sample resistance and current):

$$U_{II-I} = -IR - \sum_n \int_I^{II} \frac{t_n}{z_n e} d\mu_n \tag{7.17}$$

The voltage of a fuel cell or battery is thus composed of a thermodynamic part that arises from the chemical gradient, and the well-known “IR”-term which arises from the limiting kinetics (transport) in the sample. The letters “IR” may be taken to reflect current times resistance, or to mean “internal resistance”. We will later see that the direction of the current ends up such that the IR-term reduces the voltage over the fuel cell or battery.

Flux of a particular species

One of the general expressions for the electrical potential gradient can now be inserted in an expression for the flux density of an individual species (Eq. 7.4) or the corresponding partial current density (Eq. 7.5). Since these two sum over chemical potential gradients of charged species, we may conveniently use Eq. 7.7 for our purpose. Inserting this into Eq. 7.4 and rearranging yields

$$j_i = \frac{t_i\, i_{tot}}{z_i e} - \frac{\sigma_i}{(z_i e)^2} [\frac{d\mu_i}{dx} - z_i \sum_k \frac{t_k}{z_k} \frac{d\mu_k}{dx}] \tag{7.18}$$

The first term in the right hand side simply says what we expect: If there is a net current, a flux density of species i will be set up proportional to the total current density and the transport number t_i, divided by the species’ charge.

The equation above is a rather general expression that we can use to calculate flux densities of one charged species in the company of many other species. However, it reflects the flux density and gradients and properties at a particular point through the membrane. The gradients will adjust according to the varying materials properties so as to maintain a constant flux density everywhere – what we call *steady state*. In order to implement this, we integrate the flux density expression over the thickness of the membrane and require that the flux density remains constant:

$$\int_I^{II} j_i dx = j_i X = \int_I^{II} \frac{t_i\, i_{tot}}{z_i e} dx - \int_I^{II} \frac{\sigma_i}{(z_i e)^2} [d\mu_i - z_i \sum_k \frac{t_k}{z_k} d\mu_k] \tag{7.19}$$

If the transport number t_i and the total current density i_{tot} can be taken as constant through the membrane we further obtain:

$$j_i X = \frac{t_i\, i_{tot}}{z_i e} X - \int_I^{II} \frac{\sigma_i}{(z_i e)^2} [d\mu_i - z_i \sum_k \frac{t_k}{z_k} d\mu_k] \tag{7.20}$$

Wagner-type transport theory case studies

Oxides with transport of oxide ions and electrons

General equations

We will now apply the general theory to a number of cases. The cases we will consider first is an oxide that conducts oxygen ions and electrons. Flux equations for the two are

$$j_{O^{2-}} = \frac{-\sigma_{O^{2-}}}{4e^2} [\frac{d\mu_{O^{2-}}}{dx} - 2e\frac{d\phi}{dx}] \tag{7.21}$$

$$j_{e^-} = \frac{-\sigma_{e^-}}{e^2} [\frac{d\mu_{e^-}}{dx} - e\frac{d\phi}{dx}] \tag{7.22}$$

By summing up the currents from the two we obtain (as in Eq. 7.7) the expression for the potential gradient:

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} + \frac{t_{O^{2-}}}{2e}\frac{d\mu_{O^{2-}}}{dx} + \frac{t_{e^-}}{e}\frac{d\mu_{e^-}}{dx} \tag{7.23}$$

In order to relate the potential gradient to the chemical potential of neutral species we introduce electrochemical equilibria between neutral and charged species of oxygen:

$$O_2(g) + 4\,e^- = 2\,O^{2-} \tag{7.24}$$

for which equilibrium can be expressed by

$$d\mu_{O_2(g)} + 4d\mu_{e^-} = 2d\mu_{O^{2-}} \tag{7.25}$$

Introducing this we get

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} + \frac{t_{O^{2-}}}{4e}\frac{d\mu_{O_2(g)}}{dx} + \frac{t_{O^{2-}}}{e}\frac{d\mu_{e^-}}{dx} + \frac{t_{e^-}}{e}\frac{d\mu_{e^-}}{dx} \tag{7.26}$$

and knowing that the sum of transport numbers for oxygen ions and electrons in the oxide we have defined is unity, we further get

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} + \frac{t_{O^{2-}}}{4e}\frac{d\mu_{O_2(g)}}{dx} + \frac{1}{e}\frac{d\mu_{e^-}}{dx} \tag{7.27}$$

whereby we have exemplified Eq. 7.11. We now integrate over the thickness of the sample to obtain the voltage over it, eliminating the chemical potential of electrons (as in Eq. 7.14 and onwards):

$$U_{II-I} = -IR + \frac{1}{4e}\int_{I}^{II} t_{O^{2-}}\, d\mu_{O_2(g)} \tag{7.28}$$

Using

$$\mu_{O_2(g)} = \mu^0_{O_2(g)} + kT \ln p_{O_2} \tag{7.29}$$

$$d\,\mu_{O_2(g)} = kTd\ln p_{O_2} \tag{7.30}$$

we get

$$U_{II-I} = -IR + \frac{kT}{4e}\int_{I}^{II} t_{O^{2-}}\,d\ln p_{O_2} \tag{7.31}$$

If $t_{O^{2-}}$ is independent of the oxygen partial pressure, or if we can assume as an approximation that an average $\overline{t_{O^{2-}}}$ is constant over the pressure range applied, then

$$U_{II-I} = -IR + \overline{t_{O^{2-}}}\frac{kT}{4e}\ln\frac{p_{O_2}^{II}}{p_{O_2}^{I}} = -IR + \overline{t_{O^{2-}}}E_N \tag{7.32}$$

where E_N is the Nernst-voltage of the oxygen concentration cell.

Open circuit voltage of concentration cell for transport number measurements

The equation above can be used to measure the average oxygen ion transport number, by measuring the open circuit voltage (OCV) over a sample exposed to a small, well-defined gradient in partial pressure of oxygen:

$$\overline{t_{O^{2-}}} = \frac{U_{II-I}}{E_N} \tag{7.33}$$

Galvanic oxygen sensor

By using a material with $t_{O^{2-}} = 1$, i.e. a solid electrolyte, and a well-defined reference partial pressure of oxygen, the OCV can be used for measuring an unknown partial pressure:

$$U_{II-\mathrm{Ref}} = \frac{kT}{4e}\ln\frac{p_{O_2}^{II}}{p_{O_2}^{\mathrm{Ref}}} \tag{7.34}$$

Fuel cell

If we apply a solid electrolyte with $t_{O^{2-}} = 1$ and expose it to a gradient in oxygen partial pressure, we get

$$U_{II-I} = -IR + \frac{kT}{4e} \ln \frac{p_{O_2}^{II}}{p_{O_2}^{I}} = -IR + E_N \qquad (7.35)$$

This would be the voltage of a fuel cell. With I = 0, we obtain the OCV of the cell to be the Nernst voltage, of course. We furthermore see that the cell voltage diminishes due to the IR loss at increasing current, I. A plot of U vs I gives –R as the slope.

The power delivered by the cell is the current multiplied with the voltage;

$$P = UI = -I^2 R + IE_N \qquad (7.36)$$

and it goes through a maximum at $U=E_N/2$. Under these conditions, the power is split half-half between external work and internal loss, so the electrical efficiency is 50%. The efficiency increases as the power and current decrease, and commonly fuel cells are designed to run at ca. $U=2/3\ E_N$.

Mixed conducting oxygen permeable membrane

Now we will show how we obtain the flux of oxygen through a mixed conducting oxygen membrane material. As mentioned in the general section this is done by inserting the electrical potential gradient back into the flux equation for oxygen ions. Using e.g. Eqs. 7.21, 7.23, and 7.25 we get:

$$j_{O^{2-}} = -\frac{t_{O^{2-}} i_{tot}}{2e} - \frac{\sigma_{O^{2-}} t_{e^-}}{4e^2} [\frac{d\mu_{O^{2-}}}{dx} - 2\frac{d\mu_{e^-}}{dx}] = -\frac{t_{O^{2-}} i_{tot}}{2e} - \frac{\sigma_{O^{2-}} t_{e^-}}{8e^2} \frac{d\mu_{O_2(g)}}{dx} \qquad (7.37)$$

The steady state restriction is introduced by integrating over the thickness of the membrane:

$$\int_I^{II} j_{O^{2-}} dx = j_{O^{2-}} X = -\int_I^{II} \frac{t_{O^{2-}} i_{tot}}{2e} dx - \int_I^{II} \frac{\sigma_{O^{2-}} t_{e^-}}{8e^2} d\mu_{O_2(g)} \qquad (7.38)$$

$$j_{O^{2-}} = -\frac{i_{tot}}{2eX} \int_I^{II} t_{O^{2-}} dx - \frac{kT}{8e^2 X} \int_I^{II} \sigma_{O^{2-}} t_{e^-} d\ln p_{O_2} \qquad (7.39)$$

If i_{tot} is not zero, i.e., some external current is drawn, then the first term says that the oxygen ion flux is proportional to that current and to the oxygen ion transport number. If the latter is independent of x, then the integral is straightforward. The last term adds to the flux; it is driven by an oxygen pressure gradient, and it is proportional to the oxygen ion conductivity and the transport number of electrons. This term describes permeation of oxygen ions, requiring transport of both oxygen ions and electrons. Moreover, it is inversely proportional to the thickness of the sample, X. Under open circuit conditions we may simplify to get:

$$j_{O^{2-}} = -\frac{kT}{8e^2 X} \int_I^{II} \sigma_{O^{2-}} t_{e^-} d\ln p_{O_2} = -\frac{kT}{8e^2 X} \int_I^{II} \sigma_{tot} t_{O^{2-}} t_{e^-} d\ln p_{O_2} \qquad (7.40)$$

We now apply our derived expression for the example case of an oxygen-deficient oxide MO_{1-x}. This contains oxygen vacancies compensated by electrons, and we have earlier shown that the defect concentrations are given by $n = 2[v_O^{\bullet\bullet}] = (4K'_{v_O^{\bullet\bullet}})^{1/3} p_{O_2}^{-1/6}$. Since the mobility of electrons is usually much higher than that of ionic defect, and with similar concentrations of electrons and oxygen vacancies, we may assume that the material is mainly an n-type electronic conductor. Therefore $t_{e^-} \approx 1$. Furthermore, $\sigma_{O^{2-}} = \sigma_{v_O^{\bullet\bullet}} = 2e[v_O^{\bullet\bullet}]u_{v_O^{\bullet\bullet}} = \sigma_0 p_{O_2}^{-1/6}$, and when we insert these in the flux expression we obtain

$$j_{O^{2-}} = -\frac{kT\sigma_0}{8e^2 X} \int_I^{II} p_{O_2}^{-1/6} d\ln p_{O_2} \qquad (7.41)$$

We substitute $d\ln p_{O_2} = \frac{1}{p_{O_2}} dp_{O_2}$ and obtain

$$j_{O^{2-}} = -\frac{kT\sigma_0}{8e^2X}\int_I^{II} p_{O_2}^{-7/6}\,d\,p_{O_2} \tag{7.42}$$

which integrates into

$$j_{O^{2-}} = \frac{6kT\sigma_0}{8e^2X}\left\{(p_{O_2}^{II})^{-1/6} - (p_{O_2}^{I})^{-1/6}\right\} \tag{7.43}$$

One may note that we here have utilised the Wagner approach for electrochemical transport of two defect species and combined it with the solution of the defect chemical equilibria to obtain an accurate description of the flux of oxygen ions. Numerical values for the flux of course require that one knows σ_0, which contains many parameters relating to formation and mobility of the defects. In the right-hand parenthesis the negative exponent –1/6 has the consequence that the smaller of the two partial pressures of oxygen has the biggest influence. The larger partial pressure makes a comparatively smaller difference and may be neglected to a first approximation for large gradients. For instance, consider a gas separation membrane exposed to air on the one side and a very reducing gas (e.g. methane) on the other. The flux would then be essentially proportional to $(p_{O_2}^{red})^{-1/6}$ while it would largely be independent of $p_{O_2}^{ox}$. It would thus not help much to compress the air to 100 atm, while removal of oxidised products like H_2O on the reducing side would lower $p_{O_2}^{red}$ and have a large effect.

Ambipolar diffusion and conduction

We have seen above that the transport coefficient term that enters in the flux of oxygen ions in a mixed oxygen ion conductor is a product which we may call the ambipolar conductivity

$$\sigma_{amb} = \sigma_{tot}\, t_{O^{2-}} t_{e^-} = \sigma_{O^{2-}} t_{e^-} = \frac{\sigma_{O^{2-}}\sigma_{e^-}}{\sigma_{O^{2-}} + \sigma_{e^-}} \tag{7.44}$$

This reflects the fact that so-called ambipolar conduction of two species is necessary (in this case for charge compensation).

The corresponding terms transformed to diffusivities are called the ambipolar diffusion coefficients for oxygen:

$$D_{amb} = D_{O^{2-}} t_{e^-} \tag{7.45}$$

Such terms, like partial conductivity and self-diffusion, reflect thermal random mobility as before, but are restricted by the slower of the two ambipolar diffusing species (ions or electrons).

Many other cases of combined transport coefficients are in use, e.g. the combined (additive) transport of oxygen and metal ions commonly that we shall address later (and exemplify by the high temperature oxidation of metals), the combination of two diffusivities involved in interdiffusion (mixing) processes, and the mass transport in creep being rate limited by the smallest out of cation and anion diffusivities in a binary compound. As some of these sometimes are referred to as ambipolar or chemical diffusivities, we want to stress the above simple definition of ambipolar transport coefficients as relevant for membrane applications using mixed conductors.

Chemical diffusion and tracer diffusion

General case

The *chemical diffusion coefficient* is the *phenomenological* coefficient that enters a Fick's 1st law of diffusion in a concentration gradient:

$$j_i = -\tilde{D}_i \frac{dc_i}{dx} \tag{7.46}$$

Unlike our previous treatment in this chapter (which uses D_r, B, u or σ), Fick's 1st law and the chemical diffucion coefficient says in general nothing directly about the self-diffusion coefficient. Neither is, as we shall see, the chemical diffusion coefficient in general a constant in any sense. Only when the species *i* is an effectively neutral particle that moves independently of other species will the chemical diffusion coefficient be identical to the self-diffusion coefficient. If the flux and the concentration gradient are in terms of an isotope (tracer), the coefficient is the tracer diffusion coefficient, which is almost equal to the self-diffusion coefficient:

$$D_t = f\ D_r \tag{7.47}$$

where the correlation factor f is a number usually equal to or somewhat smaller than unity, depending on structure and diffusion mechanism, reflecting the difference of transport between distinguishable and indistinguishable diffusing species. For vacancy diffusion, f usually falls around 0.75 depending on structure.

Chemical diffusion is sometimes referred to as diffusion under a chemical potential or concentration gradient. However, we have seen that self-diffusion properly describes this, and that all diffusion phenomena we consider here are merely minor perturbations of thermally induced self-diffusion. Why then do we need to be concerned with chemical diffusivity? We use it whenever we want to or have to relate our flux to the concentration gradient. Under steady state processes, our flux vs force equivalents provides sufficient description, and can be integrated to yield steady state fluxes through membranes, as we have shown; Thus, for this purpose, the self-diffusion coefficients and conductivities are useful since they behave in a simple manner vs defect concentrations, and since we in general know the chemical and/or electrical potential gradients applied. However, if we know the concentration gradients and how the chemical diffusivity varies with concentration, we *can* use chemical diffusivity.

In a transient experiment, however, parts of the material take up or give away matter (leading to changes in the concentration of species with time). This is sometimes referred to as a capacitive effect and is expressed in Fick's 2nd law:

$$\frac{dc}{dt} = -\frac{dj}{dx} = \frac{d(D\frac{dc}{dx})}{dx} \tag{7.48}$$

If D is independent of c and thus of x and t, then we can simplify to obtain:

$$\frac{dc}{dt} = -\frac{dj}{dx} = D\frac{d^2c}{dx^2} \tag{7.49}$$

Mathematical solutions for interpretation of transient tracer, thermogravimetric and electrical conductivity experiments are based on this. The diffusion coefficient obtained is the tracer diffusion coefficient in case of a transient in isotope composition (whether measured radiographically, mass spectrometrically, thermogravimetrically or electrically) or chemical diffusion coefficients in case of a transient in chemical composition (measured thermogravimetrically or electrically or by other means). The latter is usually executed by stepping the oxygen activity by gas composition, total pressure, or electrochemical means.

It may be noted that certain tracer experiments actually yield an interdiffusion coefficient between two tracers. The diffusion coefficient for oxygen in practice cannot and need not distinguish between the different isotopes in use (usually ^{18}O and ^{16}O, more rarely ^{17}O). However, for protons, the difference to deuteron (or triton) transport coefficients is considerable (e.g. a factor of 2) and should be taken into account when interpreting the results.

The multitude of transport coefficients collected can thus be divided into self-diffusion types (total or partial conductivities and mobilities obtained from equilibrium electrical measurements, ambipolar or self-diffusion data from steady state flux measurements through membranes), tracer-diffusivities, and chemical diffusivities from transient measurements. All but the last are fairly easily interrelated through definitions, the Nernst-Einstein relation, and the correlation factor. However, we need to look more closely at the chemical diffusion coefficient. We will do this next by a specific example, namely within the framework of oxygen ion and electron transport that we have restricted ourselves to at this stage.

Chemical diffusivity in the case of mixed oxygen vacancy and electronic conductor

In a system with oxygen vacancies as the only oxygen point defects, the concentration of oxygen ions is the concentration of oxygen sites *minus* the concentration of vacancies:

$$c_{O^{2-}} = c^{0}_{O^{2-}} - c_{v_O^{\cdot\cdot}} \qquad (7.50)$$

such that

$$\frac{dc_{O^{2-}}}{dx} = -\frac{dc_{v_O^{\cdot\cdot}}}{dx} \qquad (7.51)$$

We manipulate this in the following manner:

$$\frac{dc_{O^{2-}}}{dx} = -\frac{dc_{v\ddot{o}}}{dx}\frac{d\ln p_{O_2}}{d\ln p_{O_2}} = -\frac{dc_{v\ddot{o}}}{d\ln p_{O_2}}\frac{d\ln p_{O_2}}{dx} \tag{7.52}$$

and insert the result into the flux equation for oxygen ions, Eq. 7.37 (open circuit, $i_{tot} = 0$):

$$j_{O^{2-}} = -\frac{\sigma_{O^{2-}}\, t_{e^-}}{8e^2}\frac{d\mu_{O_2(g)}}{dx} = -\frac{\sigma_{O^{2-}}\, t_{e^-} kT}{8e^2}\frac{d\ln p_{O_2}}{dx} = \frac{\sigma_{O^{2-}}\, t_{e^-} kT}{8e^2}\frac{d\ln p_{O_2}}{dc_{v_O^{\bullet\bullet}}}\frac{dc_{O^{2-}}}{dx} \tag{7.53}$$

We have thus derived an expression for the flux that relates it to the concentration gradient of oxygen ions. We may now therefore compare this expression with that of the oxygen flux in terms of Fick's 1st law involving the concentration gradient in oxygen ions and the chemical diffusion coefficient:

$$j_{O^{2-}} = -\tilde{D}_{O^{2-}}\frac{dc_{O^{2-}}}{dx} \tag{7.54}$$

and we obtain

$$\tilde{D}_{O^{2-}} = -\frac{\sigma_{O^{2-}}\, t_{e^-} kT}{8e^2}\frac{d\ln p_{O_2}}{dc_{v_O^{\bullet\bullet}}} \tag{7.55}$$

Further, by using the Nernst-Einstein-relation and inserting the relation $D_{O^2} c_{O^{2-}} = D_{V\ddot{o}} c_{V\ddot{o}}$ we get

$$\tilde{D}_{O^{2-}} = -\frac{D_{O^{2-}} c_{O^{2-}}\, t_{e^-}}{2}\frac{d\ln p_{O_2}}{dc_{v_O^{\bullet\bullet}}} = -\frac{D_{v_O^{\bullet\bullet}} c_{v_O^{\bullet\bullet}}\, t_{e^-}}{2}\frac{d\ln p_{O_2}}{dc_{v_O^{\bullet\bullet}}} = -\frac{D_{v_O^{\bullet\bullet}}\, t_{e^-}}{2}\frac{d\ln p_{O_2}}{d\ln c_{v_O^{\bullet\bullet}}} = -\frac{D_{v_O^{\bullet\bullet}}\, t_{e^-}}{2\dfrac{d\ln c_{v_O^{\bullet\bullet}}}{d\ln p_{O_2}}}$$

(7.56)

Thus, depending on the p_{O_2}-dependency of the vacancy concentration, the chemical diffusion coefficient becomes a varying function of the vacancy (defect) self diffusion coefficient. When the sample is mainly an electronic conductor and the transport number for electrons is unity, the collection of remaining terms (the enhancement factor) takes absolute values from 1 to 3 when the sample has a variable non-stoichiometry, while it increases beyond this (and can take on very large values) as the stoichiometry approaches a constant value, as in certain doped materials or materials with prevailing point defect (Schottky or Frenkel) disorder; the vacancy concentration is then close to independent of p_{O_2}.

The full range of aspects of chemical diffusion coefficients make them rather difficult to translate into other transport coefficients, and they tend to live their own lives in the literature. They can, as said before, be applied directly to calculate fluxes in membranes if the concentration gradients through the membranes are known. However, we note that the chemical diffusion coefficient (for oxygen ions) is

- *proportional* to the self diffusivity of the oxygen *defect* (here the vacancy),
- *proportional* to the electronic transport number (but which is often unity),
- *enhanced* by the term $d\ln p_{O_2}/2d\ln c_{V_O}$, which for various limiting cases of simplified defect situations takes on values of 1, 3 or ∞, and
- fully *forwards and backwards transformable into defect diffusivities* (and in turn self diffusivities) if we know the transport numbers and how defect concentrations vary with pO_2.

Surface and interface kinetics limitations

The flux of matter through a fuel cell or electrolyser is limited by the electrolyte (ionic) resistance, the electrode kinetics, and the external electronic load resistance. We commonly express the steady state situation and the fact that the current is the same through the entire closed circuit in terms of the voltage drops around the circuit:

$$E = E_N - I R_i - (\eta_c + \eta_a) \qquad (7.57)$$

where I is the current, E_N is the open circuit (Nernst) voltage, R_i is the ionic electrolyte resistance, and η_c and η_a are the cathode and anode overpotentials, respectively. (The conventions and practices on measuring and reporting the sign of various voltages may vary, so that the equation may take various forms with respect to signs).

Similarly, the current of matter through an electrodeless gas separation, mixed conducting membrane will be limited by ionic conduction, the surface kinetics, and by the *internal* electronic resistance, the latter replacing the external circuit in the previous case.

However, we may need to express the loss of potential in terms of the chemical potential of oxygen.

If we look at the system without any gradients or currents applied, there will be equilibrium at both the electrodes and the surfaces. Depending on the application and the side of the membrane, the electrochemical processes in operation on an oxygen ion conductor may be:

$$O_2(g) + 4e^- = 2O^{2-} \tag{7.58}$$

and

$$H_2(g) + O^{2-} = H_2O(g) + 2e^- \tag{7.59}$$

In all of these the electrons can be supplied to or from an electrode or a mixed conductor surface. Thus, basically, the reactions are the same in the electrolyte/electrode as in the mixed conductor, although the physical presence of the electrode can promote (by catalysis) or obstruct the reaction.

Most surface reactions are believed to be made up of a sequence of reactions steps. Some possible step reactions in the case of the reduction of oxygen gas are given by the following reaction sequence;

$$O_2(g) = O_{2,ads}$$

$$O_{2,ads} + e^- = O^-_{2,ads}$$

$$O^-_{2,ads} + e^- = 2\,O^-_{ads}$$

$$O^-_{ads} + e^- = O^{2-}_{ads}$$

$$O^{2-}_{ads} = O^{2-}_O$$

The last of these steps may involve bulk defects (e.g. a vacancy that the adsorbed ion may occupy) but we will not attempt here an analysis of surface or electrode kinetics in terms of defect chemistry. Determination of the nature of the rate limiting step of a reaction is a difficult task. Studies of surface exchange kinetics using ^{18}O or by relaxation methods can give information about the temperature and pressure dependence of the rate determining step. However, it should be pointed out that information from such methods do not give direct information about the exact mechanism and which intermediate species that are involved. Each step will have its own specific rate and resistance to the mass transport across the membrane. In lack of detailed knowledge of the rate limiting step(s) we normally have to relate to the overall reaction.

At equilibrium the overall reaction takes place forth and back at equal rates r_0 as a result of thermal energy. The activation process forth and back may not be symmetrical in this case, but that will not affect us at present. Similarly to the random diffusion coefficient connected with these thermal disorders in bulk ($D_r = 1/6\ r_0\ s^2$) we introduce an exchange coefficient for interfaces; $k_i = r_0 s_i$ which says something about the thermal fluctuations of the reaction across the interface of thickness s_i. We can alternatively express this as an exchange flux density

$$j_0 = k_i c = r_0 s_i c \qquad (7.60)$$

or an exchange current density

$$i_0 = j_0 ne = k_i nec = r_0 s_i nec \qquad (7.61)$$

where n is the number of electrons involved in the reaction. If we now apply a small force over the interface, we get a perturbation and a net flux, given by

$$j = -c\, k_i \frac{dP}{kT} \qquad (7.62)$$

where c is the volume concentration of species in the interface, and dP is the potential step over the interface. The net current density is

$$i = jne = cne\, k_i \frac{dP}{kT} = i_0 \frac{dP}{kT} \tag{7.63}$$

If *dP* is taken as the electrical (over-)potential energy step affecting the reaction, $dP = \eta ne$ and the overpotential becomes

$$\eta = \frac{kT}{ne}\frac{i}{i_0} \tag{7.64}$$

This is the linear or ohmic (small overpotential) form of the kinetics expression, as we expect from our specification of small forces and perturbations. We note that the interface thickness s_i has dropped out and is mostly not an interesting parameter at this level.

By measuring the overvoltage vs current density by voltammetric DC or impedance spectroscopy AC methods, we may obtain the charge transfer resistance

$$R_e = \frac{\eta}{i} = \frac{kT}{ne\, i_0} = \frac{kT}{n^2 e^2 k_i c} \tag{7.65}$$

The overpotential enters directly into the sum of voltages over the fuel cell or electrolyser circuit, and we can solve the system in terms of current and voltage. As the force and perturbations of the reaction becomes larger, the full so-called Butler-Volmer exponential dependence of current and flux on overpotential (and force) must be taken into account:

$$\eta = \frac{kT}{ne}\ln\left(\frac{i}{i_0}\right) \tag{7.66}$$

, but we will not elaborate on this here.

In order to approach an interpretation of transport measurements, we take as starting point that the flux(es) through the interface must be the same as that in the bulk of the membrane next to the interface in all cases (and in the entire membrane in the steady state cases). This is done in reality and in the mathematical solutions by adjusting the potential at the border between the interface and the bulk so as to equalise the two fluxes. However, this soon becomes difficult, as it is not trivial to choose the proper set of potentials to apply for different interface processes.

We resort first to chemical diffusion and assume that the interface process is one which can be described as species flowing from a high to a lower concentration. Then

$$- D(\frac{dc}{dx})_{bulk} = -k\Delta c_{interface} \tag{7.67}$$

, and D as well as k must refer to chemical coefficients. This requirement can be included in the solution of Fick's 2nd law for transients, and D and k can in principle be found independently by fitting transient data to this solution. The k's obtained in this way can be used to predict steady state fluxes if they are measured as a function of e.g. p_{O_2} or at the p_{O_2} extremes which the membrane is going to be subjected to. Otherwise, the k values are not immediately easy to relate to other data.

In a tracer experiment, we obtain D_t which is close to the self diffusivity ($D_t = f\ D_r$) and then the exchange coefficient should also reflect the tracer exchange coefficient $k_t \approx k_i$ and this should be comparable to that calculated as above from measurements of R_e; the difference should say to what extent the presence of an electrode has changed the kinetics of the rate limiting step of the surface reaction.

In order to analyse the effect of interface kinetics on fluxes through a membrane under steady state conditions, we may take an approach similar to that for fuel cells and electrolysers. Principally, no electrodes are involved, and voltages are therefore not directly appropriate as in the other cases. Instead we may sum up chemical potential changes of oxygen (or hydrogen) through the membrane and its interfaces. We start out with the membrane interior, and from earlier we have for a mixed oxygen ion and electron conductor of thickness L (earlier also referred to as X or Δx):

$$j_{O^{2-}} = -\frac{1}{L}\int_{I}^{II} \frac{\sigma_{O^{2-}}\, t_{e^-}}{8e^2} d\mu_{O_2(g)} \tag{7.68}$$

If we introduce ambipolar diffusivities and conductivities and assume these constant, the equation may be integrated to obtain

$$j_{O^2} = \frac{D_{amb}\, c_{O^{2-}}}{2kT}\frac{\Delta\mu_{O_2}}{L} = \frac{\sigma_{amb}}{8e^2}\frac{\Delta\mu_{O_2}}{L} \tag{7.69}$$

and, by rearrangement, the chemical potential drop over the bulk of the membrane is

$$\Delta\mu_{O_2} = \frac{-2\, j_{O^{2}}\, kTL}{D_{amb}\, c_{O^{2-}}} = \frac{-\, j_{O^{2}}\, 8\, e^{2}\, L}{\sigma_{amb}} \tag{7.70}$$

This drop over the membrane bulk is given by the external full gradient minus the losses at the interfaces. In the present terms, the flux through an interface is, from Eqs. 7.62 and 7.65 given by

$$j_{O^{2-}} = \frac{-\, c_{O^{2-}}\, k_i \Delta\mu_{O_2}^{interface}}{2kT} = \frac{-\,\Delta\mu_{O_2}^{interface}}{8\, e^{2}\, R_e} \tag{7.71}$$

such that, by rearrangement,

$$\Delta\mu_{O_2}^{interface} = \frac{-2\, j_{O^{2}}\, kT}{k_i\, c_{O^{2-}}} =-\, j_{O^{2}}\, 8\, e^{2} R_e \tag{7.72}$$

Now, we sum up the total gradient in chemical potential of oxygen over the bulk of the membrane and the two interfaces

$$\Delta\mu_{O_2}^{tot} = \Delta\mu_{O_2}^{bulk} + 2\Delta\mu_{O_2}^{interface} \tag{7.73}$$

and after insertion of the above expressions for bulk and interface, we can solve for the flux of oxygen ions through a membrane with two equally contributing interfaces under the conditions given:

$$j_{O_{2-}} = \frac{-\Delta\mu_{O_2}^{tot} c_{O_{2-}}/2kT}{\dfrac{L}{D_{amb}} + \dfrac{2}{k_i}} = \frac{-\Delta\mu_{O_2}^{tot}/8\, e^{2}}{\dfrac{L}{\sigma_{amb}} + 2\, R_e} \tag{7.74}$$

This result simply states that the flux is given by the driving force (which easily transforms to Nernst voltage) divided by the total ambipolar resistance, in our case namely that of the bulk and twice that of one interface.

In lack of full understanding of the relations involved in interface kinetics of membranes, we may refer to a critical thickness; the thickness of the bulk membrane where the interface and bulk impose equally large restrictions on the flux. In any membrane application, going much below this thickness is not of much use. We see from the above that this thickness is L_{crit}=D/k for the thickness of the part of a membrane relating to one interface, or L_{crit} =2D/k for thickness of a membrane with two equally limiting interfaces, typically the two surfaces.

Experimentally the value of L_{crit} has been found to be of the order of 100 μm for many polished samples of different fluorite and perovskite materials, indicating that there is a close relationship between e.g. vacancy concentration in bulk and the surface exchange. Roughing the surface should increase relatively the surface flux, and this has been shown to work.

Ionic transport of both anions and cations

General expressions

We have in the previous section considered ionic transport by oxygen ions (anions) only. This was done for simplicity, because it represents many real applications, and because oxygen ions relate most directly to the oxygen (non-metal) activity which we may control most conveniently.

Now we will include cation (metal ion) transport. Some systems have predominant transport of cations, and we thus need to see how this relates this to the non-metal activity. Some systems have both anion and cation transport so we need to take both into account. In some cases we may not know which one is dominating and as we shall see it is not always easy or necessary to distinguish them.

In an inorganic compound the total ionic current density, i_{ion}, is given by the sum of the current densities of anions, i_{an}, and the cations, i_{cat}. From our previously derived expressions for partial current densities we thus obtain

$$i_{ion} = i_{an} + i_{cat} = z_{an}ej_{an} + z_{cat}ej_{cat} = \frac{-\sigma_{an}}{z_{an}e}\frac{d\eta_{an}}{dx} + \frac{-\sigma_{cat}}{z_{cat}e}\frac{d\eta_{cat}}{dx} \qquad (7.75)$$

where z_{an} and z_{cat} represent the valences and thus ionic charges of the anions and cations, respectively.

In order to relate the current density of the cations and anions let us consider the equilibrium involving the formation of the compound M_aO_b from its ions:

$$aM^{z_{cat}} + bO^{z_{an}} = M_aO_b \tag{7.76}$$

Through the Gibbs-Duhem relation, equilibrium in this reaction may be expressed as

$$ad\mu_{M^{z_{cat}}} + bd\mu_{O^{z_{an}}} = d\mu_{M_aO_b(s)} \tag{7.77}$$

The fact that $d\mu_{M_aO_b(s)} = 0$ arises from $M_aO_b(s)$ being a pure condensed phase. It may further be noted that

$$az_{cat} = -bz_{an} \tag{7.78}$$

so that we obtain the very important expression

$$\frac{d\mu_{cat}}{dx} = -\frac{b}{a}\frac{d\mu_{an}}{dx} = \frac{z_{cat}}{z_{an}}\frac{d\mu_{an}}{dx} \tag{7.79}$$

From this we note that the cation and anion chemical potential gradients are the negative of each other in a binary compound, a relation that is used extensively, mathematically or by intuition.

By adding $z_{cat}e\frac{d\phi}{dx}$ to both sides of the equation and rearranging on the right hand side we obtain

$$\frac{d\eta_{cat}}{dx} = \frac{z_{cat}}{z_{an}}\frac{d\eta_{an}}{dx} \tag{7.80}$$

By combining Eqs 7.75. and 7.80 the total ionic conductivity may be expressed by

$$i_{ion} = \frac{\sigma_{ion}}{z_{cat}e}\frac{d\eta_{cat}}{dx} = \frac{\sigma_{ion}}{-z_{an}e}\frac{d\eta_{an}}{dx} \qquad (7.81)$$

Thus, the ionic current density can be expressed in terms of the ionic conductivity (sum of cationic *and* anionic conductivities) and the gradient in the chemical potential of cations *or* anions.

From this it becomes clear that the derivations of oxygen ion fluxes and currents in various situations done earlier could have been done for the case of ions (sum of cations and anions) simply by inserting σ_{ion} instead of $\sigma_{O^{2-}}$, but with for instance the oxygen activity gradient as the driving force.

Membrane "walk-out"

Metal ions are transported from the low oxygen pressure to the high oxygen pressure side of the membrane while metal vacancies are transported in the opposite direction. During the process oxygen is liberated and oxygen and metal sites are annihilated at the low oxygen pressure side, e.g. for a metal deficient oxide $M_{1-\partial}O$ with doubly charged metal vacancies:

$$O_O + V_M^{2'} + 2h^{\cdot} = \frac{1}{2} O_2(g) \qquad (7.82)$$

while equivalent number of lattice sites are formed at the high oxygen pressure side:

$$\frac{1}{2} O_2(g) = O_O + V_M^{2'} + 2h^{\cdot} \qquad (7.83)$$

In this way the oxide membrane is actually moving in laboratory space in the direction of the higher oxygen potential.

From this we see that even a very minor transport of metal ions may be detrimental to the operation of an oxygen separation membrane or fuel cell: Over the many years of operation the electrolyte or membrane will simply "walk out" of its housing, towards the high oxygen pressure.

Demixing of oxide solid solutions

An additional important aspect of diffusional transport of metal ions in a chemical potential gradient is that in a homogeneous crystal of an oxide solution, e.g. $(A,B)_{1-\delta}O$ such as for instance $(Co,Ni)_{1-\delta}O$, a demixing process begins to take place. Both cations move by vacancy diffusion. When one of the cations in $(A,B)_{1-\delta}O$, e.g. A^{2+}, has a higher mobility than the other, the A^{2+} ions will move faster to the side of the higher oxygen potential. The solid solution is enriched in AO at this side and thus becomes kinetically demixed. After extended time a steady state concentration profile is reached. This can be formulated through the use of the appropriate transport equations and by taking into account the conditions of electroneutrality and of local thermodynamic equilibrium. The process is schematically illustrated below.

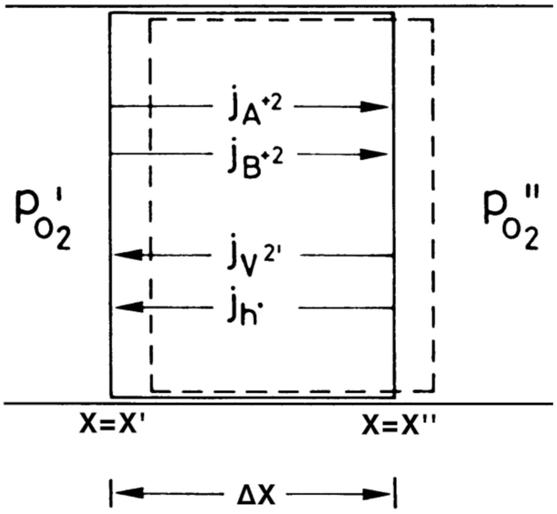

Figure 7-1. Schematic illustration of kinetic demixing of the oxide solid solution $(A,B)_{1-\delta}O$ in a oxygen potential gradient.

Multicomponent compounds, e.g. ABO_3 and AB_2O_4, can similarly be demixed in an oxygen potential gradient. In this case defects are formed due to the nonstoichiometric A/B ratio. Both in the case of the solid solution and in the case of a ternary oxide, new phases may be precipitated as soon as the concentration of solute or defects exceeds the solubility limit. We may in fact end up in the peculiar situation that the starting oxide is stable in both the low and the high oxygen activities separately, but unstable in the gradient between them.

The decomposition and precipitation of new phases may violate the functional or mechanical properties of the material. These types of phenomena are not only of importance in oxide membranes, but also in oxidation of alloys where solid solutions of two or more oxides or multicomponent compounds may be formed in the oxide scales.

High temperature oxidation of metals; the Wagner oxidation theory

When high temperature oxidation of metals results in the formation of compact scales and sufficient oxygen is available at the oxide surface, the rate of reaction is governed by the solid state diffusion of the reactants or transport of electrons through the scale. As the diffusion distance increases as the oxide grows in thickness, the rate of reaction decreases with time. When the diffusion takes place homogeneously through the scale, the rate of growth of the oxide thickness, x, is inversely proportional to the oxide thickness:

$$\frac{dx}{dt} = k_p^* \frac{1}{x} \tag{7.84}$$

In the integrated form this becomes

$$x^2 = 2k_p^* t + C_0 = k_p t + C_0 \tag{7.85}$$

where $k_p (= 2k_p^*)$ and k_p^* are expressions for the rate constants and C_O is an integration constant. The oxide thickness may alternatively be measured in terms of the oxygen uptake by the metal specimen during oxidation by using, for instance, thermogravimetry. The oxide thickness or oxygen uptake is parabolic with time, and the oxidation is termed parabolic. Accordingly k_p and k_p^* are termed parabolic rate constants. The figure below illustrates the variation in the oxide thickness or oxygen uptake as a function of time.

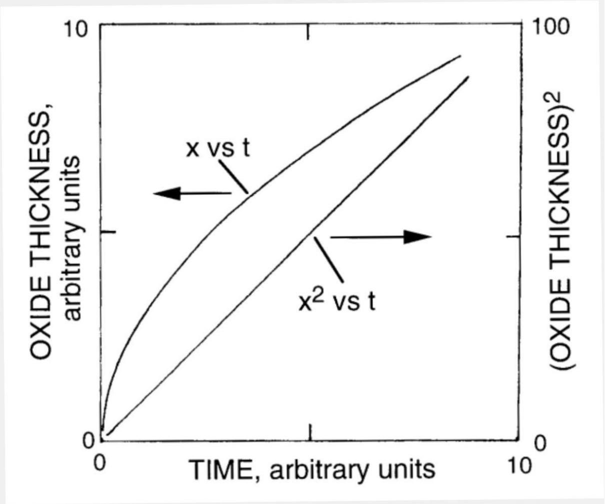

Figure 7-2. The variation in oxide thickness as a function of time (x vs t and x^2 vs t) during parabolic oxidation of metals

Carl Wagner developed the theory for parabolic oxidation assuming that the reaction is governed by lattice diffusion of the reacting ions (metal and oxygen ions) or transport of electrons. The important aspect of the Wagner oxidation theory is that the parabolic rate constant is expressed in terms of independently measurable properties, i.e. the electronic and ionic conductivity of the oxide or alternatively in terms of the self-diffusion coefficients of the reacting ions. This thus provides an interpretation of the reaction mechanism and a theoretical basis for changing and improving the oxidation resistance of metals and alloys. The Wagner theory has been one of the most important contributions to our understanding of the high temperature oxidation of metals.

The basic assumption of the theory is that the lattice diffusion of the reacting ions or the transport of electrons through the dense scale is rate-determining for the overall reaction. Lattice diffusion is assumed to take place because of the presence of point defects, and the migrating species may alternatively be considered to constitute lattice and electronic defects. Wagner further assumed that ions and electrons migrate independently of each other and that local equilibria exist within the oxide. Such transport processes through a dense, single-phase scale growing by lattice diffusion is illustrated in the figure below. One part (a) illustrates the transport of metal and oxygen ions and electrons, while the other (b) illustrates the transport processes when the predominant diffusion processes involve metal vacancies and interstitials.

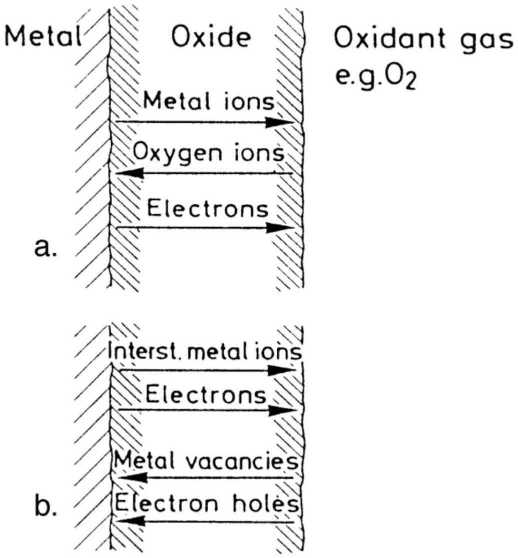

Figure 7-3. Transport processes during growth of a dense, single phase scale growing by lattice diffusion. a) transport processes illustrated by transport of reacting ions, b) transport processes illustrated by transport of diffusion of important point defects (in this case assumed to be metal vacancies and interstitials).

As diffusion through the scale is rate-determining, reactions at the phase boundaries are considered to be rapid, and it is assumed that thermodynamic equilibria are established between the oxide and oxygen gas at the oxide/oxygen interface and between the metal and the oxide at the metal/oxide interface.

The overall "driving energy" of the reaction is the Gibbs free energy change associated with the formation of the oxide, e.g. M_aO_b, from the metal M and the oxygen gas. Correspondingly, a gradient in the partial pressure (activity) of oxygen exists across the scale from the partial pressure of oxygen in the ambient atmosphere at the outer surface to the partial pressure of oxygen at the metal/oxide interface. The latter partial pressure is the decomposition pressure of the oxide in equilibrium with its metal.

The main driving force for the transport through the plane is the chemical potential gradient. But when one considers the transport of the reacting ions and of the electrons through the scale, it may also be noted that the mobilities of the cations, anions and electrons are not equal. Because of this difference, a separation of charges takes place in a growing scale. This creates a space charge (diffusion potential) that opposes a further separation of charges, and a stationary state is reached for which no net electric current flows through the scale. In describing the transport of ions and electrons through the scale, it is thus necessary to take into account both the transport due to the gradient of the chemical potential and that due to the gradient of the electrical potential, i.e. the electrochemical potential gradient. Thus, the treatment of the transport through the oxide scale is identical to that done earlier for electrochemical transport through a mixed conducting oxide.

The net ionic current is from earlier given by

$$i_{ion} = -\frac{\sigma_{ion}}{z_{an}e}\frac{d\eta_{an}}{dx} = \frac{\sigma_{ion}}{2e}\frac{d\eta_{O^{2-}}}{dx} \qquad (7.86)$$

and the electronic current is given by

$$i_{el} = -\frac{\sigma_{el}}{-1e}\frac{d\eta_{el}}{dx} = \frac{\sigma_{el}}{e}\frac{d\eta_{el}}{dx} \qquad (7.87)$$

As no net current flows through the scale during the scale growth, then $i_{tot} = i_{ion} + i_{el} = 0$. By solving this system in the usual manner (solving for the electrical potential gradient, inserting into the expression for the ionic current density and inserting the equilibrium condition between oxygen ions, oxygen molecules and electrons) we obtain

$$i_{ion} = \frac{\sigma_{ion} t_{el}}{4e} \frac{d\mu_{O_2(g)}}{dx} = \frac{\sigma_{ion} t_{el} kT}{4e} \frac{d \ln p_{O_2}}{dx} \tag{7.88}$$

The rate of oxygen ions reacting per unit area is obtained by dividing the ionic current density by the charge of the oxygen ions. Moreover, the rate dn/dt of oxide molecular units M_aO_b formed per unit area is obtained by further division by b:

$$\frac{dn}{dt} = \frac{1}{b}\frac{dn_O}{dt} = \frac{i_{ion}}{-2eb} = -\frac{\sigma_{ion} t_{el}}{8e^2 b} \frac{d\mu_{O_2(g)}}{dx} = -\frac{\sigma_{ion} t_{el} kT}{8e^2 b} \frac{d \ln p_{O_2}}{dx} \tag{7.89}$$

Similar to the integrations we have done earlier over the thickness of the oxide, we now integrate over the instantaneous thickness *Δx* and from the ambient oxygen pressure, $p_{O_2}^0$, to the partial pressure at the metal/oxide interface, $p_{O_2}^i$. The growth rate $\frac{\mathrm{dn}}{\mathrm{dt}}$ then takes the form

$$\frac{dn}{dt} = \left\{ \frac{kT}{8e^2 b} \int_{\mathrm{p}_{\mathrm{O}_2}^{\mathrm{i}}}^{\mathrm{p}_{\mathrm{O}_2}^{\mathrm{o}}} \sigma_{ion} t_{el} \, d \ln p_{O_2} \right\} \frac{1}{\Delta x} \tag{7.90}$$

We have chosen to organize the equation and direction of integration such that the rate is positive if $p_{O_2}^0 > p_{O_2}^i$. In general the directionalities and signs of the fluxes and processes that go on pose a problem for us. One may most often neglect this issue because the output, namely growth (increase in amount of oxide or thickness of scale) is intuitively positive. However, it may not be just that easy, since there is nothing preventing scale reduction (reverse of growth) if the outer oxygen partial pressure is smaller than the activity at the metal/oxide.

Many readers will probably wonder about another small apparent problem: We integrate over the scale to take into account the steady state condition, namely that the fluxes are constant throughout the scale. Still, the scale grows, so how can it be steady state? The answer is that it is steady state (constant fluxes) for that moment of time with that instantaneous thickness of the scale. So our integration is for a given moment in time. At another moment the integration will give a different flux.

The expression in the parentheses, $\left\{ \frac{kT}{8e^2 b} \int_{p_{O_2}^{i}}^{p_{O_2}^{o}} \sigma_{ion} t_{el} \, d \ln p_{O_2} \right\}$, can be considered to be one form of the parabolic rate constant and is in the following written k_t:

$$\frac{dn}{dt} = k_t \frac{1}{\Delta x} \tag{7.91}$$

k_t thus represents a time-independent coefficient, while the flux and thickness of the scale vary according to the parabolic relation above.

As written in this form $\frac{dn}{dt}$ is expressed in number of molecular units of M_aO_b per cm^2sec and Δx in cm. The derivation of Wagner's equation that we here have done for growth of oxide scales, may be applied to many other gas-metal reactions.

It may be noted that although the total particle current is equal to the rate of growth of the scale, the rate-determining process may either be diffusion of ions or transport of electrons depending on the properties of the scales. If the scale is an electronic conductor, $t_{el} \approx 1$, the diffusion of the reacting ions through the scale is rate-determining, while the transport of electrons through the scales is rate-determining if the scale is an ionic conductor, $t_{ion} \approx 1$. The oxygen pressure dependence of the reaction will depend on whether the scale is an ionic or electronic conductor and in the latter case on the type of nonstoichiometry that prevails in the oxide scale.

In our derivation, the growth rate is proportional to $\sigma_{ion}t_{el} = \sigma_{tot}t_{ion}t_{el} = (\sigma_{an} + \sigma_{cat})\, t_{el}$. However, the conductivity of the metal and oxygen ions in M_aO_b can be expressed in terms of the self-diffusion coefficients of the metal and oxygen ions in M_aO_b through the Nernst - Einstein relation ($z_i^2e^2D_ic_i = kT\,\sigma_i$). Let us insert this relationship in Eq.7.89 and assume that the oxide is an electronic conductor ($t_{el} \approx 1$). Let us furthermore take into account that concentrations of metal ions (cations), c_M, and oxygen ions, c_O, in an oxide M_aO_b with relatively small deviation from stoichiometry are related through $c_M/c_O = a/b = |z_{an}|/z_{cat}$. The expressions for the flux through the growing scale (expressed in terms of the number of M_aO_b units per cm^2sec.) and for k_t then become

$$\frac{dn}{dt} = -\frac{c_O z_{an}^2}{8b} \left(\frac{z_{cat}}{|z_{an}|} D_M + D_O \right) \frac{d \ln p_{O_2}}{dx} = -\frac{c_O}{2b} \left(\frac{z_{cat}}{2} D_M + D_O \right) \frac{d \ln p_{O_2}}{dx} \tag{7.92}$$

$$\frac{dn}{dt} = \left\{ \frac{c_0}{2b} \int_{p^{i}_{O_2}}^{p^{o}_{O_2}} \left(\frac{z_{cat}}{2} D_M + D_O\right) d \ln p_{O_2} \right\} \frac{1}{\Delta x} = k_t \frac{1}{\Delta x} \qquad (7.93)$$

c_O denotes the number of oxygen ions in M_aO_b per cm^3 and D_M and D_O are respectively the self-diffusion coefficients of the metal and oxygen ions in M_aO_b. It should be noted that D_M and D_O are the self-diffusion coefficients for random diffusion of the respective ions.

In Eqs.7.92 and 7.93, k_t represents the rate of formation of M_aO_b units per cm^2 per second for an M_aO_b scale of thickness 1 cm. One may as said before alternatively express the parabolic rate constant in terms of the rate of growth of the oxide thickness

$$\frac{dx}{dt} = k_p^* \frac{1}{x} \qquad (7.94)$$

The dimensions of k_p^* are then $cm^2 sec^{-1}$, and Eq.7.93 then takes the form

$$\frac{dx}{dt} = \frac{b}{c_0}\frac{dn}{dt} = \left\{ \frac{1}{2} \int_{p^{i}_{O_2}}^{p^{o}_{O_2}} \left(\frac{z_{cat}}{2} D_M + D_O\right) d \ln p_{O_2} \right\} \frac{1}{\Delta x} = k_p^* \frac{1}{\Delta x} \qquad (7.95)$$

Growth of metal-deficient $M_{a-y}O_b$ on M

Let us make use of Eq. 7.95 to illustrate the temperature and oxygen pressure dependence of the parabolic rate constant. Consider the growth of the metal deficient oxide $M_{a-y}O_b$ on high purity M. Let us assume that the predominant defects throughout the entire scale are metal vacancies with an effective charge α, i.e. $V_M^{\alpha'}$. Let us further assume that the metal ion diffusion is much faster than the oxygen diffusion, i.e. $D_M >> D_O$. Eq. 7.95 then reduces to

$$\frac{dx}{dt} = \left\{ \frac{z_{cat}}{4} \int_{p^{i}_{O_2}}^{p^{o}_{O_2}} D_M \, d \ln p_{O_2} \right\} \frac{1}{\Delta x} = k_p^* \frac{1}{\Delta x} \qquad (7.96)$$

In order to integrate this we need to analyse the defect structure of the oxide. The formation of the metal vacancies may be written as

$$\frac{b}{2a}O_2(g) = v_M^{\alpha/} + \alpha h^{\bullet} + \frac{b}{a}O_O^x \qquad (7.97)$$

If other defects may be neglected and the electroneutrality condition is given by $\alpha[V_M^{\alpha'}] = p$, the concentration of metal vacancies is given by

$$\alpha[v_M^{\alpha/}]^{\alpha+1} = K_{V_M^{\alpha/}} p_{O_2}^{b/2a} \qquad (7.98)$$

where $K_{V_M^{\alpha/}}$ is the equilibrium constant for the formation of the metal vacancies (Eq. 7.97) and $[v_M^{\alpha/}]$ denotes the fraction of the metal ion sites that are vacant. On the basis of this relation, $d\ln p_{O_2}$ may be written

$$d\ln p_{O_2} = \frac{2a(\alpha+1)}{b} d\ln[v_M^{\alpha/}] \qquad (7.99)$$

When this expression for $d\ln p_{O_2}$ is introduced in Eq.7.96 , the expression for k_p^* becomes

$$k_p^* = (\alpha+1)D_{v_M^{\alpha/}}\{[v_M^{\alpha/}]^{\mathrm{o}} - [v_M^{\alpha/}]^{\mathrm{i}}\} \qquad (7.100)$$

If the concentration of metal vacancies at the oxide/oxygen interface is much larger than at the metal/oxide interface, i.e. $[v_M^{\alpha/}]^{\mathrm{o}} >> [v_M^{\alpha/}]^{\mathrm{i}}$ then

$$k_p^* = (\alpha+1)D_{v_M^{\alpha/}}[v_M^{\alpha/}]^{\mathrm{o}} = (\alpha+1)D_M^o \qquad (7.101)$$

where $D_M^o = D_{v_M^{\alpha/}}[v_M^{\alpha/}]^{\mathrm{o}}$ is the self-diffusion coefficient of the metal ions in $M_{a-y}O_b$ at the oxide/oxygen interface. Thus under these conditions the parabolic rate constant and the self-diffusion coefficient of M in $M_{a-y}O_b$ has the same temperature and oxygen pressure dependences.

Growth of oxygen-deficient M_aO_{b-y} on M

The same procedure may be used to derive the parabolic rate constant for the growth of oxygen deficient M_aO_{b-y} on M. let us assume that the oxygen vacancies have α charge α, i.e. $V_O^{\alpha\cdot}$, and the formation of the oxygen vacancies may then be written

$$O_O = V_O^{\alpha\cdot} + \alpha e' + \frac{1}{2}O_2 \qquad (7.102)$$

If $D_O >> D_M$ and $(\overset{i}{p}_{O_2} << \overset{o}{p}_{O_2})$ then k_p^* becomes

$$k_p^* = (\alpha+1)D_{v_O^{\alpha\bullet}}[v_O^{\alpha\bullet}]^{i} = (\alpha+1)D_O^{i} = (\alpha+1)D_O^{0}(p_{O_2}^{i})^{\frac{-1}{2(\alpha+1)}} \qquad (7.103)$$

where $D_{v_O^{\alpha\bullet}}$ is the self-diffusion coefficient of the oxygen vacancies, $[v_O^{\alpha\bullet}]^{i}$ the concentration of oxygen vacancies and D_O^i the self-diffusion coefficient of oxygen in M_aO_{b-y} at the metal/oxide interface. D_O^0 is the oxygen self-diffusion coefficient at 1 atm. O_2, and $\overset{i}{p}_{O_2}$ is the oxygen activity at the metal/oxide interface. For this case the parabolic rate constant is independent of the ambient oxygen pressure, and the temperature dependence is given by that of the oxygen self-diffusion coefficient in M_aO_{b-y} at the metal/oxide interface. In this respect it may be further noted that the oxygen self-diffusion coefficient at the metal/oxide interface is given by that the product of the self-diffusion coefficient of oxygen at constant oxygen pressure (at 1 atm O_2) and $(\overset{i}{p}_{O_2})^{-1/2(\alpha+1)}$, i.e. $D_O^i = D_O^0 (\overset{i}{p}_{O_2})^{-1/2(\alpha+1)}$. In these terms the activation energy of the parabolic rate constant is under these conditions given by the activation energy associated with D_O^0 and the enthalpy term associated with $(\overset{i}{p}_{O_2})^{-1/2(\alpha+1)}$.

Similar treatments may be given for oxide scales growing by interstitial metals ions or interstitial oxygen ions.

Scales with ionic conductivity predominant.

Most oxides encountered in high temperature oxidation of metals are electronic conductors. In the literature there are no examples of high temperature metal-oxygen reactions involving conventional metals or alloys with essentially pure ionic conductivity over the entire existence range of the oxide. Such type of reactions are, however, found in metal-halogen reactions, e.g. $Ag+Br_2$ to form AgBr. The same treatment as given above may be applied to reactions involving formation of ionically conducting scales, and the important feature of these reactions is that it is the electronic transport through the scale that is rate-determining.

Varying defect structure situations through the scale.

In the preceding examples it has been assumed that the same defect structure prevails throughout the entire scale from the scale/gas to the metal oxide interfaces. This is an oversimplified model for many systems. The charge on the defects, or even the predominant defects may change with changing oxygen activity. Furthermore, the presence of impurities and dopants may significantly affect the defect structure situation. Thus, following the discussion in Ch.4 one part of the scale may possibly have intrinsic properties and another part extrinsic properties. As part of such behaviour one part of the scale may have significant ionic conductivity while the rest has electronic conductivity.

In the presence of water vapour, hydrogen defects may affect the diffusional behaviour in growing oxide scales in many different ways.

Sintering

Stages of the sintering process

Sintering is the process whereby powders and small particles agglomerate and grow together to form a continuous polycrystalline body. As a rule of thumb the material must be heated to 2/3 of the melting temperature to achieve considerable sintering.

The overall sintering process may conveniently be divided into different stages. The principal stages involved is illustrated schematically below.

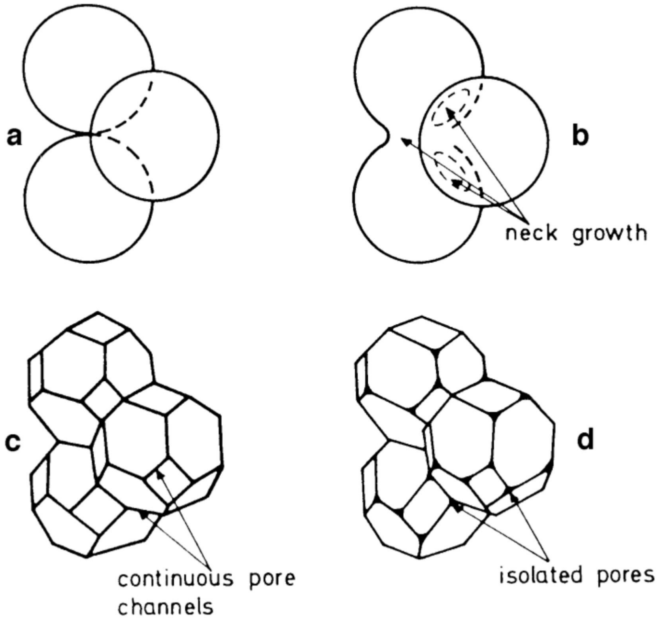

Figure 7-4. Principle stages during the sintering process: a) particles in contact prior to sintering; b) neck growth between particles during initial stage of sintering; c) further neck growth and formation of continuous pore channels along grain boundaries; and d) further sintering in closing channels and formation of closed pores at grain corners.

During an initial stage the surface roughness is decreased and the particles begin to adhere and grow together. At the end of this stage grain growth begins to occur and grain boundaries and metastable pore phase is established. Only modest shrinkage of a few percent takes place during the initial stage.

During an intermediate stage grain growth continues, and the cross-sectional area of the pore phase decreases. The final stage begins when the body achieves 90-95% theoretical density. At this point the pore phase becomes discontinuous, and the closed pores usually remain at grain boundaries. The final stage may involve complete removal of the remaining pores, leading to a completely dense material. Alternatively, it may involve discontinuous growth of the large grains at the expense of the small ones, and closed pores may as a consequence be isolated inside the grains. If the latter process occurs, complete densification becomes extremely difficult.

Driving force for sintering

During sintering, as for other spontaneous processes, the total free energy of the system decreases. For elemental solids and homogeneous compounds the only free energy change involved in sintering is that of the surface free energy or surface tension, γ, of the particles.

Any surface atom or molecule is subjected to a resultant inward attraction because of the unsaturated chemical bonds at the surface. The surface therefore tends to contract to the smallest possible area. In the case of a spherical particle with radius r, the interior is subjected to an excess force of $\pi r^2 p$, where p is the pressure. This is counteracted by the surface tension acting along the circumference: $2\pi r\gamma$. When equating these opposing forces, the pressure is given by

$$p = \frac{2\pi r\gamma}{\pi r^2} = \frac{2\gamma}{r} \tag{7.104}$$

From this it is seen that the vapour pressure of a sphere is larger the smaller the radius and as a result a large sphere will grow at the expense of a neighbouring small sphere. This type of grain growth in polycrystalline solids is termed Ostwald's law. Similarly the vapour pressure of convex surfaces is larger than the vapour pressure of a concave surface and as a result surface roughness of particles is reduced during the initial stage of sintering.

Transport processes during sintering

As for creep, sintering of oxides implies that molecular units of the oxide are transported from areas from high to low surface free energies, e.g. from convex to concave surfaces. Generally it is concluded that transport of the slower moving species determines the overall sintering rate. Thus, in a binary oxide where oxygen ions diffuse much more slowly than the metal ions, it is generally to be expected that the sintering rate is determined by the transport, e.g. lattice or grain boundary diffusion of oxygen.

Sintering governed by lattice diffusion will be dependent upon the concentration of point defects in oxides. Accordingly, sintering rates of an oxide can be optimised by close control of impurities or dopants and the ambient partial pressures of oxygen and of water vapour in cases where proton defects affect the defect structure of the oxide.

Oxides with additional transport of protons

If an oxidic material conducts protons in addition to oxygen ions and electrons we need to introduce electrochemical equilibria between neutral and charged species of both oxygen and hydrogen:

$$O_2(g) + 4e^- = 2O^{2-} \tag{7.105}$$

$$H_2(g) = 2H^+ + 2e^- \tag{7.106}$$

for which equilibria can be expressed by

$$d\mu_{O_2(g)} + 4d\mu_{e^-} = 2d_{O^{2-}} \tag{7.107}$$

$$d\mu_{H_2(g)} = 2d\mu_{H^+} + 2d\mu_{e^-} \tag{7.108}$$

We insert these into the flux equations for the species and utilise that the sum of all transport numbers equals unity. We furthermore use that dμ = kTdlna ≈ kTdlnp, where a and p are activity and partial pressure, respectively, and obtain

$$\frac{d\phi}{dx} = -\frac{i_{tot}}{\sigma_{tot}} + \frac{t_{O^{2-}}kT}{4e}\frac{d\ln p_{O_2}}{dx} - \frac{t_{H^+}kT}{2e}\frac{d\ln p_{H_2}}{dx} + \frac{1}{e}\frac{d\mu_e}{dx} \tag{7.109}$$

This can now be integrated to obtain the voltage of a cell, or inserted into a flux equation of a species of interest in the usual manner. We now have a system with two chemical driving forces for electrochemical transport; that of oxygen activity and that of hydrogen activity.

Other cases

We have in the present version of the treatment of electrochemical transport omitted many cases. These include a more full coverage of proton transport in oxides and transport by other hydrogen species and other foreign species. Moreover, we have left out the case of solid-solid reactions. Also, creep and sintering were given more phenomenological than defect-chemical treatments.

We have throughout assumed isothermal conditions. If we instead of (or in addition to) chemical driving forces considered a temperature gradient, our integrations over membranes would have been from one temperature to another. The voltage we would have obtained would be the thermoelectric voltage. It has several contributions, but in general it is related to the transport numbers and the gradients in concentrations of defects. The variations of concentration is in turn given by the entropy of the defects.

More on these themes will be included in future updates.

Excercises

1. Electrochemical cell with stabilised zirconias (YSZ and CSZ)

a) Consider an yttria-stabilised zirconia (YSZ) material with 90 mol% ZrO_2 and 10 mol% Y_2O_3. What is the site fraction of oxygen vacancies in the material? We often use and refer to the YSZ as an electrolyte. What is an electrolyte? What requirements does it put on a material?

b) We make ourselves a disk of YSZ, attach Pt electrodes to both faces, and place the disk in a cell where we can expose the two faces to different gases. If the gases have different partial pressures of oxygen, a voltage will be set up over the electrodes, and we can draw current from the cell. Write equations for the two half-cell-reactions that run when current is drawn. Write the expression for the voltage over the cell given that the material is a purely ionic conductor. Which (high or low pO_2) is the positive pole?

c) Smith, Meszaros and Amata (*J. Am. Ceram. Soc.,* **49** (1966) 240) found that the permeabilities at high temperatures of Ar and N_2 through calcium-stabilised zirconia (CSZ) were immeasureable, while that of O_2 was significant and proportional to $p_{O2}^{1/4}$. Suggest an explanation for these observations.

2. Electrochemical cell for determination of thermodynamic properties

a) Indicate how you can use a YSZ cell with Pt electrodes to determine the thermodynamic parameters for formation of a binary oxide (e.g. NiO).

b) J.J. Egan (J. Phys. Chem. **68** (1964) 1978) studied an electrochemical cell built up as follows:

- (Mo)Th(s),ThF_4(s)|CaF_2(s)|ThF_4(s),ThC_2(s),C(s)(Mo) +

in order to measure the thermodynamic properties of ThC_2. The Mo serves as inert electrodes. The cell was kept under an inert atmosphere. The CaF_2(s) is a solid electrolyte – a fluoride ion conductor – in the temperature range of operation of the cell (700-900°C). The signs indicate the polarity of the cell. Write the reactions at each electrode when current is drawn from the cell. Also, express the voltage of the cell as a function of the Gibbs energy, enthalpy and entropy of formation of ThC_2.

3. Wagner's oxidation theory

Wagner's equation for the oxidation rate of a metal, forming the oxide M_aO_b, can in one form be written

$$\frac{dn}{dt} = \frac{RT}{8bF^2} \int_1^2 \sigma(t_{cat} + t_{an}) t_{el} d \ln p_{O_2} \frac{1}{\Delta x}$$

a) Specify what the different symbols denote in this equation.

b) Examine the equality of units on both sides of the equation.

c) What are the model and requirements for the theory. Why do the different terms enter the equation as they do?

d) This type of behaviour is called parabolic. What does it reflect? Identify the parabolic rate constant.

e) Which deviations from the requirements of the theory may give rise to non-parabolic (e.g. linear) behaviour?

f) Assume that the oxide has formula MO and is a p-type semiconductor with doubly charged metal vacancies as dominating point defects ($D_M >> D_O$). Assume that $pO_2^o >> pO_2^i$ and find how the oxidation rate varies with the outer oxygen pressure pO_2^o.

g) Assume that the oxide has formula M_aO_b and that it is an n-type semiconductor with $M_i^{\cdot\cdot\cdot}$ as dominating point defects and that $D_M >> D_O$. How does the oxidation rate in this case vary with the outer oxygen partial pressure (assuming also now that $pO_2^o >> pO_2^i$)?

h) The tracer self-diffusion coefficient for Co in CoO in air at 1200°C is $9*10^{-9}$ cm^2/s. Calculate the parabolic rate constant in cm^2/s (which is k_p in $dx/dt = k_p/x$) when cobalt oxidizes to CoO under the same conditions. The dominating point defects in CoO are assumed to be singly charged metal vacancies v_{Co}'.

4. Sintering

a) What happens during sintering? What are the driving forces? What are the contributing mechanisms and what may be the rate limiting factors?

b) Powders of CoO can be sintered to dense samples. The defect structure of CoO is characterized by the following: There is a deficiency of metal and the dominating point defects are singly charged metal vacancies. $D_{Co} >> D_O$.Assume that the oxygen defects are interstitials with one negative effective charge. Draw a Brouwer-diagram for the oxide. Under what conditions should the oxide be sintered to achieve the highest sintering rate?

5. Creep

a) What is creep? What are the driving forces, mechanisms, and rate limiting factors?

b) TiO_2 is an n-type electronic conductor with oxygen deficiency, dominated by doubly charged oxygen vacancies and triply charged titanium interstitials. Assume that the concentration of oxygen vacancies is 1000 times higher than that of titanium interstitials at $pO_2 = 1$ atm. Draw a Brouwer diagram for the oxide from $pO_2 = 10^{-20}$ to 1 atm. How would you qualitatively expect that the creep of TiO_2 single crystals varies with pO_2?

6. General considerations

b. The expression "uphill diffusion" is used for diffusion against a concentration gradient, referring to Fick's 1st law. From what you know about electrochemical transport, what might cause uphill diffusion?

c. Eq. 7.4 is written with conductivity as the transport coefficient. Flux of a neutral species is not affected by the electrical potential gradient and can not be expressed by conductivity. It is therefore meaningful to express the equation using the random diffusion coefficient for the chemical potential gradient and the conductivity for the electrical potential gradient. Do this splitting/substitution. (This is mainly a simple exercise in the Nernst-Einstein relation). Check that both parts of the resulting equation will have units that confer with flux density.

7. Transport number and partial conductivity measurements

d. A membrane of a mixed oxygen ion and electron conducting oxide is equipped with Pt electrodes and exposed to two different oxygen partial pressures: On one side is dry air ($p_{O2} = 0.21$ atm) and on the other is pure oxygen ($p_{O2} = 1.00$ atm). What is the Nernst voltage of this cell at 1000°C?

e. A voltage of 0.01 V is read over the cell at 1000°C using a high impedance voltmeter. What is the average ionic transport number of the oxide under these conditions?

f. Make a sketch of the sample and gases and external circuitry during the voltage measurement. Indicate internal and external fluxes and currents (with directions).

g. The total conductivity of the sample is measured and the result is 0.01 S/cm. What are the partial ionic and electronic conductivities?

h. Assuming that the sample is 0.5 mm thick, what is the flux of oxygen gas through the sample during the voltage measurements?

i. If you had measured this flux in some way, and used this as your (only) transport measurement, what would be the transport parameter that you find?

8. Sensor

a) A sensor is made out of a membrane of yttria.stabilized zirconia (YSZ) with Pt electrodes and is operated at 700°C. As reference is used pure oxygen at 1 atm. A voltage of -950 mV is read over the sensor (measuring on the "unknown" side vs the reference side). What is p_{O2} at the "unknown" side?

b) If you assume that the "unknown" side is hydrogen gas at 1 atm, what does the sensor voltage say about the water vapour content of that hydrogen gas?

9. Fuel cell

a) A fuel cell made of YSZ electrolyte with Pt electrodes is operated at 700°C and runs on hydrogen as fuel and air as oxidant, both at atmospheric pressure. Both gases contain 2% water vapour. What is the open circuit voltage of the fuel cell?

b) Assuming that the electrolyte has a conductivity of 0.1 S/cm and is 100 μm thick, calculate the short-circuit current density, the maximum power density and the current and power densities at a voltage of 2/3 of the open circuit voltage. (Assume in all cases that the gas compositions remain unaffected by the current in the cell, and that electrode polarization is insignificant).

c) Assuming the operation at 2/3 of the OCV voltage, what is the area of electrolyte needed for a power plant of 300 MW capacity?

d) The assumption that the gas compositions are unaffected is not realistic for a real power plant; what would it mean in terms of fuel utilization?

e) The system we have described can be operated at other temperatures. Still neglecting electrode polarization, what are the factors and processes that enter into the temperature dependency of the power density of the fuel cell? (Hint: refers to defects and transport.)

10. Gas separation membranes

a)

A gas separation membrane that can separate oxygen from air is constructed from La_2NiO_{4+d}, a material where we for simplicity may assume that the defect structure is dominated by doubly ionized oxygen interstitials compensated by electron holes. In the following assume further that the cell is operated at 1000°C and that the membrane is 100 μm thick and has no surface kinetics limitations. Assume also that the oxygen ion conductivity (by interstitials) is 1 S/cm at 1 atm O_2 and that the electronic (hole) conductivity is 200 S/cm at 1 atm O_2. The membrane is operated with atmospheric air on one side and pumped to effectively 0.1 atm at the other side. What is the area-specific flux of oxygen?

b)

What would increase the flux more: Compress the air to 10 atm or pump the vacuum side to 0.01 atm?

1. Find an expression for the chemical diffusion coefficient for oxygen ions.
2. During operation the membrane may possibly suffer from both membrane "walk-out" and decomposition. Explain both phenomena qualitatively.

c) Solid –solid reactions

1. La_2NiO_4 may be synthesized by solid-state reaction between La_2O_3 and NiO. Imagine that you investigate this reaction by making tablets of the two reactants and holding them together at high temperature. Sketch how the product La_2NiO_4 is formed as a phase between the reactants and what diffusion process (or processes) that take place to constitute the reaction.
2. Suggest how you might use a marker (e.g. a Pt wire) to identify the rate limiting diffusion process.
3. The reaction rate can be measured as the thickness of the product layer. It would be expected to grow parabolically. Explain.

d) High temperature oxidation of metals

1. Zr metal is used at high temperatures in nuclear installations. It is protected by an oxide layer of ZrO_{2-y}, dominated by oxygen vacancies and electrons.
2. What is rate limiting for the parabolic oxidation process in your opinion?
3. Nb is often used as allyoing element in Zr in order to reduce oxidation of the metal. Explain how this dopant might work when it ends up in the oxide scale.
4. In nuclear reactors it is a problem that hydrogen diffuses as protons through the oxide scale, dissolves in the Zr alloy and makes it brittle. How would you expect Nb alloying to affect the proton transport through the oxide scale?
5. Ni oxidizes to $Ni_{1-y}O$, i.e. an oxide dominated with Ni vacancies and electron holes. How would you suggest to dope this oxide to reduce oxidation?

Tables

Greek letters

Α	α	alfa
Β	β	beta
Γ	γ	gamma
Δ	δ	delta
Ε	ε	epsilon
Ζ	ζ	zeta
Η	η	eta
Θ	θ	theta
Ι	ι	iota
Κ	κ	kappa
Λ	λ	lambda
Μ	μ	my
Ν	ν	ny
Ξ	ξ	ksi
Ο	ο	omikron
Π	π	pi
Ρ	ρ	rho
Σ	σ	sigma
Τ	τ	tau
Υ	υ	ypsilon
Φ	φ	phi
Χ	χ	khi
Ψ	ψ	psi
Ω	ω	omega

Constants

Gravitational acceleration	g	9,80665	m/s^2
Atomic mass unit	u	$1,6605*10^{-27}$	kg
Electron mass	m_e	$9,110*10^{-31}$	kg
Elementary charge	e	$1,602*10^{-19}$	C
Electron volt	eV	$1,602*10^{-19}$	J
Proton mass	m_p	$1,673*10^{-27}$	kg
Neuytron mass	m_n	$1,675*10^{-27}$	kg
Speed of light in vacuum	c	$2,99792*10^{8}$	m/s
Boltzmann's constant	k	$1,381*10^{-23}$	J/K
		$8,6174*10^{-5}$	eV/K
Planck's constant	h	$6,626*10^{-34}$	Js
Rydberg's constant	$\mathfrak{R}$	$1,097*10^{7}$	m^{-1}
Bohr's constant	B	$2,18*10^{-18}$	J
Avogadro's number	N_A	$6,022*10^{23}$	mol^{-1}
Gas constant	R	8,31451	J/molK
		0,0820578	Latm/molK
Faraday's constant	F	96485	C/mol
Gravitational constant	γ	$6,672*10^{-11}$	Nm^2/kg^2
Permeability in vacuum	μ_0	$1,257*10^{-12}$	H/m
Permittivity in vacuum	ε_0	$8,854*10^{-12}$	F/m
Electrical constant $=(4\pi\varepsilon_0)^{-1}$	k_e	$9,0*10^{9}$	Nm^2/C^2
Magnetic constant	k_m	$2*10^{-7}$	N/A^2
Stefan-Boltzmann constant	σ	$5,67*10^{-8}$	W/m^2K^4
Wien's constant	a	$2,90*10^{-3}$	mK
Volum of 1 mol ideal gas	$V_{m,298K}$	24,4651	L