Summer 8-31-2003

Understanding the origins of metastability in thin film growth; tantalum and the early group VB-VIB metals

Neme Okechukwu Nnolim
New Jersey Institute of Technology

ABSTRACT

UNDERSTANDING THE ORIGINS OF METASTABILITY IN THIN FILM GROWTH; TANTALUM AND THE EARLY GROUP VB-VIB METALS

by
Neme Okechukwu Nnolim

Tantalum (Ta) is a metal that is highly prized for its applications in a variety of industries, including the microelectronics industry, where it is largely used in thin film modifications in order to achieve various electronic, magnetic and structural capabilities in solid-state devices. Ta frequently forms metastable phases, including the well-known β-phase, during the preparation of thin Ta films by standard film deposition methods such as sputtering and electro-deposition. In order to gain insight into Ta metastable phase formation, the Bain transformation mechanism is studied for Ta and the neighboring body centered cubic (bcc) transition metals in Groups 5 and 6 of the Periodic Table, resulting in the prediction of hypothetical body centered tetragonal (bct) and face centered cubic (fcc) phases for all the studied metals at c/a ratios greater than 1.6 and equal to $\sqrt{2}$ respectively.

Density Functional Theory (DFT) in the Linearized Augmented Plane Wave (LAPW) formalism and the Lowest Order Variational Approximation (LOVA) for the solution of the Boltzmann transport equation have been used to calculate the temperature dependent resistivity for the studied metals and reveal that substrate-induced film strain is not by itself sufficient to cause the formation of β-Ta. Fermi surface calculations of the mean electron velocity in the proposed $P4_2/mnm$ structure of β-Ta suggest that the high values of resistivity frequently measured for β-Ta are not an intrinsic property.

UNDERSTANDING THE ORIGINS OF METASTABILITY IN THIN FILM GROWTH; TANTALUM AND THE EARLY GROUP VB-VIB METALS

by
Neme Okechukwu Nnolim

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey-Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Applied Physics

Federated Physics Department

August 2003

APPROVAL PAGE

UNDERSTANDING THE ORIGINS OF METASTABILITY IN THIN FILM GROWTH; TANTALUM AND THE EARLY GROUP VB-VIB METALS

Neme Okechukwu Nnolim

Dr. Trevor A. Tyson, Dissertation Advisor — Date
Associate Professor of Physics, NJIT

Dr. Kenneth Farmer, Committee Member — Date
Associate Professor of Physics, NJIT

Dr. Sergej Savrasov, Committee Member — Date
Assistant Professor of Physics, NJIT

Dr. Michael C. Y. Huang, Committee Member — Date
Assistant Professor of Chemical Engineering, NJIT

Dr. Lisa Axe, Committee Member — Date
Associate Professor of Civil Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Neme Okechukwu Nnolim

Degree: Doctor of Philosophy

Date: August 2003

Undergraduate and Graduate Education:

- Doctor of Philosophy in Materials Science and Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2003

- Master of Science in Physical Chemistry,
  Massachusetts Institute of Technology, Cambridge, MA, 1996

- Bachelor of Science in Chemical Engineering,
  University of Rochester, Rochester, NY, 1992

Major: Materials Science and Engineering

Presentations and Publications:

Neme O. Nnolim, Trevor A. Tyson and Lisa Axe,
"Theory of the structural phases of Group 5B – 6B metals and their transport properties", Journal of Applied Physics, Volume 93, Number 8, pages 4543 – 4560, April 2003.

Neme O. Nnolim, Trevor A. Tyson and Lisa Axe,
"A study of the structural phases of the Group 5B – 6B elements and their mechanical properties", Meeting of the Materials Research Society, Boston, MA, December 2002.

Neme O. Nnolim, Trevor A. Tyson and Lisa Axe,
"Determination of metastable structural phases in column 5B – 6B metals by total energy computations", Meeting of the American Physical Society, Indianapolis, IN, March 2002.

I would like to express my deepest gratitude and appreciation to my dear wife Charlotte Messina-Nnolim and my dear family including my mother Dorothy Adaku Nnolim, my father Benedict Nwankwo Nnolim, my brothers Chukwuma Nnolim and Uche Nnolim, my sister Chiemedinam Nnolim, my Aunt Guy Ilobi, my cousin Chudi Ejekam, my sister-in-law Pamela Messina and my uncle Geoffrey Nnolim for their constant encouragement, support and advice. I could not have done this without them.

Onye ria osisi oji, ya jisike kpaa ya nku ka o nwere ike: anaghi ari enu osisi oji kwa ubosi

(When one finds oneself atop a tall and mighty tree gathering wood, fatigue is not an option)

IGBO PROVERB

ACKNOWLEDGMENT

I would like to express my sincere appreciation and gratitude to my research supervisor, Dr. Trevor Tyson, who in instances too numerous to count gave me invaluable advice, support and scientific insight and who cheerfully provided resources whenever they were needed. Special thanks are due to Dr. Michael C. Y. Huang, Dr. Sergej Savrasov, Dr. Lisa Axe, and Dr. Kenneth Farmer for being active members of my committee, and to David Perel of the NJIT Computing Services Division for his assistance over the years.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

The purpose of this dissertation is to gain insight into metastable phase formation in tantalum thin films by studying the structure dependence of transport properties such as the electrical resistivity and the Fermi velocity. Special attention is paid to β-tantalum, a metastable phase with seemingly peculiar resistivity behavior that is readily formed in thin tantalum films. A wider range of structures, including metastable structures, is available to materials in thin film form due to energetic considerations having their origins in the electronic structure of materials. The metastable crystal structures are most easily formed in thin films and are materials with lattice parameters which result in a local minimum in the total energy per unit cell, and which are separated from the structure representing the absolute energy minimum by an activation energy. They are readily prepared using thin film deposition techniques such as sputtering and ultra-high vacuum evaporation and exhibit electronic, magnetic and structural properties markedly different from their bulk material precursors.

The rich variety of structural and electronic properties observed in metastable phases depend on an often bewildering number of possible unique preparation environments which are due to various combinations and permutations of a smaller number of controllable process parameters. Difficulties often arise when attempting to isolate the effects of individual process parameters because of the large number of parameters involved, whose effects on the creation of unique preparation environments for thin film growth resulting in the formation of metastable phases are often inextricably linked. The most important factors in transition metal film growth include the strain

energies induced by any mismatch between the film and substrate lattices as well as the corresponding surface free energy differences. With the realization that the origins of metastable phase formation are rooted in electronic structure phenomena, the problem of understanding it is made more tractable.

This work aims to use a theoretical approach grounded in quantum mechanics in order to predict the metastable phases of some of the most important transition metals located in groups VB-VIB of the Periodic Table, which include vanadium, niobium, chromium, molybdenum, tungsten and tantalum, as well as their electrical and transport properties in order to gain insight into some processes by which metastable phases of these materials are formed. This study was motivated by an attempt to understand the origins of metastable β-tantalum phase formation during the sputter deposition of tantalum protective coatings on steel substrates. In addition to having desirable properties such as resistance to high-temperature fatigue and oxidation, tantalum possesses a whole range of other desirable properties important in a variety of applications. In the following sections, the discovery of tantalum, its extraction by various means and the technological developments involved in its processing for various applications is discussed.

1.1 Discovery and Occurrence of Tantalum

Tantalum metal was discovered by Anders Ekeberg in 1802 [1] who found the element in two different minerals, tantalite from the town of Kimito in Finland and yttrotantalite from the town of Ytterby in Sweden. Yttrotantalite is a rare earth tantalate consisting of high concentrations of yttrium and tantalum. Prolonged experiments on the minerals eventually resulted in the isolation of a substance then believed to be a new metal oxide

and named "tantalum" because of the difficulty involved in dissolving the mineral with acids and isolating the tantalum oxide. In Greek mythology, Tantalus was the son of Jupiter and also the king of Phrygia who was condemned for eternity to stand up to his neck in water while being unable to drink it because the water receded every time he tried to drink from it.

Berzelius, a student of Ekeberg, was the first to prepare tantalum metal powder in 1825 [1]. Berzelius prepared the tantalum powder, which contained a high concentration of impurities, by reducing potassium heptafluorotantalate (K_2TaF_7) with potassium metal under a salt layer in an iron crucible. Tantalum was confused with columbium when in 1809 William Hyde Wollaston declared that columbium and tantalum were the same element.

In 1840, Heinrich Rose, a pupil of Berzelius, established that two similar metallic elements occurred in a variety of columbite and tantalite deposits, which he named tantalum and niobium [1]. However, Rose mixed up the names and assigned the name niobium to tantalum. The differences between niobium and tantalum were eventually well established by the works of Charles Broomstand and Jean Charles Galissard de Marignac who differentiated between the metals based on their vapor pressures. The first successful extraction and separation of tantalum from niobium is due to de Marignac [1], who did not obtain pure niobium, but obtained it combined with titanium impurities. The process invented by de Marignac to separate the metals exploited the different solubilities of tantalum and niobium fluorides in dilute acid solution. The de Marignac process consisted of pulverizing the tantalum and niobium ores, and then fusing the pulverized ore with sodium hydroxide in an iron crucible. The fused product was then leached with

water after which it was filtered to isolate the insoluble sodium niobate and tantalate compounds. The tantalates and niobates were converted into hydroxides by digesting the filtered product in hydrochloric acid, producing a sodium chloride waste stream. Dissolving the hydroxides in hydrochloric acid resulted in a solution containing a mixture of heptafluorotantalic acid and oxypentafluoroniobic acid.

The solubilities of the potassium salts of the two different acids were discovered by de Marignac to be different at 15°C, the tantalum salt having a solubility about 12 times less than the solubility of the niobium salt [1]. Adding potassium chloride or potassium fluoride to the mixed acid solution resulted in the potassium tantalum fluoride crystallizing out of solution, while the niobium salt remained in solution. The potassium tantalum fluoride salt so obtained was then an easier raw material to convert to tantalum metal.

In 1866, Rose produced a tantalum powder purer than that produced by Berzelius by reducing Na_2TaF_7 with sodium metal [1]. However, the tantalum powder so produced, due to the high temperatures required for the reduction reaction, was probably high in oxygen content. Between 1900 and 1903, Dr. Werner von Bolton and Dr. Otto Feuerlein, working for Siemens and Halske AG in Berlin, were able to produce, by carbon reduction of tantalum oxide, tantalum powder pure enough to be used in the manufacture of tantalum sheet and wire, leading to the use of tantalum wire in lighting devices and therefore the first commercial application of tantalum [1].

Tantalum is 54^{th} in order of abundance of the elements in the Earth's crust, and is estimated to make up about 2×10^{-4} wt% of the earth's crust [2]. It is usually found naturally occurring together with niobium in many deposits [1, 3]. The most important

tantalum–containing minerals are tantalite, wodginite, microlite and columbite [2]. Typically, tantalum pentoxide (Ta_2O_5) is found at concentrations less than approximately 0.1% in processable pegmatic deposits, which represent the major raw material for the production of tantalum. Other important tantalum containing minerals are euxenite, yttrotantalite, samarskite and strueverite. These other minerals often contain uranium and thorium as well [4] and are highly radioactive as a result, limiting the extent to which they can be safely worked to produce tantalum.

Most of the world's tantalum reserves occur in Australia, Brazil, Canada, Congo and Nigeria [5], and these countries are among the major producers of tantalum. An important potential reserve consisting of casserite (SnO_2), which contains tantalum and niobium, lies in a regional belt encompassing China, Indonesia, Thailand and Malaysia [1]. Other countries such as Bolivia, Russia and Zambia also produce tantalum, but not in any significant measure of total global annual output. The significant deposits of tantalum occurring worldwide are listed in Table 1.1.

Table 1.1 Tantalum Deposits Worldwide

Location	Reserves (10^3 kg Ta_2O_5)	Production (1991 Estimated) (10^3 kg Ta_2O_5)	Annual Production Capacity (Estimated) (10^3 kg Ta_2O_5)	Mine Production (2000)	Mine Production (2001)
Australia	17 000	235	300	485	640
Brazil	18 000	100	245	90	300
Canada	3 000	80	130	57	50
Malaysia	2 000	80	130		
Thailand	16 000	150	200		
Africa*	10 000	25	50	172	94
Russia	> 20 000	< 50			
China	17 000	100	200		
Others	10 000	< 50	< 100		

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).
*includes Congo, Ethiopia and Nigeria.

The chemical compositions of tantalum-containing minerals are shown in Table 1.2.

Table 1.2 Composition of Important Tantalum-containing Minerals

Mineral	Composition	Ta_2O_5 Content, by % Weight	Nb_2O_5 Content, by % Weight
Tantalite	$(Fe, Mn)Ta_2O_6$	42 – 84	2 – 40
Microlite	$(Na, Ca)(Ta, Nb)_2O_6F$	60 - 70	5 – 10
Columbite	$(Fe, Mn)(Nb, Ta)_2O_6$	1 - 40	40 – 75
Wodginite	$(Ta, Nb, Sn, Mn, Fe, Mn)_{16}O_{32}$	45 – 56	3 - 15
Yttrotantalite	$(Y, U, Ca)(Ta, Fe^{3+})_2O_6$	14 – 27	41 - 56
Fergusonite	$(RE^{3+})(Nb, Ta)O_4$*	4 – 43	14 – 46
Strueverite	$(Ti, Ta, Nb, Fe)_2O_4$	7 – 13	9 – 14
Tapiolite	$(Fe, Mn)(Nb, Ta)_2O_6$	40 – 85	8 – 15
Euxenite	$(Y, Ca, Ce, U, Th)(Nb, Ta, Ti)_2O_6$	1 – 6	22 – 30
Samarskite	$(Fe, Ca, U, Y, Ce)_2(Nb, Ta)_2O_6$	15 – 30	40 - 55

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).
*RE – Rare earth element.

The worldwide reserves of tantalum ore are shown categorized according to major geographical regions in Figure 1.1.

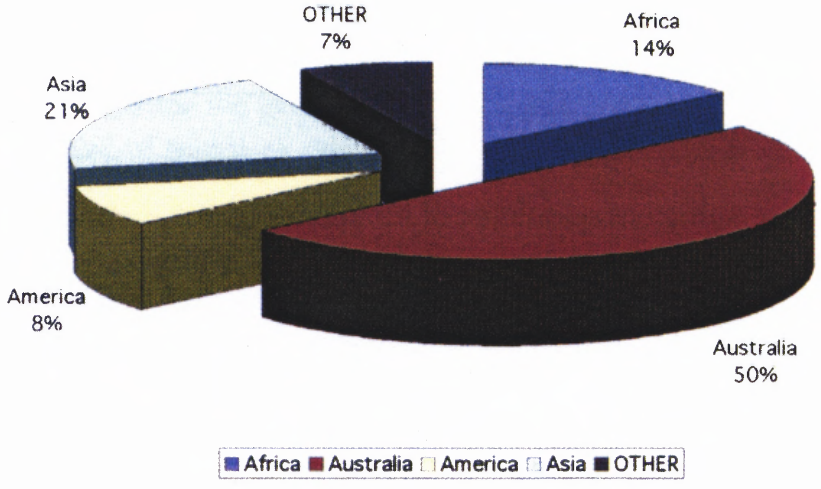


Figure 1.1 Worldwide reserves of tantalum ore in 2000 (taken from Reference [1]).

1.2 Extraction of Tantalum

Tantalum-containing ores are usually present in hard pegmatitic rock in finely divided form [2]. The hard rock is extracted by open-cast mining. Processing consists of an initial

particle size reduction to < 15 mm, followed by grinding in ball [6] or rod mills to < 1 mm. The soluble constituents are separated out from the concentrates by leaching, in a process called flotation. The concentrates are dried, and the magnetic constituents are removed [7]. The ore concentrate obtained in this manner, which can contain up to 40% Ta_2O_5, can be dissolved in several ways such as by agglomeration with an alkali metal hydroxide or carbonate, treatment with strong acids, digestion with acids and acidic mixtures, and chlorination [8].

The most significant industrial method of tantalum ore degeneration is by dissolution in concentrated hydrofluoric acid (HF) and then separation from the existing niobium by extraction with methyl isobutyl ketone [9], leading to a final tantalum product of H_2TaF_7. K_2TaF_7 is crystallized out of H_2TaF_7 by adding potassium salts [10]. K_2TaF_7 is then reduced with sodium to obtain metallic tantalum. Tantalum is a strongly electronegative metal; therefore electro-refining in molten salts consisting of K_2TaF_7-NaCl-NaF is a widely used method to refine it. Recently highly pure (highest elemental impurity level of 0.18 at.%) samples of tantalum metal have been obtained [10] by electro-refining in a NaCl-KCl-CsCl-K_2TaF_7-NaF melt. The major steps involved in the processing of tantalum, from its extraction in ore form to the production of finished products, are shown in Figure 1.2.

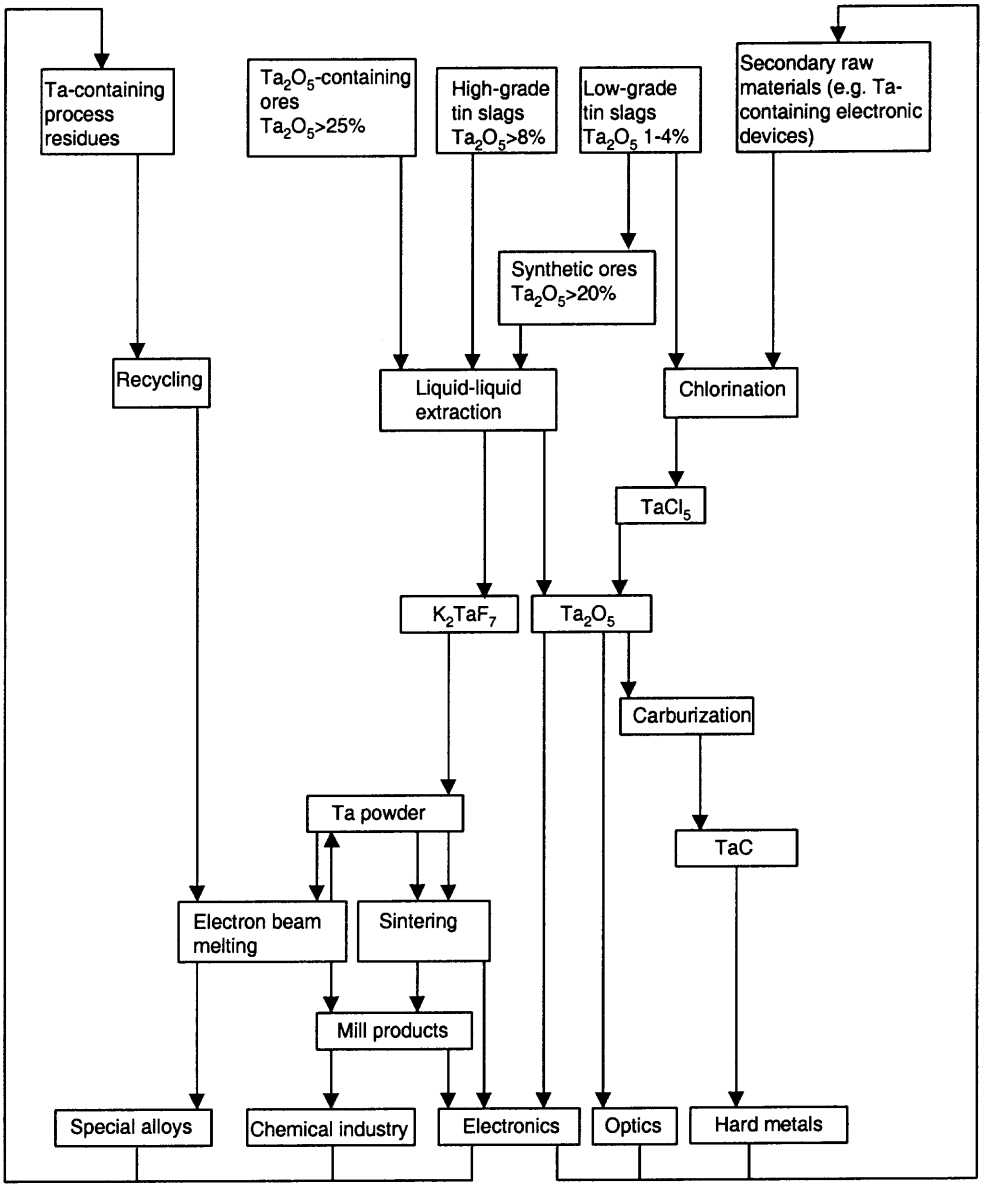


Figure 1.2 Major tantalum processing routes (taken from Reference [2]).

1.3 Important Tantalum Sources

In 1980, about 50% of the world demand for tantalum was obtained from tin slags [2]. The low price of tin due to the lowered demand for tin over the past two decades has led to reductions in tin mining and therefore a decline in the availability of tantalum-containing tin slags. The fraction of tantalum world requirements satisfied from Ta_2O_5 -

containing tin slags is expected to fluctuate around 20%, because the low price of tin does not permit the economical processing of the tantalum-rich tin ores from Malaysia and Thailand. The most important slag grades and their tantalum content are given in Table 1.3.

Table 1.3 Tantalum Content of the Most Important Slag Grades

	% Ta_2O_5 Content by Weight	% Nb_2O_5 Content by Weight
High Grade		
Thailand	13–16	9–10
Zaire	11	10
Australia	10	5-6
Medium Grade		
Thailand	4–5	4
Malaysia	3–4	3–4
Nigeria	4–5	12
South Africa	6–8	8–9
Low Grade		
Malaysia	0.8–2.5	0.8–2.5
Thailand	0.5–1	0.2–0.8
Brazil	1–2.5	1–2.5
Singapore	1–1.5	1–1.5

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).

There are two kinds of tantalum deposits: placer deposits and hard-rock deposits. Placer deposits are usually cheaper to recover because they can be naturally concentrated and more easily mined. Placer deposits can be brought to market in about a year. Hard-rock mining requires large capital outlays and usually involves deep-hole or open-pit mining. Products from hard-rock mines are usually brought to market in a period of two to three years. Most new tantalum deposits are of the hard-rock category [1]. Hard-rock mines located in Australia are the largest active tantalum mines. There is another class of tantalum deposits, still untouched, that is expensive to develop and which is referred to as

carbonatite. These deposits consist of massive bodies of ore with high complex mineral content. Tantalum ores are classified as either reserves or deposits depending on the degree of exploration of the ores and their commercial value. Commercial deposits are deposits for which the prevailing market value, which is highly dependent on market fluctuations, can be realized. Commercial deposits and reserves usually comprise less than a 30-year supply. Total commercial tantalum reserves worldwide are estimated to consist of 36,400 metric tons of tantalum, enough to satisfy world tantalum requirements for 16 years [1].

The largest tantalum mines in the world are the Greenbushes and Wodgina mines, which are both located in Australia and owned by a commercial concern known as Sons of Gwalia Limited. Resources are defined as raw material deposits that may not presently be of commercial value. Tantalum resources worldwide are estimated at 287,000 metric tons, which is enough to satisfy world tantalum demand for a period of over 125 years, assuming a worldwide tantalum consumption rate of 5 million pounds a year. The major share of worldwide tantalum resources is located in Australia, Asia, Africa and the Americas, with most of the resources being located in Australia. Compared with copper, which is a major raw material for the electronics industry and whose deposits have been estimated to be able to satisfy worldwide copper demands for the next 63 years, tantalum is a more dependable raw material. The United States Defense Logistics Agency (DLA), whose purpose was to provide a stable supply of important materials in case of war, has been buying, storing and selling the US government tantalum stockpile for many years. The DLA is estimated to have 580 metric tons of tantalum in its inventory [1].

In 1998, world mines produced approximately 2 million pounds of tantalum a year. World tantalum output increased to over 3 million pounds by the year 2002 and by the end of 2003, the mines could possibly produce more than 5.5 million pounds if the demand arises [11]. The 5 million pounds shipped to tantalum capacitor manufacturers in the year 2000 was greater than the 3 million pounds of tantalum produced by world mines, therefore a significant amount of tantalum came from DLA, miner and processor inventories, as well as from tantalum scrap recycling, which increased from 25% to 30%. The greatest supplier of tantalum ore, Sons of Gwalia Limited, increased its production capacity to over 2 million pounds per year due to recent shortages in the tantalum supply chain.

High-concentration slags are used directly in the wet chemical process that consists of dissolution of the slag in HF/sulphuric acid (H_2SO_4), followed by extraction and separation of tantalum and niobium [2, 12]. Low concentration slags are treated pyrometallurgically to produce concentrates containing 50 – 60% Ta_2O_5/Nb_2O_5, which are then dissolved in HF. In the majority of the ore-concentration processes, the slag is first melted in an electric furnace with the addition of reducing agents such as coke. The impurities are left in the slag, while the tantalum and niobium collect in a carbide containing ferroalloy. The solid ferroalloy is produced in relatively large blocks, which are then broken up and ground. A concentrate containing 50 - 60% oxides of tantalum and niobium is obtained by further processing, which sometimes entails the oxidation of the ferroalloy with Fe_2O_3 as hematite. This process concentrates undesired elements in the oxidic slag, while the Ta-Nb carbide contained in the ferroalloy is minimally oxidized and remains in the metallic state. Finally, the Ta-Nb carbide is again oxidized with Fe_2O_3

to produce a slag rich in tantalum and niobium oxides. A single smelting stage using air, pure oxygen and oxygen-enriched air could also be used to convert the tantalum to slag at a much faster rate than niobium, therefore concentrating it at the expense of that metal. Wet leaching processes can also be used to treat the Ta-Nb carbide-containing ferroalloy [2].

Physical and chemical processes are used extensively when processing waste materials containing tantalum. Tantalum scrap is usually divided into two categories according to difficulties associated with recycling the scrap material [2]. The first kind of tantalum scrap can consist of pure unoxidized metallic tantalum, such as scrap foil, sheet, wire, anodically oxidized sintered tantalum pellets containing 1.5 – 3% oxygen from capacitor manufacture and sintered tantalum powder pellets from the production of tantalum electrolytic capacitors. For these scrap materials, the following processing methods are used: ingot melting in an electron beam furnace, conversion to brittle tantalum hydride, grinding, and dehydrogenation at > 600°C under vacuum or under a protective gas to form metallic tantalum powder. The oxygen content of oxidized tantalum anodes can be reduced by deoxidation with magnesium or calcium to produce high-grade scrap.

The second kind of recyclable tantalum scrap consists of oxidized tantalum anodes with manganese dioxide or conductive silver coatings, or tantalum anodes welded to nickel conductor wires [2]. Included in this category are faulty capacitors, in which the coated and welded tantalum anodes are covered in tinned brass or synthetic resin. Resin-covered capacitors can be subjected to grinding mills for size reduction, and then sifted

for tantalum by gravity. Treatment with nitric acid (HNO_3) and HF removes foreign metals to leave a residue of oxidized tantalum anodes.

Tantalum scrap containing other impurities are roasted to form tantalum oxides, and then treated like ore concentrates. Condensates formed during the melting of tantalum ingots, impure sawings and turnings, flue dust, scrapped alloy materials and sediments from the wash liquors arising from the sodium reduction of K_2TaF_7 are included in this category. As described above, the conversion of tantalum scrap to tantalum oxides is a major step in the recovery treatment process. Tantalum carbide is an intermediate material in this process and can be formed from ground tantalum by carburization [13]. Hard metals containing tantalum carbide can then be roasted and treated with caustic soda solution, melting with sodium nitrate and anodic oxidation [2] to achieve the conversion to oxide.

1.4 Commercial Manufacture of Tantalum

The reduction of K_2TaF_7 is the method used to produce the large majority of metallic tantalum products such as ingots, powder and semifinished products [2]. The first successful deposition of tantalum by electroplating from molten fluorides containing K_2TaF_7 was achieved by Senderoff and Mellors in 1965 [14]. The chemical reaction involved in the reduction of K_2TaF_7 to produce tantalum is [2]

$$K_2TaF_7 + 5Na \rightarrow 2KF + 5NaF + Ta. \qquad (1.4.1)$$

The reaction is strongly exothermic and is controlled with inert salts such as KCl, NaCl, KF and NaF. The reduction is carried out at constant temperature by controlled heat

removal, which is achieved by the controlled addition of one of the reactants. The process of leaching is used to remove the various salts from the reaction product in order to recover tantalum powder.

Tantalum capacitor powder is obtained from the tantalum powder by two established processes. In one process, the metallic tantalum powder which has been sodium reduced is purified and converted to a compressible and free-flowing product by further processing steps, such as high temperature treatment under vacuum or protective gas, or magnesium deoxidation, or a combination of the above two processes. The addition of dopants [15] to control the sintering of porous pellets is used to obtain optimal properties. In the other established process, ingot melting of compacted slugs of tantalum powder in electron beam or electric arc furnaces cause volatile impurities to evaporate away. The ingot can then be ground after hydrogenation. The powdered hydrogen rich tantalum is then dehydrogenated under vacuum [2] and agglomerated by high-temperature treatment.

Other processes exist for the production of metallic tantalum in addition to the sodium reduction process, such as the molten salt electrolysis of K_2TaF_7 [10] and hydrogen reduction of $TaCl_5$ [16], but these processes have not yet been made economically viable. For example the carbothermic reduction of tantalum pentoxide under high vacuum at approximately 1900°C first produces tantalum carbide [2]

$$Ta_2O_5 + 2C \rightarrow 2TaC + 5/2O_2. \qquad (1.4.2)$$

The produced tantalum carbide then reacts above 2000°C with residual Ta_2O_5 to form metallic tantalum [2]

$$5TaC + Ta_2O_5 \rightarrow 7Ta + 5CO. \quad (1.4.3)$$

This process is not economically important because of the high quality standards for particle size and purity required by manufacturers of tantalum capacitors [2].

Semi-finished tantalum products such as sheet, wire and shaped articles are produced from tantalum metal power, whose production by sodium reduction has been described above. Semi-finished tantalum products are produced from tantalum powder in two ways. In the first method of producing tantalum powder, the tantalum powder is compacted and then sintered by the passage of an electric current at approximately 2000°C or by resistance/induction heating under vacuum. Tantalum wires used for the manufacture of tantalum electrolytic capacitors are produced in this manner. Adding dopants to the powder during the process controls the grain size in order to optimize the mechanical properties. In the second method of tantalum powder production, powder compression is used to produce bars which are then formed into ingots by electron beam or electric arc melting, and shaped by forging, rolling, extruding and wire drawing. A tantalum product with an initial ingot hardness of 80 – 90 HV has a typical hardness of 180 HV after 95% deformation [2]. The commercial importance of tantalum may be appreciated by the fact that the most recent published industry source on tantalum product prices [5] indicated the following selling prices per pound tantalum content for the following tantalum products: capacitor grade powder, $135 - $260, capacitor wire, $180 - $270, and vacuum grade tantalum metal for superalloys, $75 - $100.

1.5 Mechanical and Thermodynamic Properties of Tantalum

Tantalum is a dense refractory metal that is grayish silver in appearance. The mechanical properties of most refractory metals depend strongly on the intrinsic crystalline structure, defects in the crystal structure such as dislocations and vacancies, and the purity of the metal sample. Mechanical properties such as yield strength are strongly temperature dependent, the yield strength at 200°C being approximately 30% of that at 20°C [2]. Some mechanical properties of tantalum are listed in Table 1.4.

Table 1.4 Mechanical Properties of Soft Annealed Tantalum

Mechanical Property	**Soft Annealed Tantalum**	**Cold Formed Tantalum**
Tensile strength	240MPa	400 – 1400MPa
Yield strength	210MPa	300 – 900MPa
Breaking elongation	40%	2 – 20%
Vickers hardness	60 – 120HV	105 – 200HV
Modulus of elasticity	177 – 186MPa	
Ductile – brittle transition temperature	4 K	
Poisson's ratio	0.35	
Recrystallization temperature		1200 - 1500K

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).

Tantalum has a very high melting point of 2996°C, which is exceeded only by tungsten and rhenium [2]. When pure, tantalum is ductile and can be drawn into fine wire, which can be used as a filament for evaporating metals such as aluminum. Even at low concentrations, interstitial impurities increase the hardness or brittleness and decrease the ductility of tantalum. Cold forming is used to improve the hardness, tensile and yield strength to create high performance tantalum for use in applications where fatigue and heat resistance qualities are critical. Some thermodynamic and crystallographic properties of tantalum are summarized in Table 1.5.

Table 1.5 Crystallographic, Thermodynamic and Electronic Properties of Tantalum

Crystal structure	Body centered cubic
Lattice constant	3.3025 Å
Density	16.6 g/cm^3
Melting point	2996°C
Boiling point	5425°C
Heat of fusion	28.5 kJ/mol
Heat of vaporization	78.1 kJ/mol
Specific heat capacity at 20°C	25.41 J mol $^{-1}$K^{-1}
Linear coefficient of thermal expansion at 20°C	6.5 x 10 $^{-6}$ K $^{-1}$
Thermal conductivity at 20°C	54.4 W m $^{-1}$ K^{-1}
Specific electrical conductivity at 20°C	0.081 Ω $^{-1}$ cm $^{-1}$
Superconducting transition temperature	4.3 K
Temperature coefficient (0–100°C)	0.00383

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).

1.6 Chemical Properties of Tantalum

Tantalum is located in group 5B and period 6 of the Periodic Table. Its electronic configuration is $1s^2, 2s^2p^6, 3s^2p^6d^{10}, 4s^2p^6d^{10}f^{14}, 5s^2p^6d^3, 6s^2$. There are two naturally occurring isotopes of tantalum, ^{181}Ta and ^{180}Ta. ^{180}Ta is a weakly radioactive isotope. Tantalum, like niobium, has two principal oxidation states, +2 and +5 [2]. At temperatures below 100°C, tantalum metal is extremely resistant to corrosion by most organic and inorganic acids, except HF, acidic solutions containing the fluoride ion, and free sulfur trioxide [2]. The exceptional resistance of tantalum to corrosion is due to a dense passivating layer of tantalum oxide, which forms on all tantalum metal surfaces upon exposure to air [2]. Tantalum tends to form oxides, hydrides, carbides and nitrides above approximately 300°C. These tantalum compounds are discussed in following sections.

1.6.1 Tantalum Oxides

Tantalum (II) oxide (TaO) is the only stable oxide of tantalum with an oxidation state lower than that of Ta_2O_5. It is produced from Ta_2O_5 by reduction with carbon at 1900°C or with hydrogen at 1100°C. The possibility of introducing interstitial tantalum atoms into the Ta_2O_5 structure leads to the production of TaO_x compounds (x=2 - 2.5). Ta_2O_5 occurs in two thermodynamically stable phases, namely the α and β phases [17]. The transition temperature from the orthorhombic β phase to the tetragonal α modification is approximately 1400°C. A ε phase of Ta_2O_5, which is produced hydrothermally from tantalic acid at 300 – 340°C, also exists. ε-Ta_2O_5 is isomorphous to β-Nb_2O_5 and is transformed by heating in air above 886°C to β-Ta_2O_5 [18]. There are two major processes [2] used to manufacture Ta_2O_5.

In one process, ores and tin slag concentrates are dissolved in HF at approximately 100°C. From the resulting acidic aqueous solution, tantalum and niobium are extracted with methyl isobutyl ketone. Impurities such as AlF_6^{3-}. FeF_6^{3-}, SiF_6^{2-} and SbF_6^{2-}, are removed with H_2SO_4, subsequent to which the tantalum and niobium are extracted in a multistage process. Ammonia (NH_3) gas or solution is added to the aqueous H_2TaF_7 solution to form hydrated tantalum (V) oxide ($Ta_2O_5 \cdot nH_2O$). The hydrated Ta_2O_5 is washed, dried and heated above 800°C to form dry Ta_2O_5 [2] according to the wet chemical method. In the other process used to manufacture Ta_2O_5, the ferroalloy produced by the pyrometallurgical method is chlorinated in an iron chloride sodium chloride melt at 550 – 600°C. The iron in the complex salt $NaFeCl_4$ loses a chlorine atom and is reduced to the divalent state, while the added chlorine gas causes the continuous reoxidation of the iron to Fe(III). The resulting $TaCl_5/NbCl_5$ mixture is distilled to obtain

very pure tantalum and niobium chlorides. The chlorides are then ground up, hydrolyzed with steam in a fluidized bed, and heated to obtain Ta_2O_5.

1.6.2 Tantalates

Tantalates are produced by heating of oxide mixtures and by the dissolution of Ta_2O_5 in molten alkali metal carbonates or hydroxides. The formation of water-soluble tantalate isopolyanions occurs in an excess of the alkali carbonates or hydroxides. These isopolyanions have the form $(H_x Ta_6O_{19})^{(8-x)-}$, where x can be 0, 1, or 2 [2].

1.6.3 Tantalum Halides

Tantalum pentafluoride (TaF_5) can be produced from tantalum chloride by adding anhydrous hydrogen fluoride [2] or by passing fluorine gas (F_2) over metallic tantalum. Tantalum also forms oxyfluorides such as $TaOF_3$ and TaO_2F, which have no economic importance. Tantalum pentachloride ($TaCl_5$) is formed by the chlorination of metallic tantalum, ferrotantalum or tantalum metal scrap, in a process used to produce oxides of very high purity [2].

1.6.4 Tantalum Carbides

Tantalum carbide (TaC) is a high-melting point (3985°C) metal. It is produced by the carburization of tantalum metal, or by reaction of Ta_2O_5 with carbon at approximately 1900°C [2]. Tantalum carbide is present in concentrations below 10% in hard metal, tungsten carbide-based mixtures [19].

1.6.5 Tantalum Nitrides

Tantalum nitride (TaN) is a tough durable material. It is produced by heating tantalum in a pure nitrogen atmosphere at approximately 1100°C [2]. Small amounts of tantalum nitride are often found in cermets [20].

1.6.6 Tantalum Silicides

The tantalum silicides $TaSi_2$ and Ta_5Si_3 are produced by heating tantalum and silicon under vacuum at 1000°C – 1500°C [2].

1.6.7 Analytical Chemistry of Tantalum

The most prevalent method of analyzing for tantalum as a component in raw materials is by extracting it as a fluoro-compound on an ion-exchange resin and then performing gravimetry. Trace concentrations are analyzed by photometric means [2, 21]. In scrap and ores, tantalum is analyzed for by fusion with lithium or sodium borate, followed by X-ray fluorescence analysis, which can determine tantalum concentrations as low as 0.01%. At lower concentrations, tantalum can be determined by emission spectroscopy at 296.5 nm [22].

Sensitive multi-element methods, mainly based on spectroscopy, have been developed for the determination of trace impurities in commercial products such as K_2TaF_7, tantalum oxide, and metallic tantalum. Such methods include plasma optical emission spectrometry subsequent to HF dissolution under pressure [2], flame atomic absorption spectroscopy for the analysis of alkalis, hydride techniques for the determination of arsenic and other hydride forming elements, and photometry for the analysis of phosphorus, which is used as a dopant in tantalum powders. Combustion

analysis is used to analyze for carbon and sulfur, while hot carrier gas extraction is used for nitrogen and oxygen. Glow discharge mass spectrometry is now among the most sensitive methods for trace analysis in the sub-ppm region [23 - 25].

1.7 Uses of Tantalum

1.7.1 Corrosion Protection

Semi-finished tantalum is used in the chemical industry because of its excellent resistance to corrosion by hot concentrated mineral acids, except HF, which can corrode tantalum. The strong corrosion resistance of tantalum is due to the passivation of tantalum surfaces by a dense oxide film 1 - 4 nm thick, which forms spontaneously in air and in acidic environments, and which is stable up to about 260°C [2]. Tantalum, in the pure form and in the form of the alloy Ta-2.5W, is used to construct complex heat exchangers and intricate heating equipment for use in the petrochemical and nuclear industries.

A particularly important use of tantalum in the nuclear industry is for the construction of reaction vessels for the chemical treatment of nuclear fuel [8]. The alloy Ta-7.5W, which is produced by powder metallurgy, is harder than Ta-2.5W and is a preferred material for valves in chemical equipment. Large vessels are often merely applied with a thin coating of tantalum because of the high price of the metal. The coating of steel with tantalum is usually desirable in order to achieve corrosion protection. Tantalum coating of surfaces can be carried out by the explosion method [26], by sputter deposition [27], or by build up welding of tantalum [28].

1.7.2 Capacitors

Semi-finished tantalum is used in the manufacture of electrolytic capacitors [2]. Ever since the discovery that tantalum was a suitable material for electrolytic capacitors [29] this application of tantalum has grown to the point where it now accounts for the largest consumption of the metal [2, 11]. The capacitor consists of an anode of tantalum metal powder compressed into a 20 – 200mg pellet and sintered to achieve electrical continuity and a density of approximately 6 - 9g/cm^3 [11]. Tantalum wire, attached by welding or sintering, is used to provide the electrical connection. Using anodic oxidation and an electrolyte of dilute H_3PO_4/H_2SO_4, often including organic compounds [30], a coating of an amorphous dielectric layer of Ta_2O_5 is applied to the entire pore surface of the sintered anode. Ta_2O_5 has a resistivity of approximately $10^{15}\mu\Omega$cm at room temperature and a dielectric constant ε of 25 [2, 11]. The anode of the capacitor is coated with a semiconducting layer of MnO_2 by impregnation with an aqueous solution of manganese, followed by pyrolysis. This solid electrolyte provides the cathodic contact. A typical design is illustrated in Figure 1.3.

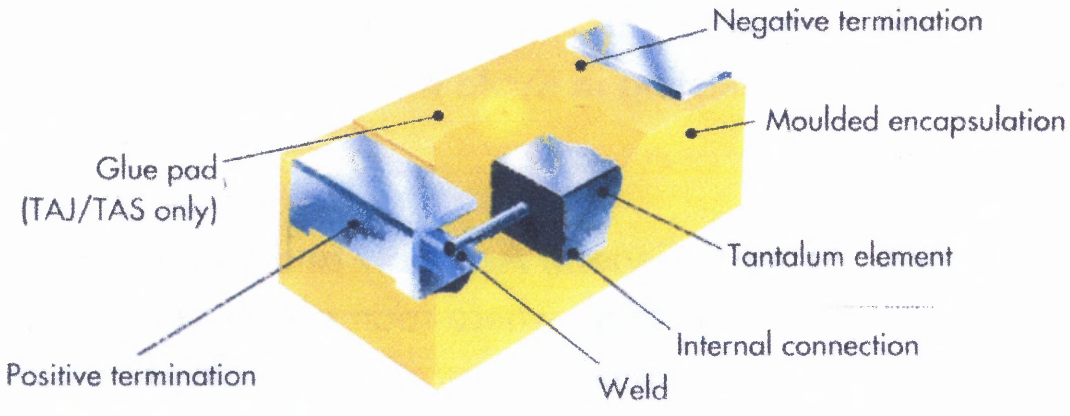


Figure 1.3 Cut-away schematic of a chip type tantalum capacitor design (taken from Reference [11]).

The effect of the pyrolysis temperature on the Ta_2O_5 layer and its dielectric properties, the effect of temperature on the capacity and conductivity, and the effect of all these on the leakage current of the capacitor have been extensively researched [2, 30 - 32]. Gradual improvements in powder quality [15] have resulted in the doubling of the specific capacitance (the product of capacitance and formation voltage per unit mass) associated with tantalum metal powder. Representative properties of capacitor grade tantalum powder are shown in Table 1.6.

Table 1.6 Typical Properties of Capacitor Grade Tantalum Powder

Chemical Impurities (ppm)	**Powder from Electron Beam Melting**		**Powder from Sodium Reduction**	
	QR 7*	QR 3*	VF1-18 KT**	STA-30 KD***
O	800	1500	1800	2200
C	10	10	35	40
N	20	40	50	70
H	15	15	50	50
Nb	< 20	< 20	< 10	< 10
W	< 20	< 20	< 10	< 10
Fe	< 10	< 10	25	25
Cr	< 5	< 5	10	10
Ni	< 5	< 5	20	20
Na	< 1	< 1	< 2	< 2
K	< 1	< 1	5	10
Specific charge, CV/g (μFV/g)	3000	5000	15000	30000
Application working voltage, V	50–75	35–50	25–35	< 25

Source: K. Andersson, K. Reichert, R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).
* H. C. Starck, Newton, USA.
** H. C. Starck, V-tech, Tokyo.
*** H. C. Starck, Goslar, Germany.

Tantalum capacitors are highly valued because of their high reliability, low temperature coefficient of capacitance, and high volumetric efficiency [15]. They add functions to circuits that are beyond the capability of dielectrics manufactured from other materials [11]. Tantalum capacitors have an oxide film that is stable between –55°C and 125°C [2] and are used in laptop computers, automotive applications, communications systems,

aircraft instrumentation and control systems, missile technology and other weapons and maritime systems applications. A recent five-year trend in the shipment and manufacturing capacity of tantalum capacitors is shown in Figure 1.4.

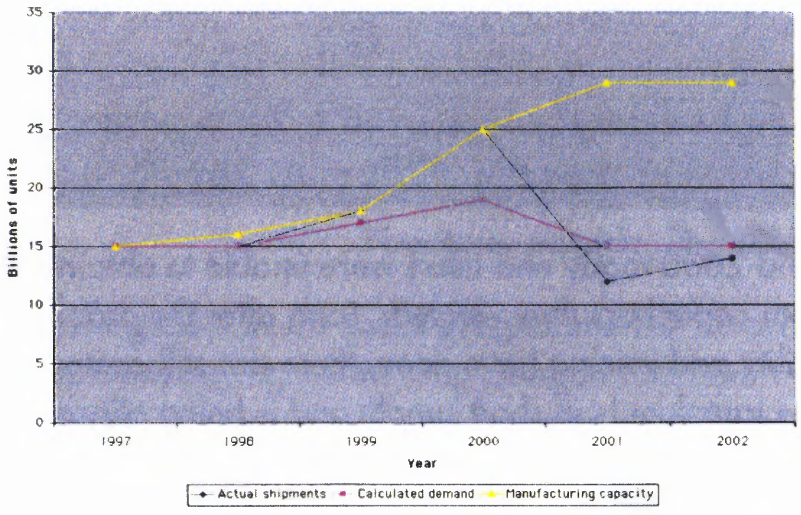


Figure 1.4 Trends in shipment and manufacturing capacity of tantalum capacitors (taken from Reference [1]).

The last decade has seen an increase in demand for tantalum capacitors in high technology applications such as cellular telephones and pagers, personal computers, disk drives and subassemblies in automotive electronics. These applications now represent major end uses for tantalum capacitors [5]. In 1983, the capacitance of a standard 3216 tantalum chip capacitor could reach 1μF. The same 3216 case size capacitor could achieve a capacitance of 47μF in 1999 [15].

1.7.3 Medical Uses of Tantalum

Metallic tantalum is highly inert and nontoxic, in addition allergic reactions to the metal have not been reported. It has very good compatibility with human tissue. These qualities make it a very attractive metal for use in the manufacture of surgical instruments and for

X-ray diagnostics. It is used clinically in sutures and plates, and when alloyed with titanium, in orthopedic medicine and dentistry [33]. Inhaled tantalum powder has been investigated for its suitability as a contrast medium in X-ray diagnostics of the throat, trachea, larynx, bronchi, esophagus and stomach [2], but the material has not seen widespread use in these applications, mostly for reasons of cost. Small tantalum spheres or wires are used in organs such as the heart, in bone, or in implants when it is desired to control the function and position of such implants by X-rays [2].

The introduction of a foreign body such as a metal into living tissue will subject that metal to corrosive processes, whose end result on the surrounding tissue will be denaturation. The corrosive processes consist of redox reactions, which entail electron exchange at the metal-tissue interface, hydrolysis, which entails proton exchange, and the creation of metal-organic complexes around the tissue in close proximity to the metallic implant. As a consequence of the corrosive processes, the metallic implant can become surrounded by a pseudo-membrane [34] in a process known as sequestration. With implants made of metals such as Fe and Ag, the corrosion is so rapid that the tissue in contact with the implant becomes oxygen-starved. Metallic ions, some of which are toxic to cells, are also released during the *in-vivo* corrosion process. These ions do not usually follow the same metabolic pathway as metallic compounds ingested from dietary sources. The redox reaction needs the passive metallic oxide layer formed on metal implants to be electrically conducting, so that electrons are available at the outer interface. The passivating layers formed on iron, chromium and stainless steel are conductive [34], however the passivating layer of tantalum (V) oxide formed on tantalum surfaces is a good insulator, which inhibits the redox reactions associated with corrosion.

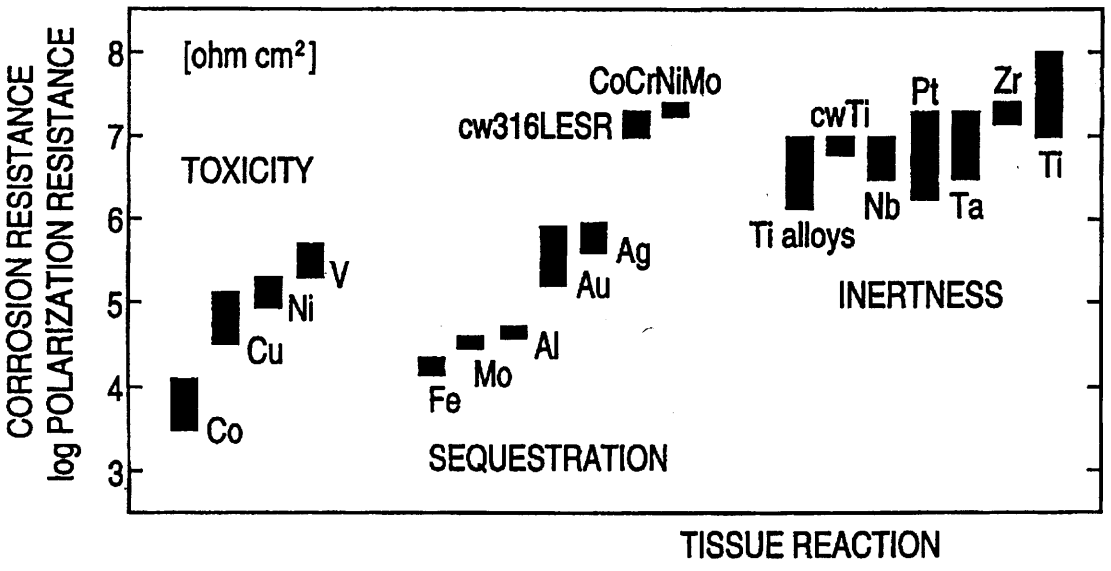


Figure 1.5 Data collected from in vivo corrosion experiments showing three different categories of tissue reaction to implants made from various metallic elements and alloys (taken from Reference [34]).

Tantalum metal is virtually bioinert in the human body because of its stable oxide layer [2]. Coating AISI 316L stainless steel medical implants, composed of austenitic chromium, nickel and molybdenum steel with a sandwich layer of tantalum and tantalum oxide improves the pitting corrosion resistance of such implants, and avoids the deleterious effects of such corrosion in human tissue, such as formation of connective tissue between the implant surfaces and body tissue, change of pH, release of metallic ions from the implant and allergic reactions to the nickel and chromium components of stainless steel [35]. α and β type titanium alloys such as Ti-6Al-4V ELI [36] are widely used for implants because they possess excellent corrosion resistance, biocompatibility and mechanical properties well suited for hip joints and dental applications. However, these alloys contain toxic elements such as aluminum and vanadium. Aluminum is a possible cause of Alzheimer's disease [36] and is known to inhibit bone growth, while

vanadium is known to be highly cytotoxic [36]. Much less toxic β type titanium alloys containing tantalum instead of the toxic metals aluminum and vanadium are being developed which retain all the desirable properties of the old type β titanium alloys [37]. The figure below shows that except for the area near the incision made to insert the implant, little inflammatory response occurs in subcutaneous rat tissue after insertion of a Ta implant for 4 weeks. Also new bone formation is observed around a Ta implant after 4 weeks following insertion into rat femoral bone marrow. The area of direct contact at the bone-metal interface is observed to increase [37] after a 2-week period.

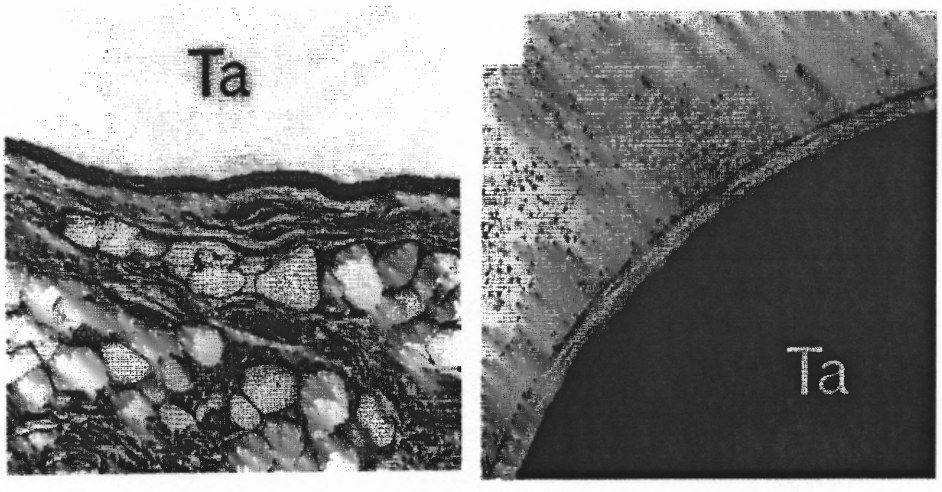


Figure 1.6 Tissue response to Ta implant in subcutaneous rat tissue after a period of 4 weeks (left panel), and new bone formation around a Ta implant (right panel) in rat femur bone marrow after the same period of time.

Tantalum shows excellent bone-bonding ability due to its ready formation of biologically active bonelike apatite material upon treatment [33] and has been considered for application as an artificial implantation material in humans where qualities such as the ability to bond directly to bone and high fracture toughness are required. Animal experiments have shown that broken bones held in position with tantalum rods heal better

than bones held together with chromium-nickel steel or niobium, even when the tantalum rods are used in septic conditions with corrosive tissue reactions [2]. Jaw implants made of tantalum have exhibited good metal-bone contact over periods of many years with no adverse reactions [32]. Tantalum mesh and plate is also implanted in damaged areas of the skull in some cases [2]. In addition, tantalum is used in miniature clips for the rapid temporary or permanent occlusion of vessels during surgery [2].

Tantalum is being investigated as an important component in methods for depositing onto medical instruments highly biocompatible and bioactive surface coatings that can promote and stabilize cell attachment [38]. The method and the materials could have important ramifications in a number of areas of research and biotechnology, for example for chronic implantation of microelectrode arrays in the cerebral cortex for neuroprosthetic and neural monitoring applications, as well as for research on the human central nervous system. Other possible applications of these new materials are for coronary artery stents and pacemaker electrodes.

1.7.4 Tantalum-based Alloys and Superalloys

Ever since the end of World War II, requirements have existed for materials able to withstand extremely harsh conditions such as encountered in jet engines, propulsive and nuclear systems, and space launch rockets and vehicles. Turbine buckets must for example withstand, often for hundreds of hours at a time, tensile stress conditions, creep, fatigue, oxidation, and sporadically, impact, thermal shock and thermal fatigue [39]. Important alloys containing tantalum include Ta-8% W-2% Hf (T-111) and Ta-9.6% W-2.4% Hf-0.01% C which were developed for use in aircraft jet nozzles and for cladding the interior of combustion chambers and the exterior of rockets and spacecraft [40 - 41].

These alloys allow aircraft and land-based turbine engines to operate at higher temperatures and therefore higher efficiencies. The so-called superalloys are another important class of alloys containing tantalum. Superalloys [42 - 43] were formulated to satisfy the ever-increasing requirements for materials that facilitate higher operating efficiencies of gas-turbine engines and are metallic multicomponent materials having nickel, cobalt, or iron as the main constituent, with metals such as tantalum, chromium, aluminum, hafnium, molybdenum, titanium, tungsten etc being added to enable use of the superalloys at conditions of high temperatures and high mechanical stress.

Superalloy components such as TRW-NASA VI A, MAR-M 302 contain up to 9% tantalum [2] and are produced by forging, casting and powder metallurgy. Such superalloys are used in aircraft engines as well as stationary gas and steam turbines. In superalloys tantalum contributes primarily to properties such as high-temperature strength, resistance to high-temperature fatigue, and resistance to corrosion by sea air.

1.7.5 Other Uses of Tantalum

The tantalum oxide Ta_2O_5 in its hydrated form is used as an acid catalyst [2]. It exhibits very good catalytic activity for alkylation, esterification, Beckmann rearrangement and aldol condensation reactions [44 - 48]. Ta_2O_5 is used in lanthanum borate glasses [2], characterized by their low optical scattering properties and high refractive index, although niobium pentoxide (Nb_2O_5) is replacing Ta_2O_5 for this application because it is 3 - 5 times less expensive. TaO is a major component of fiber-optic materials used to facilitate dense wavelength divisional multiplexing and has been investigated for possible use in the production of heat reflecting glass for windows [2].

Single-crystal wafers of lithium tantalate ($LiTaO_3$) are produced by the Czochralski crystal growth method [49 - 50] and are used as surface acoustic wave (SAW) filters for high-frequency applications in communications systems and as intermediate-frequency filters in television and video equipment [2]. $LiTaO_3$ is also used together with lithium niobate in the fabrication of waveguides, frequency multipliers, optical modulators and optical switches [2]. Yttrium tantalite powder ($YTaO_4$) is applied as an emulsion to a film, and is used in medical diagnostics as a phosphorescent substance for X-ray amplification [2]. $YTaO_4$ enables the reduction of X-ray radiation doses received by patients because it converts X-rays to fluorescent light in very high yield [51 - 52].

Barium magnesium tantalate ($Ba_3MgTa_2O_9$) and barium zinc tantalate ($Ba_3ZnTa_2O_9$) have the perovskite structure and are highly promising materials for frequency filters as well as base material to be used in microwave resonators for the stabilization of oscillators because of their high dielectric constants and very minimal high-frequency dielectric loss [53 - 56]. The high-quality resonators produced from these materials will be invaluable in the area of satellite communications where high carrier frequencies are the norm.

TaF_5 as well as niobium pentafluoride (NbF_5) are used in the petrochemical industry as catalysts for isomerization and alkylation processes. The fluorides of tantalum and niobium are also important as fluorination catalysts in the manufacture of fluorochloro-hydrocarbons and fluorinated hydrocarbons [57]. K_2TaF_7 is an industrially important intermediate compound in tantalum metal production. $TaCl_5$ and tantalum alkoxides are used in the chemical vapor deposition (CVD) of tantalum metal or oxide.

TaC is used as powder in the manufacture of hard metals, such as tungsten carbide based metal mixtures [2]. It improves the thermal shock resistance and the high-temperature fatigue strength of cutting tools. TaN films are used as diffusion barriers in integrated circuits in semiconductor technology [58]. They are also used as piezoresistive materials for strain gauges [59]. $TaSi_2$ targets were formerly used as sputtering targets in the semiconductor industry for the production of circuit interconnections [60].

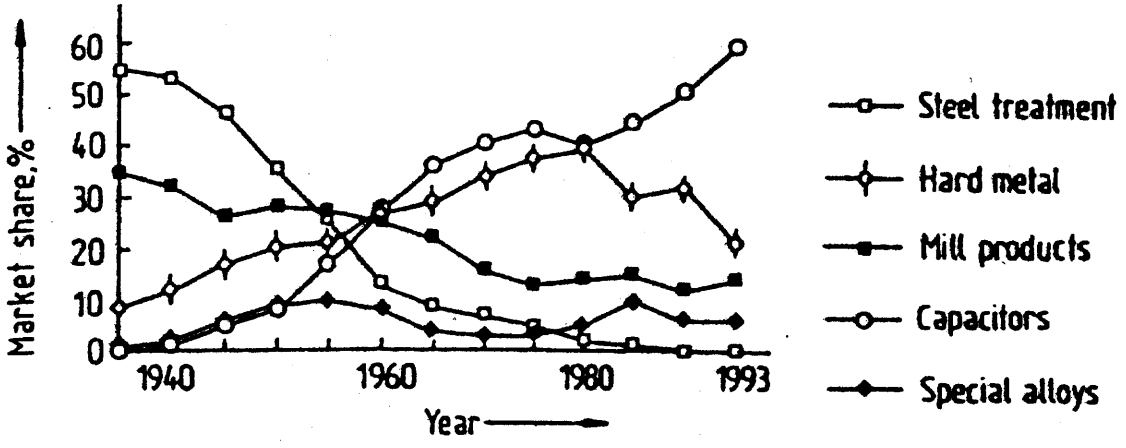


Figure 1.7 Market share of tantalum applications from 1935-1993 (taken from Reference [2]).

The importance of tantalum to a wide spectrum of industries has been discussed in this chapter. The microelectronics industry, where tantalum is largely used in thin film form, is responsible for the single largest consumption of the metal. In order to fully appreciate the importance of tantalum in thin film technology, the evolution of the technology and the physics phenomena on which it is based is covered in Chapter 2. The more specific issue of metastable phase formation in tantalum thin films with which this dissertation is concerned will arise naturally in this context.

CHAPTER 2

PHYSICS AND APPLICATIONS OF THIN FILM TECHNOLOGY

2.1 Thin Film Technology

A thin film is a liquid or solid having one of its linear dimensions much smaller than the other two dimensions. Thin films can be supported on a solid substrate or they can be unsupported and they can be produced by a variety of chemical and physical deposition processes. They find wide uses in optical, electronic, magnetic, mechanical and protective applications, for example they have long been recognized as useful materials for the fabrication of passive circuit elements in Resistive Capacitive (RC) networks [61]. Thin film technology grew rapidly between the two world wars when mirror optics and antireflection coatings were produced respectively from vacuum evaporated metal and dielectric films. In the 1960s, well after the end of World War II, thin film methods enabled the mass fabrication of semiconductor electronic components by printing techniques, and the miniaturization of circuit features by integration [61]. The increase in integrated circuit (IC) transistor integration densities over the decades is shown in Figure 2.1.

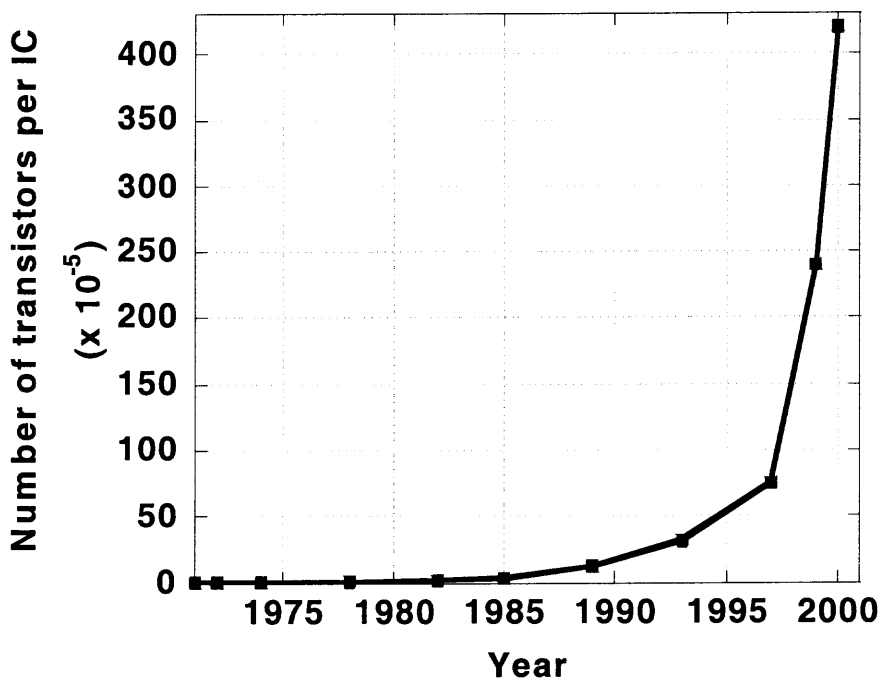


Figure 2.1 Increase in IC transistor integration densities over three decades.

The optical applications of metallic and dielectric thin films include filters, optical wave-guides for opto-electronic communication, reflection coatings, and infrared sensors. Electrical and magnetic applications of thin films include magnetic and superconducting films for memory and logical devices. Thin film properties differ significantly from bulk metal properties. One of the main goals of scientists working with thin films is to gain insight into the processes leading to their special structural, electronic, optical, and magnetic properties. Advances in the application of quantum mechanical based computational methods over the past two decades, which must often be used in tandem with experimental investigative methods, have enabled great advances in the understanding of the subtle phenomena responsible for the rich variety of electronic and magnetic properties observed in thin films.

2.2 Structural Properties of Thin Films

The various methods of preparing thin metal films, molecular beam evaporation (MBE), evaporation and sputter deposition, can be used to produce thin films representing new material phases, as well as thin films having the same crystalline structure as bulk material precursors. The crystalline size along the major dimensions of a thin film can range from a few Angstroms in amorphous structures to a few centimeters in high quality single crystal films. Proper structural characterization of thin films must include information about the crystal size and orientation, the surface roughness and the internal state of stress in the film. High-resolution techniques are usually required to determine these properties because of the often-small amounts of thin film material present.

Methods that have been used to study thin film structure include optical microscopy, electron microscopy [62] and diffraction [63], scanning electron microscopy, x-ray diffraction and electron diffraction techniques differentiated by the energy of the incident electrons used for scattering, such as low energy electron diffraction (LEED), quantitative low energy electron diffraction (QLEED), and reflection high-energy electron diffraction (RHEED). Transmission electron diffraction can be used to determine grain orientation while RHEED can be used to detect changes in orientation. Transmission electron microscopy can be used to determine the distribution of imperfections in small areas of a single crystal film. This method is also very useful for monitoring the various stages of film nucleation and growth, and for characterizing bulk material surface topology.

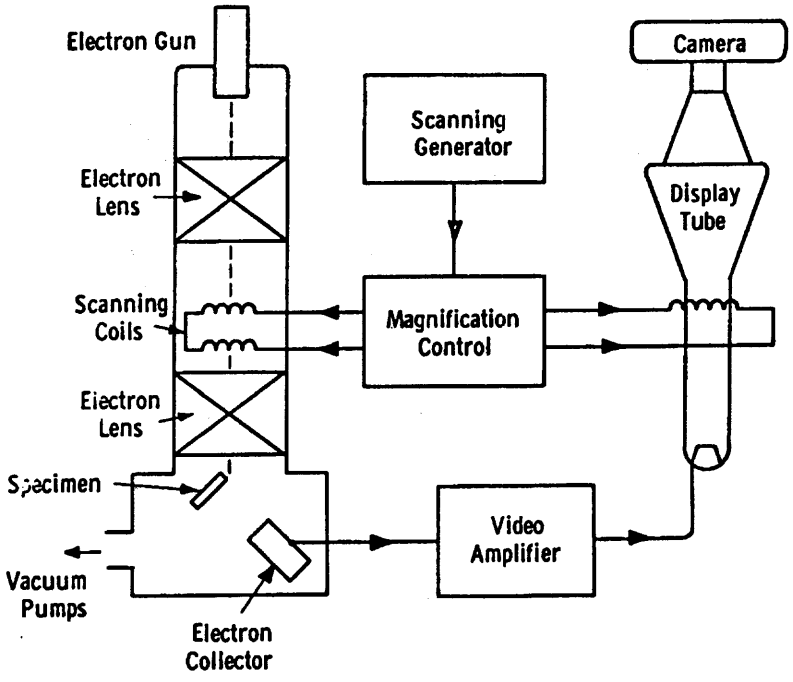


Figure 2.2 Schematic diagram of a scanning electron microscope (SEM) system (taken from Reference [66]).

The scanning electron microscope can be used to reveal film surface topology and electrical anomalies in semiconducting thin films. Recently, high-resolution instruments such as the Atomic Force Microscope (AFM) [64] and the Scanning Tunneling Microscope (STM) [65] have been used to observe and manipulate thin film crystalline structures at very high levels of resolution and detail. X-ray diffraction is a non-destructive characterization technique, which provides a much larger penetration depth than any sort of electron diffraction. A very clear picture of film orientation can be obtained from X-ray diffraction by rotating the sample with respect to the x-ray beam while monitoring the changes in diffracted intensity on a synchronously moving detector. Another technique based on X-ray radiation and known as X-ray topography can be used to determine the distribution of crystalline imperfections over large thin film volumes. Surface analysis of thin films requires a technique such as LEED, which is more surface-

sensitive than X-ray diffraction. The low energy (10 - 500eV) of the electrons used in LEED enables penetration only to a few monolayers, making LEED ideal for surface analysis. LEED has been extensively used to study the initial stages of epitaxial film growth.

Thin films with an amorphous structure can be formed when atoms are deposited onto a cold substrate and do not have sufficient mobility to move into energetically favorable long-range order positions. Amorphous thin films are useful for semiconductor applications and insulators in dielectrics. Amorphous structures are useful for capacitors and passivating layers because electrical leakage is minimized from such structures due to their lack of grain boundaries. There is usually a transition temperature associated with amorphous structures, which determines their transition to crystalline structures. For example by slowly increasing the film temperature, amorphous Ta_2O_5 films can be transformed into films with the crystalline structure of β-Ta_2O_5 [66]. Deposition of materials such as Ta_2O_5 at temperatures higher than the transition temperature results in the production of polycrystalline films. Fiber-textured films, in which the crystals grow with a common crystal axis parallel to a given direction, result if deposition occurs at temperatures higher than the transition temperature on a non-crystalline substrate such as glass. The direction of the common axis varies with deposition temperature and film structure. GaAs films produced by sputtering show an abrupt transition from a strongly [110] film texture to a strongly [111] film texture when the substrate temperature is increased from about 520 to 550°C [66]. The understanding of the origins of fiber texture changes in thin films is still not comprehensive.

2.3 Electrical Properties of Thin Films

Thin films have played a major role in the development of major technological innovations such as the semiconductor IC (see Reference [61] for a full review). They have been used extensively to make passive circuit elements consisting of resistors and capacitors. Several factors combine to make thin film electrical properties different from bulk metal electrical properties. One such factor, which is to an extent dependent on preparation conditions, is that thin films do not grow as continuous layers in the first stages of growth. In the course of a typical thin film growth process, islands of atomic aggregates with an average thickness of a few thousand Angstroms can be formed before island coalescence and the formation of a continuous layer. Another consideration that is important in determining thin film electrical properties is that the possibility of contamination of the films is always high when they are prepared by deposition methods such as sputtering. Yet another consideration is that thin films grown under high rates of deposition tend to have small grain sizes and a significant distribution of structural imperfections, with drastic effects on electrical properties such as the resistivity. The description of thin film electrical devices in the following sections is taken largely from References [61], [66] and [67].

2.3.1 Thin Film Resistors

Thin film resistors are manufactured by deposition of film material between two suitable electrodes placed on a much less conductive substrate. Resistors formed from a thin film material in this manner are usually much more resistive than the corresponding bulk metal. The sheet resistance of a thin film is an important parameter and is defined as

$$R_s = \rho / A, \tag{2.3.1.1}$$

where ρ and A are respectively, the resistivity and the surface area of the thin film. Desirable properties of thin film resistors include high resistivity and low temperature coefficient of resistance (TCR). The most stable resistors are generally formed from metallic oxides and nitrides. The resistivity and TCR of thin films has always been observed to change upon heating, the resistivity decreasing and the TCR increasing. Thin film resistivity has also been observed to decrease with increasing film thickness. The high resistivity of thin films relative to bulk materials is usually explained by modification of the electronic distribution function due to electronic scattering from imperfections, and the lower mean free paths available to electrons in thin films compared to bulk materials. The evolution of the electronic distribution function and hence the electrical resistivity under the influence of an external field is described by the Boltzmann equation.

An early investigation [61] of the suitability of different classes of metals for the manufacture of thin film resistors showed that refractory metals such as tantalum, molybdenum and tungsten were best suited for this purpose, because they had the highest values of resistance per unit area and the lowest values of TCR before and after annealing. Thin film resistors made from refractory metals can also be anodized to produce a dielectric layer having high permittivity and insulation for conductor interconnects and capacitors in thin film ICs. Tantalum resistors are usually deposited with resistance values lower than required and then anodized to increase the resistance to the required level. They are stable within 1% of their resistance value when used at

temperatures below 100°C [61]. Extremely stable resistors have also been produced by sputtering tantalum in O_2-Ar and N_2-Ar mixtures. Films sputtered in nitrogen usually have resistivity values 10 – 40 times the value of that of the bulk metal, while films sputtered in argon have resistivity values on the order of twice the bulk values. Migration of oxygen into thin film grain boundaries has been identified as the major cause of resistance destabilization in thin films. One remedy for the thin film grain boundary oxygen diffusion problem consists of diffusing a non-oxidizable metal such as gold into the grain boundaries to prevent oxidation [66].

2.3.2 Thin Film Capacitors

Thin film capacitors are indispensable components in the ever more miniaturized circuits that have facilitated the information and communications revolution of the past decade. Important properties of thin film insulator films used in capacitor dielectrics include the dielectric constant ε, the dielectric loss as represented by the dissipation factor tan δ and the breakdown voltage V_D. The dielectric constant of excellent dielectrics such as amorphous Al_2O_3 and Ta_2O_5 thin films remain essentially unaltered down to 10 – 50nm [67]. Thin film effects on dielectric loss can be attributed to the different structure of thin films as compared to corresponding bulk materials. Dielectric loss in thin film capacitors is frequency dependent, the complexity of this dependence depending on the particular thin film material. Good dielectrics such as amorphous Al_2O_3 and Ta_2O_5 thin films have virtually constant losses in the audio frequency range. The value of tan δ for Ta_2O_5 thin films is about 4 x 10^{-3} in this frequency range [67]. Thin film dielectrics usually have appreciably higher values of the breakdown voltage than bulk material dielectrics. This has been attributed to the inhibition of current avalanche due to the comparative absence

of leakage pathways in thin films. The usual mechanism for voltage breakdown in thin films is thermal runaway [67], and this occurs when at voltages near breakdown, small areas in the film draw a disproportionately large amount of current leading to local heating and the enhancement of current due to a negative TCR. Thin film capacitors having low loss will have a dielectric tan δ value of approximately 10^{-3}, a temperature coefficient of capacitance (TCC) of < 20ppm K^{-1}, and a very high capacitance per unit area, an example of a dielectric thin film material suitable for fabricating capacitors with these properties is amorphous Al_2O_3. Thin film capacitors having high capacitance, *i.e.* charge storing ability will have a dielectric tan δ value of approximately 10^{-2}, a temperature coefficient of capacitance (TCC) < 500ppm K^{-1}, and a very high capacitance per unit area, an example of a dielectric thin film material suitable for this kind of capacitor is amorphous Ta_2O_5.

Thin film capacitors used in microelectronic applications are usually deposited in three layers, with an intermediate layer of dielectric between top and bottom metallic electrodes. Materials that have been used to produce thin film capacitors include alkali halides, metal oxides and silicon oxides. Substrate imperfections are the main cause of faults in dielectrics made from silicon oxides, making a clean substrate and dust free production environment essential. Sputtered metal oxides of metals such as tantalum, titanium and niobium can have high permittivity values. Reactive sputtering highly favors the formation of highly oxidized thin films because of the higher pressures involved compared to techniques such as evaporation.

The need for Dynamic Random Access Memory (DRAM) capacities greater than 64 megabytes was recognized in the mid-1990s [11]. In the course of the search for

alternate dielectric materials, Ta_2O_5 thin films were discovered to possess a much higher dielectric constant (20 – 27) than SiO_2 films (dielectric constant 3.9) [68], which were the dominant dielectric material at the time. Ta_2O_5 dielectrics could therefore enable a 5 to 6 times increase in the capacitance density compared with SiO_2 dielectrics. Higher packing densities of DRAM charge storage cells could then be achieved without having to produce impractically thin dielectric films. Ta_2O_5 thin films have subsequently been produced by anodic oxidation [69] and thermal oxidation [70] of tantalum films, reactive sputtering [71 - 72], plasma deposition [73], and chemical vapor deposition [74 - 76]. Ta_2O_5 thin films have also been produced from metal halides and organometallic reactants using CVD, for example from the reaction of tantalum pentaethoxide ($Ta(OC_2H_5)_5$) and oxygen using a low pressure CVD (LPCVD) reactor [77].

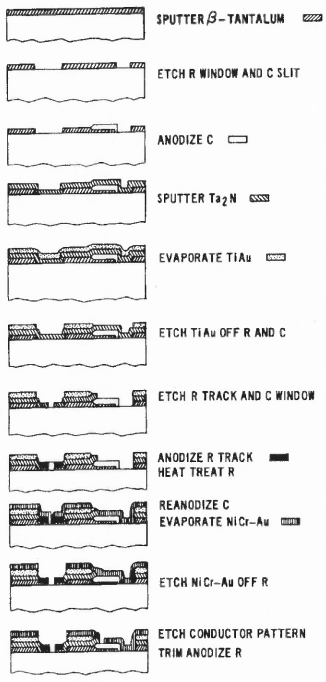


Figure 2.3 Process sequence for the fabrication of resistors and capacitors from different tantalum thin films on the same substrate (taken from Reference [83]).

2.3.3 Thin Film Superconductors

The phenomenon of superconductivity occurs in a large number of different metallic and alloy systems. Superconductivity results in a breakdown of the electrical resistance of a material below the superconducting transition temperature, T_c. Magnetic flux is emitted from the interior of superconductors to a narrow surface layer approximately 400Å thick, and decreases exponentially with distance into the surface of the superconductor. Above a critical magnetic field strength, superconductivity breaks down in type I superconductors, while type II superconductors develop vortices exhibiting normal state and superconductive regions. At the superconducting transition temperature, conduction electrons begin to condense into pairs known as Cooper pairs [67]. Electrons in a Cooper pair have opposite spins and magnetic moments and each electron in the pair is lower in energy than a single electron. The Cooper pair electrons are coupled by virtual phonons due to lattice vibration and the coupling mechanism ranges over a distance of about 10^8Å. All electrons are assumed to be condensed to Cooper pairs at $T = 0$. The binding energy of Cooper pairs appears as an energy gap near the Fermi level, E_F. Above $T = 0$, the pair bonds can be broken with energy release of the order of mega electron volts (MeV).

Differences in T_c of a few degrees between thin film and bulk superconductors are common. A variety of factors are responsible for this and other property differences between thin film superconductors and bulk material superconductors. Some thin film materials such as indium and aluminum exhibit an increase in T_c with temperature while other materials such as strontium exhibit the opposite trend. These effects can be due to internal film stresses accumulated during the film growth process. T_c has also been found to depend on the intrinsic structure of materials, for example Al has a bulk T_c of 1.2K

and a thin film T_c of 4.2K, while Bi and Be do not exhibit superconductivity in their bulk states but do so when they are deposited as amorphous thin films with T_c of 6K and 8K respectively. Low concentrations of impurities generally lead to a linear decrease in T_c, therefore superconductive films such as Ta, Nb and V thin films with high oxygen gettering ability can exhibit values of T_c vastly different from the corresponding bulk values.

The thin film upper critical field H^{cf} depends on the thickness and type of superconductor according to [67]

$$H^{cf} = H^{cb}\sqrt{6\lambda}/z, \tag{2.3.3.1}$$

for $z < \lambda$, and [67]

$$H^{cf} = H^{cb}(1+\lambda/z), \tag{2.3.3.2}$$

for $z > \lambda$, where z is the film thickness and the subscripts f and b represent respectively, the film and bulk. Quantum theory predicts the following dependence for the upper critical field [67]

$$H^{cf} = H^{cb}(z/b), \tag{2.3.3.3}$$

where b is the width of the film. Type II superconductors have a more complex expression for the upper critical field, which must account for the dependence of the Ginzburg-Landau parameter on z. Associated with the critical field is a critical current that generates the critical field. High T_c superconductors are usually more likely to

exhibit high critical field currents in thin film form than in bulk form. For $z << \lambda$, the critical field current I^{cf} is given by [67]

$$I^{cf} = H^c b \tanh z/2\lambda . \tag{2.3.3.4}$$

Memory storage devices that have the capability to switch between normal and superconducting states are one of the possible applications of superconducting thin films. High speed elements with switching times around 50ps can be produced, which can result in computing devices 10 times faster than currently available, with less than 0.1% of current power usage. Another application of superconducting thin films is the thin film cryotron. A cryotron is a device in which a control coil produces a field higher than the gate critical field, effecting a switch from the superconductive state to the normal state. Logical units may be built by amalgamating several cryotrons to create for example a single bit memory device consisting of a flip-flop and "read" and "write" units.

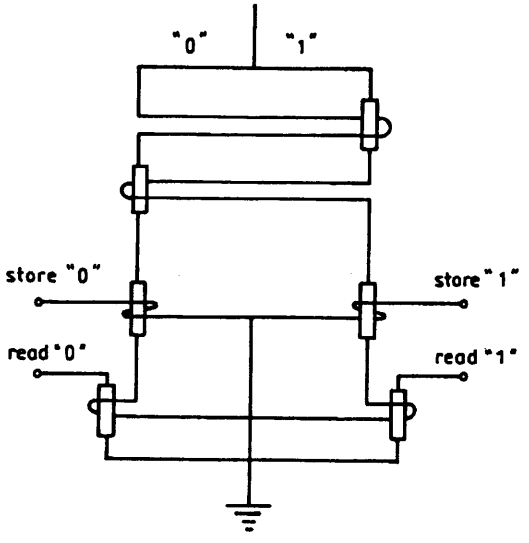


Figure 2.4 Memory device comprising of "read" and "write" modules and a flip-flop (taken from Reference [67]).

Yet another application of thin film superconductors is the class of devices known as Super-Conducting-Quantum-Interference-Devices (SQUIDS), which are used in fluxmeters and galvanometers. The principle here is that introducing a double junction in a superconducting loop increases the sensitivity of the Josephson tunneling current on an external magnetic field. The electronic wave functions undergo phase changes with changes in the magnetic flux, resulting in much higher sensitivity of detection down to 10^{-11}Gauss [67]. The Josephson effect is also useful for generating microwaves with very high spectral resolution at a frequency range up to 1000GHz. The high spectral resolution, which can be up to 10^{-7}, corresponds to infinitesimal temperature full width at half maximum (FWHM) and can therefore be used to measure temperatures down to the milli-Kelvin range.

2.3.4 Thin Film Semiconductors

Thin film technological production steps are an integral part of modern semiconductor device fabrication from start to finish of the device fabrication process. Many semiconducting devices, such as the insulated gate field effect transistor (FET) incorporate thin films as essential parts of their function [61, 67]. At $T=0$, a semiconductor has a filled valence band and an empty conduction band, both bands being separated by a band gap region. The magnitude of the energy separating the top of the valence band from the bottom of the conduction band is ΔE_g. As the temperature increases, electrons can acquire enough energy to overcome the band gap with probability $\exp\left(\Delta E_g / k_B T\right)$ in order to be promoted to the conduction band [67]. Doping a semiconductor material with an impurity having an excess valence electron forms the type of semiconductor known as an n-type semiconductor. The extra electrons from the

impurity dopants are free to move into the conduction band, where they are available to conduct electricity.

In p-type semiconductors, impurities needing a single electron to close their outer electron shell are used to dope semiconductor materials. These electron-poor impurities then take up electrons from the valence band, leading to the proliferation of electron holes. The product of the hole and electron concentrations is always constant by the law of mass action. The equilibrium concentration of electrons is practically equal to the equilibrium concentration of donors. The hole concentration in an n-type semiconductor is always much lower than the hole concentration in an intrinsic semiconductor. Likewise, the electron concentration in a p-type semiconductor is always much lower than the electron concentration in an intrinsic semiconductor. Recombination processes act to reduce excessive concentrations of holes or electrons. The lifetime of the minority charge carriers describes the change from non-equilibrium to equilibrium concentrations. The diffusion length is the average distance traveled during the lifetime of the minority charge carriers as a result of movement due to statistical thermal fluctuations.

Some examples of semiconductor devices are the $p-n$ junction, the bipolar transistor, and the insulated gate field effect transistor (IG-FET) [67]. A $p-n$ junction is depleted of mobile carriers by a process whereby a mobile space charge is created due to the diffusion of charge carriers through the $p-n$ diode junction where recombination occurs. The depletion zone is expanded and a positive and negative external voltage on n-type and p-type semiconductors respectively, facilitates the diffusion voltage and restricts carrier diffusion. Equilibrium is reestablished by switching the polarity of the

external voltage bias, increasing the diffusion current to counterbalance the diffusion voltage.

A bipolar transistor is formed by having an array of $p-n$ junctions in a *pnp* or a *npn* sequence. An array of $p-n$ junctions in a *pnp* sequence forms an emitter diode and is forward biased, while an array of $p-n$ junctions in an *npn* sequence is biased in blocking direction and is a collector diode. If the diffusion length is greater than the thickness of the semiconductor base, the base may be penetrated by minority carriers, which then may also penetrate the collector junction. The success with which amplification of electrical signals is achieved is determined by the degree of bias applied to the emitter-base and base-collector diodes, which controls the concentration gradient, and hence the flux of minority carriers.

The IG-FET is an example of an FET device. FET devices are routinely manufactured from thin films. An example of an FET device is a p-Si-SiO_2-Al structure [67]. In order to operate the FET device, Band bending is used to deplete the hole concentration at the SiO_2-Si interface. By inducing positive bias in the Al electrode, the Si conduction band is bended until it overlaps the Fermi level. Since the Si conduction band overlaps the Fermi level, it will be filled by electrons, leading to an n-conducting channel in close proximity to the surface. An n-conducting region can now be used, instead of a thermal source, to fill and drain the band. The amount of electrical current can be controlled by the insulated metal electrode, which allows for switching and amplification. Inversion in p-type silicon can now be present at zero bias voltage because the Si surface is usually of n-type due to the surface states. Enhancement type

FETs do not conduct at zero voltage, while self-conducting depletion type FETs conduct at zero voltage.

Semiconductor silicon wafers are starting materials for the photolithographic manufacture of IC devices, and metal oxide semiconductor field effect transistor (MOSFET) ICs. Thin films of refractory metals such as tantalum, or other metals such as aluminum can be deposited on an oxide mask formed by a photolithographic process in order to produce metallic interconnections for IC devices. There is also substantial interest in gallium arsenide (GaAs) IC devices because of the higher values of the electron mobility and peak velocity in GaAs compared to Si. The semiconductor on insulator technology (SOI) employs oxides as substrates in large and very large scale integrated (VLSI) circuits. This technology enables dielectric insulation between adjacent devices in ICs. Thin film circuits of this type have several advantages over corresponding circuits on bulk silicon; they produce higher speed signals because of their lack of parasitic capacitances and because of their lower power dissipation enabling higher density of packing, and much higher resistance to deleterious ionizing radiation. Single crystalline films are essential for this technology due to the higher charge carrier mobilities and hence higher circuit speeds obtainable with these sorts of films compared to amorphous or smaller-grained polycrystalline films. SOI technology will also make it possible to integrate optical and electronic devices in three dimensions, improving data processing, display and control functions significantly over the current capabilities afforded by two-dimensional device integration. Diamond films are being investigated for SOI substrates for application in three dimensional high packing density SOI ICs because of their high electrical resistivity and heat conductivity [67].

Another class of semiconductor devices consists of thin film transistors (TFTs), which are composed of a glass substrate upon which a semiconducting film such as CdS or CdSe deposited between two metallic strips is supported. The gate electrode is deposited on a dielectric encasing the semiconducting film. TFTs have been used as integrated switches to produce matrix addressed display panels [67].

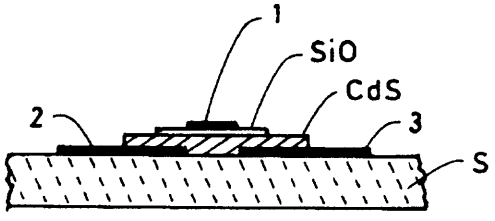


Figure 2.5 A thin film transistor (taken from Reference [67]).

Photovoltaic cells, used to harness solar energy for conversion are another very important application of thin films. If α is the linear radiation absorption coefficient of a semiconductor [67],

$$I = I_0\left(1-e^{-\alpha l}\right) \approx I_0 \alpha l \text{ for } \alpha l << 1, \tag{2.3.4.1}$$

where l is the penetration distance into the semiconductor and I_0 is the incident radiation intensity. The number of electron-hole pairs is increased from the equilibrium level at a rate [67]

$$r_G = I_0 \alpha q / \hbar\omega, \tag{2.3.4.2}$$

where q is the quantum efficiency and the number of photons per unit volume is equal to $I_0\alpha/\hbar\omega$. The created charge carriers increase the current density of a $p-n$ junction by the quantity [67]

$$j_L = \frac{I_0 \alpha q e}{\hbar\omega}(L_e + L_h), \tag{2.3.4.3}$$

which is proportional to the incident radiation intensity.

Thin film hetero- and homo-junctions are very useful for power conversion applications. Several thin film hetero-junctions include CdS-Cu_2S film junctions consisting of type p Cu_2S, mixed type or type i Cu- and CdS and type n CdS, InP-CdS, $InCuSe_2$-CdS, GaAs-ZnSe and CdTe-CdS systems. A common homo-junction is the silicon homo-junction prepared by epitaxial CVD growth. Amorphous films of silicon are now being used for large-scale photovoltaic cell applications. Conversion efficiencies in these cells have been improved from 2% to 8% over the decades. These cells are used in pocket calculators and in electronic watch battery power chargers. Possible future materials for high efficiency photovoltaic cells include amorphous Si-N and Si-Ge films, produced by glow discharge decomposition of appropriate gas mixtures [67].

2.4 Magnetic Properties of Thin Films

Every material consists of atoms in which electrons having a gyroscopic spin also execute orbits around the atomic nucleus. Two magnetic moments, the spin and the orbital magnetic moment, are associated with these two kinds of electronic motion. Magnetism, which is quantum mechanical in nature, is therefore an electronically driven effect and results in a number of interesting properties. Some of these properties are

discussed in the following sections, which draw heavily from the treatment in References [67] and [79].

Magnetism occurs due to electron spin and the Pauli exclusion principle, which forbids electrons of like spin to occupy any given energy level. In free space, a magnetic field of intensity H induces a magnetic flux density B given by

$$B = \mu_0 H , \tag{2.4.1}$$

where μ_0 is the permittivity of free space. If the magnetic field is acting in a general material medium the magnetic flux density becomes

$$B = \mu_0 H + M , \tag{2.4.2}$$

where M is the magnetic moment per unit volume of the material. This equation defines an absolute permittivity given by

$$\mu = B/H = \mu_0 + \kappa , \tag{2.4.3}$$

with $\kappa = M/H$ being the absolute magnetic susceptibility of the material.

The relative permeability of the material μ_R is the ratio of the absolute permeability of the material to the absolute permeability of free space and is represented as

$$\mu_R = 1 + \kappa_R , \tag{2.4.4}$$

and $\kappa_R = \kappa/\mu_0$ is the relative magnetic susceptibility. A magnetic dipole has magnetic poles of equal strength h but opposite polarities. If the dipole has length l and cross-sectional area a, the pole strength per unit area is $M = h/a$, or expressing the magnetic strength in terms of the volume al of the dipole material, the magnetic volume per unit volume is obtained as $M = hl/al$. The energy due to the interaction of M with an external magnetic field H is given by

$$E_H = -M \cdot H . \tag{2.4.5}$$

The saturation magnetization is given by the sum of the moments of the number density of perfectly aligned magnetic dipoles. Magnetic poles result in regions where $\nabla \cdot M = 0$. These poles usually occur at surfaces and at irregularly shaped regions. The field due to the magnetic poles is demagnetizing inside the sample and tends to stray outside it. Representing the demagnetizing and stray fields by H_{mag} and H_s respectively, the demagnetizing field interacts with the magnetic moment to produce a potential energy V_{mag} given by [67]

$$V_{mag} = -\frac{1}{2} V_{mag} \cdot M . \tag{2.4.6}$$

The potential energy due to magnetic anisotropy V_a depends on the orientation of M relative to the sample coordinates [67]

$$V_a = K_a f(\alpha_1, \alpha_2, \alpha_3), \tag{2.4.7}$$

where K_a is the anisotropic constant for the material. The direction of maximum potential energy is referred to as the hard axis, while the direction of minimum potential energy is referred to as the easy axis.

There are a wide variety of classical and quantum mechanical manifestations of magnetism such as diamagnetism, paramagnetism, ferromagnetism, antiferromagnetism, ferrimagnetism, magnetorestriction, magneto-optical phenomena, and magnetoresistance. Free atoms of some elements do not have a magnetic moment because the various orbital and spin magnetic moments cancel each other. However atoms of some elements do possess a non-zero net magnetic moment. Bulk magnetic properties depend on crystal specific inter-atomic potentials and temperature as well as on the magnetic moments of constituent atoms. Diamagnetic materials consist of atoms or molecules that do not have a magnetic moment, but when exposed to an applied magnetic field, produce a component of magnetic moment that opposes the applied field direction, while having their orbital and spin axes precess about the applied field direction with a constant angle and with an angular velocity proportional to the field intensity. Diamagnetic materials therefore have negative absolute and relative susceptibilities and their values of the relative permittivity are usually much less than unity. Paramagnetic materials are composed of atoms or molecules that each possesses a magnetic moment, which aligns in the same direction as an applied magnetic field. Paramagnetic materials have positive temperature dependent susceptibilities that are of much greater magnitude than diamagnetic susceptibilities.

Ferromagnetic materials have relative susceptibilities much greater than unity so that the relative susceptibilities of these materials are almost equal to their relative

permeabilities. Ferromagnetism is due to the alignment of uncompensated $3d$ electron spins by quantum-mechanical exchange in materials such as elemental Fe, Co, Ni and alloys of these elements. Magnetic saturation in these materials can be achieved in fields having an order of 10^4 lower intensity than fields required to achieve the same level of saturation in paramagnetic materials. Powerful internal fields of force exist in ferromagnetic materials that act to align the atomic magnetic moments, overcoming the disordering effects of temperature at moderate temperatures. Examples of ferromagnetic materials are the oxide Fe_3O_4 and ferrites.

Antiferromagnetic materials consist of atoms that have magnetic moments and that are subject to strong internal fields, which result in antiparallel directions for the magnetic moments of near-neighbor atoms. The atomic magnetic moments are therefore ordered, but in two opposite directions leading to a zero resultant magnetic moment over any finite volume of the antiferromagnetic material. The temperature dependence of the susceptibility in these materials exhibits a sharp discontinuity at a temperature known as the Néel temperature. Below the Néel temperature, antiferromagnetic materials are ordered and exhibit antiferromagnetic behavior, while above it they are disordered and behave paramagnetically. Manganese and chromium are examples of antiferromagnetic materials. Ferrimagnetic materials consist of two sets of atoms having antiparallel magnetic moments of unequal magnitudes, resulting in a non-zero net magnetic moment.

Magnetorestriction is the shape-change of a body induced by a magnetic field. Linear magnetorestriction is the strain derivative of the anisotropy energy, being the coupling between the elastic strain and the direction of the magnetic moment. The strain of lattice mismatch at substrate-film interfaces greatly influences magnetic properties by

magnetorestriction. Magnetic anisotropy is the energy resulting from a specific orientation of the magnetic moment relative to the principal crystal axes. Magneto-optical phenomena are due to coupling between electron spin and charge polarization. Magnetoresistance is the phenomenon whereby the resistance of a non-ferromagnetic metal increases with the intensity of an applied magnetic field. Magnetoresistance is used to measure Fermi surfaces for pure, high electronic mean free path materials.

Thin films of metal on metal substrates exhibit a wide range of interesting and useful magnetic properties, for example ultrathin (3 monolayer) fcc Fe(001) films grown on bulk Cu(001) substrates exhibit magnetic anisotropies [78] and Fe/Cu/Fe and Co/Cu/Co multilayers exhibit antiferromagnetic coupling [79]. Ferromagnetic thin films are currently being investigated for use in creating spin-polarized electrical current in a semiconductor substrate, a process known as spin-injection [80]. Spin-injection will be technologically useful if three criteria are satisfied, namely substantial spin-polarization of the spin-injected current, having electrons as the dominant charge carriers, and having source Curie temperatures not less than room temperature. The search is now on for materials satisfying the above three criteria. Other interesting magnetic materials include magnetic films grown pseudomorphically on semiconductor substrates, rare-earth metal films on transition metal substrates, multilayer systems consisting of combinations of ferromagnetic and normal metals, ferromagnetic and superconducting metals and layers of rare-earth films. Magnetic systems such as these are used in applications such as surface and low dimensional magnetism, magneto-optics and many others and are the focus of ever-increasing scientific research efforts as capital investment and research in

magnetic recording technology continues to grow with the ever-increasing necessity to acquire, store and retrieve information, as demanded by the information revolution.

It is inevitable that thin film magnetic materials will be integrated into advanced solid state devices in the near future. A fundamental understanding of thin-film magnetism is essential for the further development of applications such as these in magnetics technology. A few technologically important magnetic phenomena occurring in thin films and their applications are discussed in the following sections.

2.4.1 Reversal of Magnetism in Uniaxial Thin Films

Perfect parallel alignment of a magnetic moment occurs at the minimum magnetic exchange energy. Any deviation from this parallel alignment increases the exchange energy by the amount [67]

$$V_{ex} = \frac{K_e}{M}\left|\nabla \cdot M\right|^2, \tag{2.4.1.1}$$

where K_e is the exchange constant characterizing the resistance of the magnetic moment to attempts to change its direction. The magnetic moment equals the saturation magnetization inside magnetic domains, which are separated by walls. Magnetic domain wall energy depends on thin film thickness. There are several kinds of domain walls including the Bloch wall, where the magnetic moment rotates perpendicularly to the film plane, the two-dimensional Bloch wall, where the magnetic moment rotates in two directions perpendicular to each other in the film plane, the Néel wall, where the magnetic moment rotates in the film plane, and the Cross tie wall, which is an array of Néel walls separated by Bloch lines where the magnetic moment rotates out of the film

plane. The magnetic moment orientation in a given thin film sample is the direction corresponding to a minimum in potential energy. The magnetic moment in thin films therefore normally lies on the film surface, where the stray field intensity is not very strong.

Binary switching in thin films exploits the phenomenon of uniaxial anisotropy. Anisotropy in thin films can have several origins, one being the intrinsic crystalline anisotropy, such as occurs in Fe and Ni, where the directions for minimum potential energy are the [100] and [111] directions respectively. For example, cubic magneto-crystalline anisotropy occurs in bulk Fe, which causes four-fold in-plane anisotropy in Fe(001) thin films, with the hard axis oriented parallel to the [110] direction and the easy axis oriented parallel to the [100] direction. The other causes of anisotropy are magnetorestriction and anisotropy induced by external fields during film growth. Switching by coherent rotation of M in single domain films can be described by the Stoner-Wohlfahrt model according to [67]

$$V_e = -M_s H_x \cos\theta - M_s H_y \sin\theta + K_a \sin^2\theta . \tag{2.4.1.2}$$

The equilibrium orientation of M is achieved when [67]

$$\partial V/\partial\theta = M_s H_x \sin\theta - M_s H_y \cos\theta + 2K_a \sin\theta\cos\theta = 0 . \tag{2.4.1.3}$$

The stability of the equilibrium orientation of M depends on the sign of the second derivative of V [67],

$$\partial^2 V/\partial\theta^2 = M_s H_x \cos\theta + M_s H_y \sin\theta + 2K_a\left(\cos^2\theta - \sin^2\theta\right). \tag{2.4.1.4}$$

When the external field is parallel to the easy axis, two stable orientations: $\theta = 0$ and $\theta = \pi$, are possible for $-H_k < H_x < H_k$, where H_k is the extrinsic coercivity. Outside this range, the only stable orientation is that parallel to the orientation of H. Coherent 180° rotation of M in the whole film causes switching at the critical fields $\pm H_k$. When the external field is parallel to the hard axis, all orientations θ are stable. M rotates toward the direction of H and achieves complete alignment with H for $H \geq H_k$. Extrinsic coercivity is not observed in this case. For an arbitrary H direction, all magnetic field vectors point in a direction defined by the magnetic energy. The H orientation becomes unstable for fields at which $\partial^2 V / \partial \theta^2 = 0$.

The Stoner-Wohlfahrt curve, which obeys [67]

$$H_x^{2/3} + H_y^{2/3} = H_k^{2/3}, \tag{2.4.1.5}$$

and is referred to as the switching curve, results from the above condition. M_s is oriented parallel to the tangent from the tip of H on the curve. When H is inside the switching curve, two stable directions of M result, making switching impossible while H remains within the curve boundaries. If the absolute value of H exceeds the value of the extrinsic coercivity, then rotation of H just induces a continuous rotation of M without any switching.

Magnetic moment reversal by domain growth occurs after a threshold value of the magnetic field is exceeded, so that domains of opposite polarity migrate in the film until the direction of M is reversed. Magnetic moment reversal by partial incoherent rotation and domain motion takes place when local perturbations cause M to fluctuate locally

around the average orientation parallel to the direction corresponding to minimum potential energy. M is rotated in either a clockwise or anti-clockwise direction by an induced anti-parallel filed, causing the creation of domains as the external field strength is increased. The reversal is completed by domain growth when the nucleation and wall motion thresholds are surpassed. The switching time of M can be studied by several methods such as by using the magneto-optic Kerr effect to perform high-speed film imaging of domains. Setting the bias field perpendicular to the easy axis, making H_y constant, while the pulsed field is parallel to the easy axis, making H_x a square wave pulse.

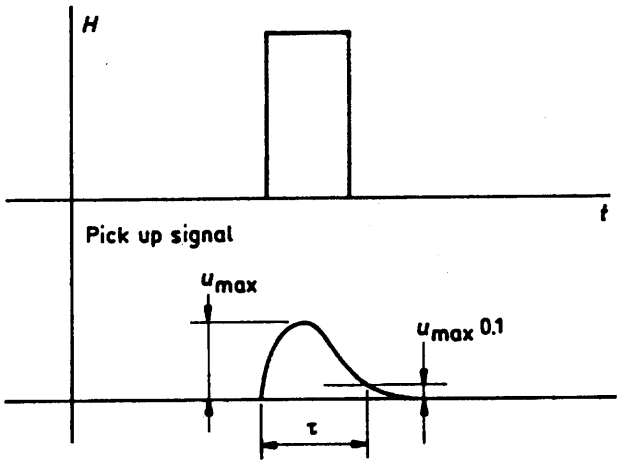


Figure 2.6 Fast switching with pulsed fields (taken from Reference [67]).

Three types of processes causing magnetic moment switching can be differentiated and they are M reversal by wall motion, M reversal by diffuse movement in the direction of minimal potential energy and M reversal by its coherent rotation in the whole film. The fastest moment reversal mechanism is the mechanism mediated by coherent rotation of M where the switching times are on the order of nanoseconds.

2.4.2 Applications of Magnetic Switching

Permalloy, 81Ni19Fe, thin films used in magnetic switching devices can be produced by evaporation or by magnetron sputtering at high temperatures [67]. For high speed random access memories, two kinds of currents, the word current I_W and the digit current I_D are used to generate magnetic fields H_x and H_y in the x and y direction respectively [67].

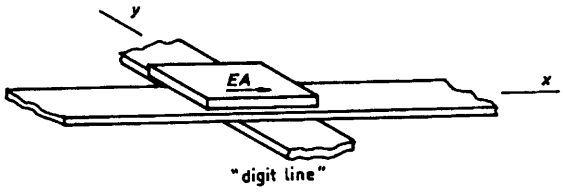


Figure 2.7 Configuration for high speed RAM (taken from Reference [67]).

In "write" mode, the magnetic moment is rotated in the direction of the hard axis by H_x while H_y pushes it in the positive direction of the easy axis representing store 1, or in the negative direction of the easy axis representing store 0. The rotation of M induces a voltage, which is picked up by the digit line in "read" mode. Ideally $I_{D,\min} \geq 0$ while I_W is at its maximum value and $I_{D,\max} \leq H_k 2b$, where H_k is the extrinsic coercivity and b is the conduction line width, while I_W is equal to zero. Practically, to avoid wall motion switching of M, $I_{D,\min}$ must not be greater than $H_{cw} 2b$ and also must be able to switch regions which are not completely aligned to the hard axis. Pulse sequences in "read" and "write" mode are shown in Figure 2.8.

Domain storage devices exist in two forms; the class of domain wall motion devices utilizes M parallel to the film plane, while the class of bubble devices utilizes M perpendicular to the film plane.

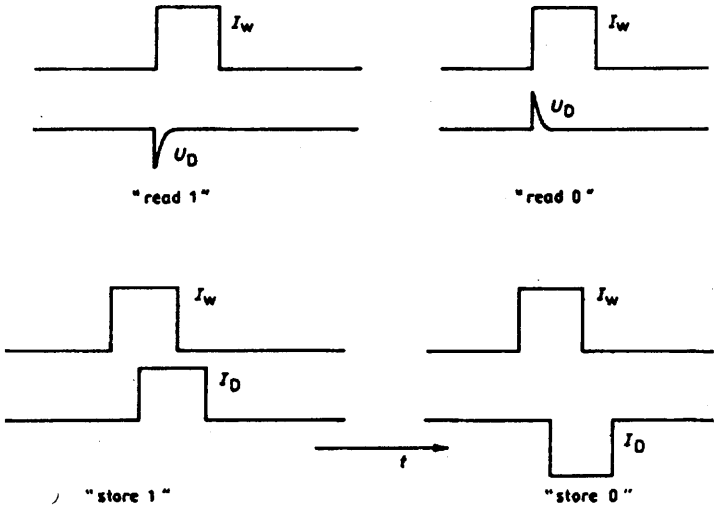


Figure 2.8 Pulse sequences in "read" and "write" mode (taken from Reference [67]).

In both device classes, domains are moved electronically to desired locations all over the film. Domain wall devices are composed of overlaid thin uniaxial films with the easy axis lying on the film plane. Fields used in these devices must lie inside the range $H_{cw} < H < H_k$ in order to achieve fast domain movement and to avoid the creation of new domains. The bit rates and bit densities of domain shift registers are respectively, 500kHz and 300 bits/cm^2 [67]. Bubble devices consist of thin films having spontaneous magnetization perpendicular to the film surface. Materials with this property have a high and constant magnetic anisotropy and low saturation magnetization and include rare earth ferrites and amorphous films. The magnetization condition for bubble domain formation is shown in Figure 2.9.

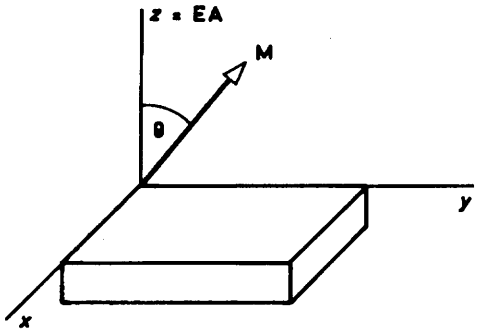


Figure 2.9 Magnetization condition for bubble domain formation (taken from Reference [67]).

Minimization of the stray field causes the formation strip domains to have antiparallel orientations. Strips with undesired orientations are shrunk by applying external fields until the spherical domains become bubble-like in shape. The bubble diameter is a technically important parameter for storage capacity. The characteristic length [67]

$$l = V_{wall} / 4\pi M_s^2 , \tag{2.4.2.1}$$

arises in the course of the variational determination of the equilibrium shape and distribution of the domains. Optimal values of the film thickness z are attained for $z = 4l$, resulting in a bubble radius of $4l$. Bubble boundaries can be Bloch walls of unique rotation sense or split rotation sense. Bubble boundaries with split rotation sense lead to bubbles that require higher intensity fields to effect their movement. Bubble movement can be achieved by local fields produced by thin film conductor loops or by rotating fields. In high capacity bubble devices, which usually incorporate a major-minor loop mechanism, bubbles are locally created by a bubble generator and when needed, are introduced into the major loop at a transfer gate while bubble transport into minor loops

occurs at other gates. To produce a read-out in these devices, magneto-optical means, the Hall effect or magnetic resistance are used. Bit rates and packing densities for bubble devices are respectively, of the order of MHz and Mbits/cm^2.

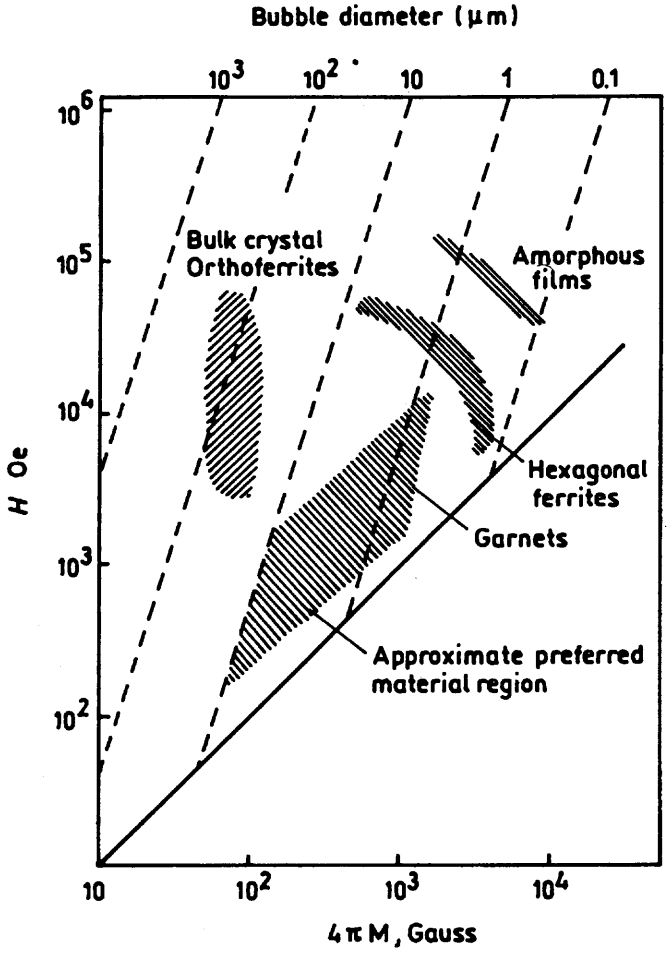


Figure 2.10 Some materials used for bubble domains (taken from Reference [67]).

2.4.3 Structure Dependence of Thin Film Magnetic Properties

The subtle effects of film growth and morphology on the magnetic properties of thin films are still not completely understood partly because magnetic and structural properties are strongly interdependent. In the past, the relevant coefficients describing subtle magnetic phenomena such as magnetorestriction and anisotropy were extracted

from combinations of physical models and experimental data [79] because the theoretical study of these subtle magnetic properties with electronic structure calculations at the time suffered from several limitations including the lack of exchange-correlation functionals accurate enough to describe these properties and the computational difficulties involved in the inclusion of the necessary relativistic spin-orbit coupling terms in the Hamiltonian equation used to describe these systems.

A variety of exchange-correlation functionals accurate enough to treat all kinds of subtle quantum-mechanical effects are now available and recent electronic structure calculation methodologies routinely include the spin-orbit coupling terms in the Hamiltonian operator occurring in the Schrödinger equation, making accurate determination of these magnetic properties quite routine for a wide variety of materials. It is now possible to model ideal monolayers of film materials with computer simulations, while it has always been very difficult for experimentalists to produce similar systems in the laboratory. Practical difficulties arise when attempting to grow monolayers because substrates are needed to support them, and substrate-film interactions determine the film properties to a large extent. For example at high temperatures an intermixed buffer layer is formed in the Fe/Cu(100) system, which results in the stabilization of the antiferromagnetic phase of fcc Fe, while at low growth temperatures, the more discrete film substrate interface results in the stabilization of the ferromagnetic phase of fcc Fe [79].

The extent of film-substrate interaction is determined by electronic interactions, which are controlled by the degree of band filling in the substrate and film materials. Substrates with filled d bands will generally be less reactive with film materials, but

such substrates have lower surface free energies, which can cause surface segregation due to thermodynamic activation forces. Transition metal substrates generally have higher surface free energies. Hybridization between magnetic d or f electron states and substrate d electron states across the film substrate interface can be the dominant factor in determining magnetic properties. On Ru(0001) substrates, Fe(111) grows with an in-plane lattice expansion which normally should facilitate ferromagnetism, but the first monolayers are observed to be magnetically inactive. This observation is explained by the in-plane expansion, which causes an inter-planar contraction and strong Fe-Ru electron band hybridization, making a magnetic moment impossible. The opposite effect of band hybridization on magnetism is seen in the Fe/Pd(100) system where the inducement of ferromagnetism in the Pd substrate due to strong d - d hybridization is predicted [42].

The {100} surfaces of Cr have been theoretically predicted to be ferromagnetic, with relatively large surface magnetic moments. The ferromagnetic (100) sheets in adjacent Cr layers are antiparallel to each other. The surface can be demarcated into antiparallel domains if the domain walls are thicker than surface terrace widths, leading to a zero macroscopic magnetic moment. Experimental techniques with enhanced resolution are being developed to investigate this problem [79]. Magnetic moments tend to be enhanced at the free surfaces of weakly magnetic metals, or created in metals that are almost magnetic. The experimental data, although often contradictory, tend to confirm these theoretical results. Experimental confirmation of theoretical predictions by direct measurement of surface specific and interface specific moments are still a work in progress [79]. The large magnetic moments at {100} chromium surfaces have been

calculated numerous times and observed indirectly, but have not yet been directly observed experimentally.

Naturally occurring magnetic materials exist in many crystallographic and magnetic phases. Thin film deposition of these materials results in an even greater variety of magnetic structural phases, due to film-substrate stabilization forces. Sometimes many of these stabilized thin film magnetic structures are new, never previously observed structures because thin films can exhibit a far wider range of structural order than the corresponding bulk materials. These artificially stabilized structures are known as metastable phases. The study of metastable phases therefore has a special relevance to thin film magnetic properties. These metastable phases provide stiff tests for the state-of-the-art in *ab-initio* computational methods. At a minimum, these methods are able to compute the zero temperature magnetic moment of an elemental crystalline material and the total energy as a function of crystalline structure and lattice parameter. The inability of any *ab-initio* computational method to predict the known magnetic phases of elemental magnetic materials such as iron is often used to identify inadequacies in the underlying theory [79]. Work on metastable phases has shown that the cubic phases observed in nature are a special case of the more general body centered tetragonal (bct) structure in materials, and that bct phases are most likely to be stabilized epitaxially. The in-plane strain resulting from lattice mismatch between deposited thin films and their substrates will invariably lead to a relaxation of inter-planer spacing to form a variety of structures, any one of which can be stabilized if it is close to a metastable tetragonal phase correctly predicted by theory. The need to improve the accuracy and sensitivity of characterization

techniques is paramount in order for the proper characterization of the structural, electronic and magnetic properties of these new phases.

A substantial amount of research has gone into understanding the diverse structural and magnetic properties of various thin film/substrate systems. Due to the interdependence of structural and magnetic properties of thin films, a detailed knowledge of structural properties is essential for the understanding of their magnetic properties in all their subtlety. In order to study magnetic properties, it is essential to solve the structural problem first because magnetic energies are orders of magnitude smaller than cohesive energies. Energies resulting from crystalline structure modifications are about of the same order (0.1 eV) as some energies resulting from changes in magnetic states [79], therefore magnetic properties of thin films often depend on the fine details of thin film preparation such as deposition rate, substrate type and temperature, impurity levels, and so on. For example ferromagnetic or antiferromagnetic fcc Fe can result from epitaxial growth of Fe on Cu substrates depending on details such as the substrate cleanliness and temperature. *Ab-initio* quantum-mechanical based calculations have also been useful in understanding the phases of Co. The calculations predict two cubic phases, fcc and bcc for Co. The naturally observed phase of cobalt is the high temperature ferromagnetic fcc phase, the bcc phase does not occur naturally but a ferromagnetic bcc phase of Co has successfully been grown on GaAs [79], confirming the theoretical prediction. Also by comparing total energies for a subset of magnetic structures, magnetic structural phases can be predicted, as in the case of bulk manganese where the possible antiferromagnetic phases have been predicted.

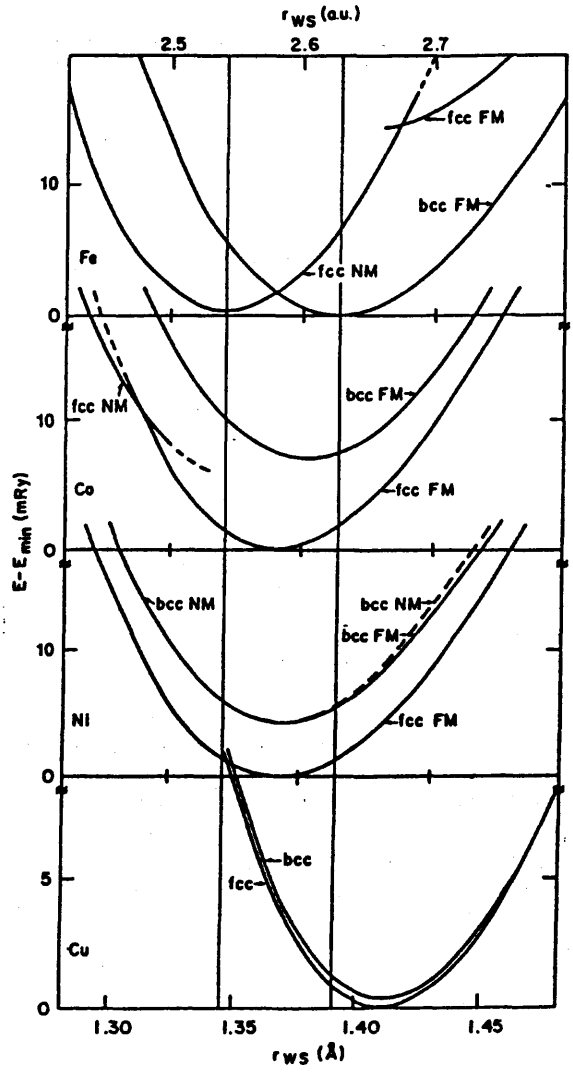


Figure 2.11 Total energy per unit cell for several metals calculated as a function of the Wigner-Seitz radius (taken from Reference [79]).

2.5 Tantalum Thin Films

Ta films are very useful in the microelectronics industry, especially as an underlayer material in computer ICs [61, 66]. The aluminum interconnects currently being used in computer ICs need to be replaced with copper (Cu) ones in order to achieve greater computing power. Cu offers lower electrical resistivity and a much higher resistance to electro migration, but it diffuses into silicon and silicon oxide much more rapidly than

aluminum [61], thereby degrading computing device properties. Diffusion barriers are therefore needed to prevent the intermixture of Cu with silicon and silicon oxide. Ta is an excellent candidate for diffusion barriers in high-speed ICs because Ta and Cu have negligible mutual inter-diffusion coefficients [66] and Ta possesses a high resistance to chemical attack while providing good adhesive properties. Ta films are usually deposited by sputtering in IC manufacturing because the electrical properties of sputtered Ta films can be adjusted by controlling the sputtering atmosphere [81], making sputter deposition the dominant technique for producing Ta films used in electronic applications. A schematic diagram of a typical plant used for large scale sputtering of Ta is shown in Figure 2.12.

An oxide of Ta, Ta_2O_5 or tantalum pentoxide, is a high-k dielectric used in Dynamic Random Access Memory (DRAM) applications. Ta_2O_5 films grown thermally from thin films of β-Ta have also been used as an etch stop layer of glass on ceramic substrates for thin film RC network applications [82].

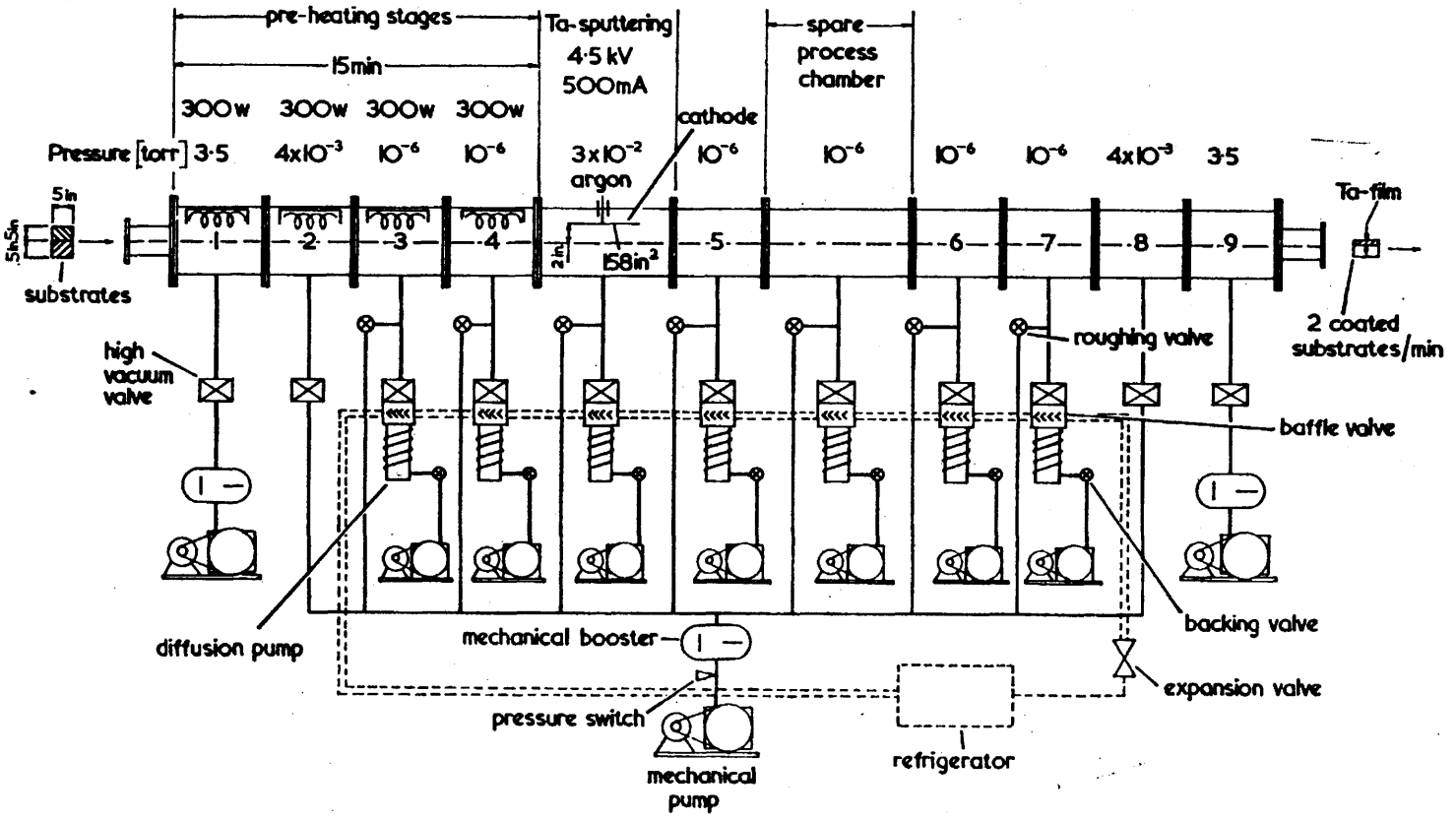


Figure 2.12 Air-to-air vacuum plant for Ta sputtering (taken from Reference [61]).

2.5.1 Electrical Conduction in Tantalum Thin Films

Ta films exhibit a negative temperature coefficient of resistance (TCR) when deposited under oxidative conditions. Thin films with a negative TCR have all been observed to possess a granular structure [83]. Granular structured films form crystallites or crystallite aggregates, which are less than 100Å in diameter, and are separated by voids. Granular structured films are usually considered as three-dimensional analogs of ultra-thin discontinuous films [83], and conduction through these films are usually described with theories used to describe conduction in discontinuous films. The TCR of oxygen–doped Ta thin films has been observed to stabilize more rapidly when heated in air at 250°C, but to a more negative value than the TCR of undoped Ta films [66]. This observation was explained by noting that in the case of oxygen doped Ta films, the Ta grains are completely surrounded by the oxide layer, which is not the case in undoped films. Subsequent studies [66] of I-V curves belonging to tunnel emission diodes formed from anodized Ta thin films confirmed that Schottky emission was the dominant conduction mechanism in oxygen doped Ta thin films.

A systematic study [81] of the effect of oxygen incorporation on the low temperature electrical properties of Ta films concluded that electrical conduction in Ta thin films was due to a dual mechanism consisting of a metallic part and a quantum tunneling part. The room temperature resistivity was found to increase directly with oxygen content, the relationship approaching an exponential one at the highest oxygen concentrations. However, the resistivity increase at high oxygen concentrations was attributed to the formation of β-Ta, and the subsequent crystal structure-induced change in the phonon spectrum, which describes scattering by charge carriers. The fact that the

quantity $\Delta\rho/\Delta T$ for Ta thin films less than 1,000Å in thickness is higher than for the bulk [84], contrary to what Matthiessen's rule would predict for an impurity dependent resistivity, suggested that other explanations of the relatively high values of $\Delta\rho/\Delta T$ and ρ in Ta thin films had to be pursued. After it was shown that quantum tunneling took place between metallic islands in a thin, discontinuous film [85], it was demonstrated [81] that ρ and $\Delta\rho/\Delta T$ can increase together when electrical conduction is due to metallic conduction and quantum tunneling. $\Delta\rho/\Delta T$ values for thin films of Ta [86] were found to be positive when metallic conduction is dominant and negative otherwise [87]. However, metallic island structures were observed [88] in thin Ta films evaporated onto sodium chloride, alumina, glass and fused quartz substrates. An increase in ρ was observed as the deposited films were exposed to oxygen and nitrogen, however, the temperature dependence of ρ was unaffected.

The electrical resistances of Ta films 250 – 2,400Å thick have been stabilized by heating to 500°C and then cooling down to 25°C [87]. It was noted that the resistivities of these stabilized films decreased with thickness and increased linearly with temperature from 25°C to 100°C. However the resistivity increased with a different slope between 180°C and 500°C. The two different slopes in the thin film Ta resistivity versus temperature diagram were attributed to the presence of two different phases in the Ta films. The parameters controlling the electrical properties of evaporated Ta films are still not well understood, however the structure of the thin film appears to be the most important factor for films less than 1,000Å [87].

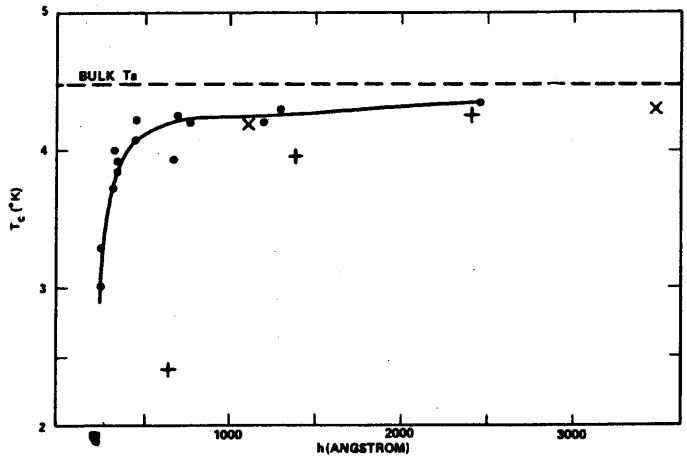
2.5.2 Superconductivity in Tantalum Thin Films

The superconducting transition temperature, T_c, for Ta has been determined [89] to be 4.477K. Ta thin film T_c values have been discovered to increase with film thickness and to approach the bulk Ta value for films thicker than 1,000Å [83]. For example, a value for T_c of 3.0K as well as an electrical resistivity value of 200μΩ cm was obtained for a 250Å thick Ta film. The high value of T_c made it improbable that the high value of the resistivity was due to impurities in the Ta film. The high value of T_c was possibly typical of islands of Ta filaments on the film, separated by air or a different phase of the metal [90]. The presence of impurities such as nitrogen is thought to decrease T_c [83]. Ta films of 9,000Å and 10,000Å thickness having impurity levels of 5% had T_c values of 4.45K and 4.51K respectively. These results could indicate that impurity levels were not affecting the value of T_c, or that the impurities were not uniformly distributed around the film, lowering their effect on T_c.

The temperature range, ΔT_c, over which evaporated Ta thin films are superconducting has been determined to increase as thin film thicknesses decrease, i.e. as T_c decreases [83]. The increase in ΔT_c was attributed to the increase in thin film impurity concentration with decreasing film thickness, as evidenced by the higher values of ρ and T_c. Bcc Ta thin films superconduct at temperatures reasonably close to the bulk T_c value, but thin films of β-Ta sputter-deposited on silicon substrates did not exhibit superconductivity above 1K [91]. Ever since the recognition that impurities cause the reduction of T_c in bulk Ta, efforts have been focused on reducing the impurity content of

sputtered Ta films in order to achieve thin Ta film T_c values very close to the bulk value [83]. Getter sputtering techniques were used extensively in this regard, resulting in thin Ta films having T_c values close to the bulk value. A low impurity content getter sputtering system was used to deposit thin Ta films on MgO substrates [92], and annealing was observed to increase the residual resistance ratio $\rho_{300K}/\rho_{4.2K}$ from 30 to 1800 by impurity removal. The activation energy for removal of the interstitial impurities was determined to be 35 kcal mol^{-1}. T_c was also observed to increase from 3.16K to 4.25K for a thickness increase of 215 - 89,000Å for the Ta films deposited on MgO substrates at 350°C. This result was attributed to the grain size, which increased from 30Å to 1,000Å. The presence of normal bcc Ta in the grain boundaries of the deposited film between superconducting Ta grains was believed to decrease the electronic mean free path to 5Å.

The observation [83] that –80 - -105V bias-sputtered Ta films with temperature coefficient of resistance α between 655 and 1675 x $10^{-6}C^{-1}$ exhibited T_c values in the range 3.4 - 3.8K, while –30V bias-sputtered Ta films with an α value of –5.2 x $10^{-5}C^{-1}$ exhibited T_c values below 2K led to the conclusion that sputtered Ta films with high resistivity usually have lower values of T_c than sputtered low resistivity films, because the temperature coefficient of resistance of Ta is known to change sign at a resistivity of about 200μΩcm. A sputtered Ta film with a resistivity of 200μΩcm has been observed to exhibit superconductivity in the temperature range 0.49 - 0.57K [93].



Superconducting transition temperature, T_c, of evaporated tantalum films as a function of film thickness (*b*).
• Marcus (1966).
X Gerstenberg and Hall (1964).
+ Khukhareva and Shalnikov (1954).

Figure 2.13 Thickness dependence of T_c for thin Ta films (taken from Reference [83]).

Another work [94] has confirmed the tendency of the T_c of thin Ta films to decrease with increasing thickness. The extrapolation of the measured critical magnetic field, H_c, of the 3,460Å thick thin films to 0K revealed that the critical field at 0K was an order of magnitude higher than that previously obtained for Ta [89]. This was attributed to the presence of a superconducting filament network, which persisted at magnetic fields higher than the value of H_c for bulk Ta. The presence of a superconducting filament network was ascribed to a random distribution of impurities in the Ta films.

2.5.3 Structural Properties of Thin Tantalum Films

The fact that thin Ta films are rarely homogeneous, *i.e.* composed of a single phase, is suggested by most of the electrical property data. In fact, Ta films thicker than 1,000Å are often not homogeneous [95], with various crystalline structures being observed in the films. However, thin films less than 2000Å deposited on freshly cleaved (001) faces of

MgO have been observed to have the bcc structure of bulk Ta, while lattice distortion occurred in the thinner films [86]. Apart from the regular bcc phase, fcc phases [87 – 89, 92, 94, 96 - 98] and the β-phase have been observed in Ta thin films. The fcc and β-phases are usually formed as relatively thin films, and are very difficult to obtain in bulk form.

2.5.3.1 Fcc Phase of Tantalum

Single crystal bcc Ta films deposited on freshly cleaved (001) MgO at 700 and 750°C have been reported, as well as an fcc Ta phase which was observed in 100Å thick films on MgO substrates at 400°C [97]. The lattice constant of the fcc phase was determined to be 4.42Å. The fcc phase was also observed on evaporated carbon and silicon oxide films. The fcc phase was observed exclusively in films less than 50Å thick, but was gradually replaced by bcc films as the film thickness increased to 250Å, where only bcc structure was seen. The fcc phase in Ta thin films was also observed by other authors [96, 98] who confirmed the replacement of the fcc phase by the bcc phase as the film thickness increased. The fcc phase was observed to be formed on a wide variety of substrates and over a wide temperature range, ruling out stress stabilization as a reason for fcc phase formation. Pseudomorphism was also ruled out because the fcc phase was observed in both epitaxial and polycrystalline films. Coulombic interaction [104] between islands was ruled out also as a stabilizing effect by observing that changes in the electron to atom ratio from 3 to 10^{-4} had no effect on the formation of the fcc phase.

Mixtures of the bcc and fcc phases have also been reported for 1.2 – 1.8μm thick Ta films deposited on sapphire [100]. The only reflection observed from the fcc phase in

the films was the (111) reflection. Fiber textured [111] bcc Ta films were obtained on a (00.1) sapphire substrate between 20 and 250°C, while epitaxial Ta(111) || Al_2O_3 (00.1) films were obtained between 400 and 450°C on the same substrate. At 600°C, a mixture of fiber textured bcc and fcc phases was obtained. At higher temperatures no fcc phase was observed. Between 800 and 1000°C, epitaxial bcc Ta(110) || Al_2O_3 (00.1) films were achieved. A 4% impurity content [86, 96] was estimated for the fcc phases. It was conjectured [86] that the fcc phase was stabilized by impurities due to the presence of a monolayer of oxygen on the substrate surface prior to deposition. The transformation of the fcc phase to bcc with increasing thickness was attributed to a decrease in the level of incorporated impurities below 4%. The bcc Ta was believed to form by a 21% contraction of the <001> lattice vector and a simultaneous 12% <110> vector elongation in two fcc unit cells, to form a bcc unit cell having almost the same volume as the fcc unit cell.

Fcc thin film structures have been observed for TaO [101 - 102] with lattice parameters of 4.47Å and 4.38Å, TaN with a lattice parameter of 4.48Å [103], and TaC with a lattice parameter of 4.46Å [104]. Island sizes of evaporated Ta were found to be independent of film thickness [105], below a film thickness of 250Å, but increased with decreasing pressure and increasing temperature during the film deposition process. It is interesting to note that fcc phases have also been observed in niobium thin films [83]. The formation of the fcc niobium phase was attributed to the effects of the interaction of the residual gas in the system with the deposited niobium.

Bcc Ta is used for thin-film interconnections, while β-Ta is used for thin-film resistors because of its high resistivity [66]. Ta metal is used in super alloys with nickel

and cobalt for aerospace applications, and in specialty metal products for the chemical industry. Ta is also an attractive material for use as an anti-corrosive coating, because of its high ductility and chemical stability.

2.5.3.2 The β Phase of Tantalum

The β phase of Ta (β-Ta) in thin films was first observed in 1965 [93], and then it was found in thicker bcc Ta films and also in an fcc and bcc thin film mixture deposited on MgO. The fact that epitaxial bcc films were formed from very thin films of Ta on cleaved (001) MgO faces containing the β phase when the temperature was raised to 700°C, led to the suggestion that the β phase of Ta was an even purer form of the metal than the bcc phase.

β-Ta occurs mostly as a thin film and is not easily formed as bulk material. It is harder, more brittle, and less ductile than the bcc phase. Densities ranging from 16.3g cm $^{-3}$ [93] to 16.9g cm $^{-3}$ [106] have been calculated for this phase assuming tetragonal unit cells containing 16 atoms and 30 atoms respectively. β-Ta is known to have a 45% smaller average grain size than bcc Ta, and to frequently contain a high concentration of impurities and defects, in addition to having a resistivity (170 - 210μΩ cm) much higher than that (15 – 60μΩ cm) of the stable bcc phase [95, 107 - 109]. It has been reported to nucleate preferentially on certain kinds of substrates, [91, 110 – 113] and may be stabilized by oxygen impurities [91, 108, 110, 113 - 115].

There have been several structures proposed for β-Ta. The structure of β-Ta has been reported as a tetragonal super-lattice structure having an *a* lattice constant of 10.29Å and a *c* lattice constant of 9.2Å [114], as a tetragonal $P4_2/mnm$ structure with an

a lattice constant of 10.19Å and a *c* lattice constant of 5.3Å containing 30 atoms per unit cell [106], and an *a* lattice constant of 5.34Å and a *c* lattice constant of 9.9Å [81]. A hexagonal structure with an *a* lattice constant of 34Å containing 144 hexagonal sub-cells with an *a* of 2.83Å and *c* of 5.3Å has also been reported [116].

On silicon and other substrates, β-Ta seems to form with a strong degree of preferred orientation in the (100) direction [91, 117]. In contrast, it was observed to be formed with random orientation on amorphous substrates such as carbon, glass, quartz, Ta_2O_5 and SiO_2 [91, 110]. Polycrystalline metals such as gold, platinum, rhenium, beryllium and tungsten with strong resistance to oxidation below 673K were observed to nucleate the alpha bcc phase of Ta in sputtering processes. Conversely, β-Ta was nucleated on polycrystalline metals susceptible to oxidation [110]. Depositing niobium onto silicon and amorphous carbon substrates and then sputtering from Ta targets resulted in the formation of β-Ta. This was true as long as the niobium underlayer was deposited in an inert atmosphere. Traces of water vapor in the sputtering system resulted in the deposition of β-Ta, even with the niobium underlayer [91].

In another reported experiment [118], negative voltage bias was applied to the substrate surface during Ta film deposition. It was subsequently shown from the results of these experiments that the amount of impurities incorporated into the sputtered Ta film is proportional to the magnitude of the applied voltage bias. The structure of bcc Ta was found to be less sensitive that that of β-Ta to impurity incorporation. The link between the structures of Ta films deposited by sputtering and the amount of incorporated impurities was also proved. The bcc Ta cell parameter increases with atomic percent of oxygen in the sputtering environment up to about 20 atomic percent oxygen when Ta_2O_5

begins to form [117]. Collecting the results of the work of several authors, negative bias applied to the substrate under ultra-high vacuum conditions seems to favor formation of the β-Ta phase. One experiment in particular [114] has found that the application of a negative potential to the substrate can also change the relative interstitial content of impurity atoms. This work contends that the application of a negative bias to the substrate during thin film deposition facilitates the formation of a superlattice structure due to interstitial ordering of impurities along the *c* and *b* axes.

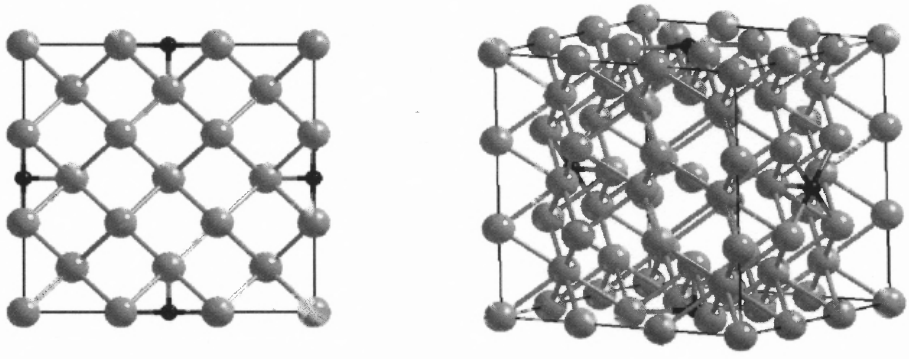


Figure 2.14 (001) (left panel) and (111) (right panel) views of unit cell of distorted bcc superlattice structure proposed (Reference [114]) for β-Ta. The dark spheres of smaller diameter represent nitrogen atoms.

Since it was determined that Ta films deposited under negative substrate bias contained up to five times more nitrogen than films deposited without substrate bias, the formation of the superlattice structure was attributed to the interstitial ordering of nitrogen atoms. Quantitative estimates of the nitrogen and oxygen impurities could not be made however. The interstitial ordering of nitrogen was believed to elongate the *a* and *b* axes to approximately three times their bcc value [114]. Interstitial ordering along the *c* axis in thin films was ruled out because of the contraction along the *c* axis produced by the

biaxial stresses in thin films. In contrast with thin films, ordering of interstitials in the bulk material takes place along the c axis, leading to its elongation and a c/a ratio greater than 1. As determined by the observed diffraction pattern [114], the superlattice structure was tetragonal, with $a=b=$ 10.29Å and c=9.2Å giving a c/a ratio of 0.894. Because of the poor x-ray scattering power of interstitials like nitrogen referred to earlier, X-ray diffraction analysis of a strong peak gave a lattice spacing of 2.63Å, which was indexed as belonging to the (200) sets of lattice planes. Electron diffraction analysis of the same peak gave a lattice spacing of 2.56Å, which was indexed as (400). The underestimation of the a and b axes in previous studies was attributed to the fact that X-ray techniques miss the inner reflections due to the interstitials. The difference in the measured values of the lattice spacing by the two techniques, X-ray diffraction and electron diffraction, was attributed to the substrate stress.

$P4_2/mnm$ Structure of β-Tantalum

The $P4_2/mnm$ structure [106] (Figure 2.15) is analogous to the structure of β–U. The $P4_2/mnm$ space group has three free atomic positional parameters, which can be optimized by subjecting the atoms to forces and allowing them to assume positions that result in the lowest energy configuration for the unit cell. The force minimization calculations are carried out using the full potential linearized augmented plane wave (FP-LAPW) method in the density functional theory (DFT) formalism. The LAPW method is described in Chapter 4 and the force minimization calculations are described in Chapters 4 and 7.

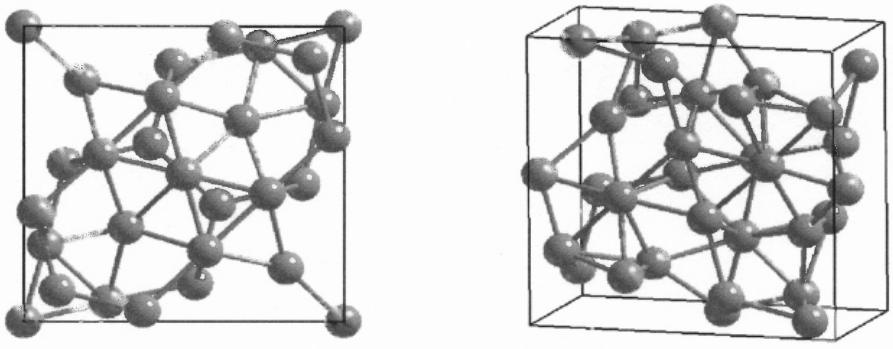


Figure 2.15 (001) (left panel) and (111) (right panel) views of unit cell of $P4_2/mnm$ structure proposed (Reference [106]) for β-Ta.

For a long time it has been known that uranium (U) has three allotropic forms. Up to 935K, the orthorhombic α form (space group No. 63 *Cmcm*) with four atoms per unit cell exists. From 935 to 1045K, it was known that the β form of U (β-U) with 30 atoms per unit cell existed, but until 1988 [119] the exact space group of this β form was not known. At 1045K, U transforms to the simple bcc γ form with two atoms per unit cell. U melts at 1405K and is known to undergo grain growth at high temperatures [119], leading to anisotropic thermal expansion in the γ phase, which gives rise to significant intergranular stresses. The grain growth also results in preferred orientation, readily observed in supposedly random X-ray diffraction powder patterns of β-U at high temperatures. X-ray diffraction patterns of β-Ta also show a (002) preferred orientation [120], suggesting that grain growth readily occurs in samples of β-Ta. The general form of the intensities of β-Ta powder lines recorded with a microdensitometer are similar to those reported for β-U. The c/a ratios and calculated densities of β-Ta and β-U are very close. Unit cell parameters for β-Ta of a = 10.194Å and c = 5.313Å were subsequently determined by least-squares refinement [106]. The uncertainty [110, 121 - 123]

surrounding the structure of β-U was not resolved until 1988 [119]. Using Rietveld profile refinements [124 - 125] of time–of–flight (TOF) neutron diffraction data, it has been determined that β-U crystallizes in the centrosymmetric $P4_2/mnm$ space group with 30 atoms per unit cell. The $P4_2/mnm$ space group has a sixteen-fold general group of equivalent atoms per unit cell. There are five groups of equivalent atoms in structures belonging to the $P4_2/mnm$ space group. In this space group, the atomic positions are distributed as follows, two atoms in position 2(b) at unit cell coordinates of (0,0,0.5) and (0.5,0.5,0), four atoms in position 4(f) consisting of the coordinates $(x,x,0)(0.5+x,0.5-x,0.5)+$ center of inversion, eight atoms in position 8(j) consisting of the coordinates $(x,x,z)(\bar{x},\bar{x},\bar{z})(0.5+x,0.5-x,0.5+z)(0.5-x,0.5+x,0.5+z)+$ center of inversion, and sixteen atoms in two sets of position 8(i) consisting of the coordinates $(x,y,0)(y,x,0)(0.5+x,0.5-y,0.5)(0.5+y,0.5-x,0.5)+$ center of inversion.

P-42_1m Structure of β-Tantalum

Recently single crystals of β-Ta have been prepared by electrodeposition from a molten fluoride bath and subjected to diffraction analysis [126]. Using 2054 reflections, tetragonal unit cell a parameters of a = 10.211Å and c = 5.3064Å were obtained. The observation of weak $(00l)$ reflections for odd l indicated that the correct space group for β-Ta was the P-42_1m space group, and not the previously reported [106] $P4_2/mnm$ space group. This structure of β-Ta is shown in Figure 2.16.

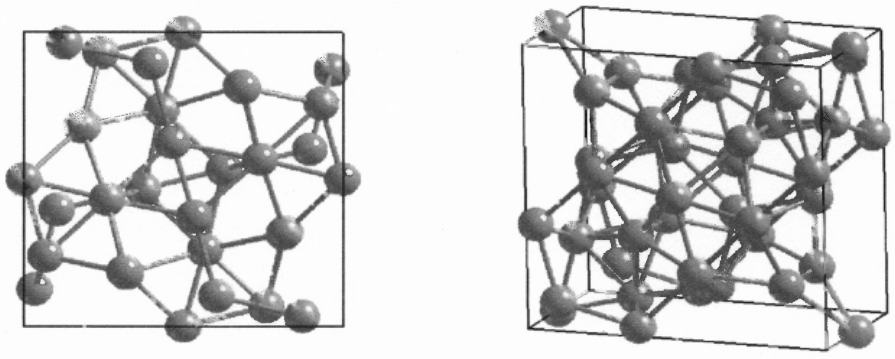


Figure 2.16 (001) and (111) views of the proposed (Reference [126]) P-42_1m structure of β-Ta.

The other crystalline structures proposed for β-Ta are shown in Figure 2.17, and the various operating conditions that have been used to produce tantalum thin films are listed in Table 2.1, which has been adapted from Reference [115].

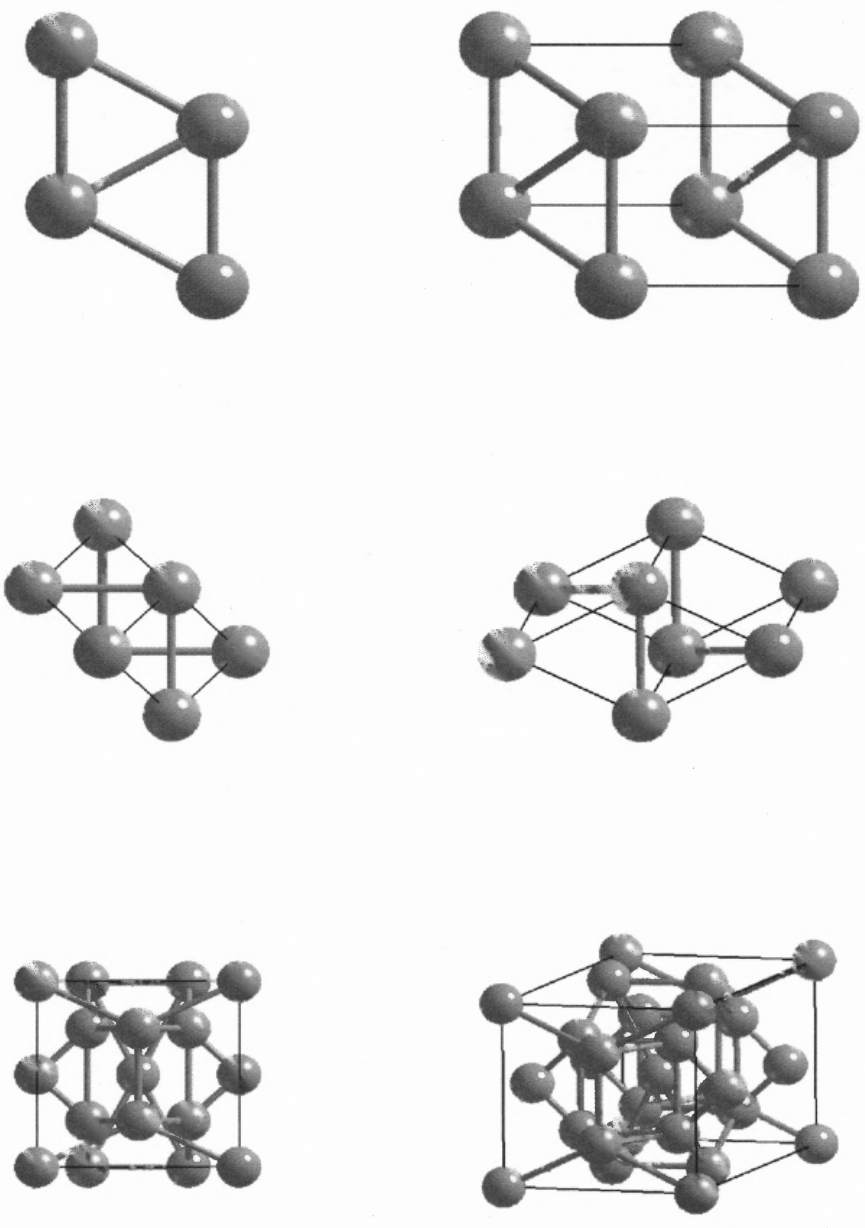


Figure 2.17 (001) (left panel) and (111) (right panel) views of (top) body centered tetragonal (bct) structure of β-Ta, (middle) hexagonal structure of β-Ta (Reference [116]) and (bottom) A15 structure of β-Ta (Reference [93]).

Table 2.1 Some Preparation Details of Sputtered Tantalum Thin Films (see Appendix for complete table)

Reference	Deposition method	Deposition pressure (torr)	Deposition rate (Å/min)	Substrate	Substrate temp. (°C)	Film Structure	Resistivity (μΩ cm)	Details	Reactive sputtering
Gerstenberg and Mayer, 1962	5 k V 0.25 mA/cm^2	1.5 x 10^{-2}	120	Glass		bcc	50		N_2
Altman, 1962	3-6 k V 0.22-0.87mA/cm^2	7 x 10^{-3}-7 x 10^{-2}	77-430	Pyrex	400	bcc	33	ρα ~ 0.05 for films 180-4,540 Å thick	
Gerstenberg and Calbick, 1964	5 k V 0.25 mA/cm^2	2 x 10^{-2}	120	Soft glass	400	bcc	50		O_2, N, CH_4
Read and Altman, 1965		2 x 10^{-2}				β	180-220	X-ray diffraction analysis Films 100-20,000 Å thick	
Krikorian and Sneed, 1966	1-6 k V 0.098-0.65 mA/cm^2	5 x 10^{-3} 1.7 x 10^{-1}	50-300	7059*	400, 450	bcc	30-8 x 10^3		
Westwood and Waterhouse, 1971	3.5 k V 0.3 mA/cm^2	2 x 10^{-2}		7059	~ 250	β + bcc	~ 250	Proportion of β-Ta increased with flow rate of O_2-Ar mixture	O_2
Calbick, 1966	1.5 k V	7 x 10^{-2}-10^{-1}		Carbon coated 7059		Possibly β		Electron microscopy	
Sosniak, 1968	Diffusion pump and turbomolecular pump system	1-2 x 10^{-2}	120	7059		β + Traces bcc	140-180	Background gases monitored during sputtering	
Cook, 1967	4-6 k V 0.3-0.5 mA/cm^2	~ 10^{-2}	~ 200	7059		β	200-700		
Vratny, Vromen and Harendza-Harinxma, 1967	4-5 k V In line and batch systems	2-3 x 10^{-2}		7059		β	180-190		
Sosniak, Polito and Rozgonyi, 1967	Oil diff. pump Sputter Ion and Ti Sub. Sputter Ion and Ti Sub.	2 x 10^{-2} 2-8 x 10^{-2} 2-8 x 10^{-2}	120 35-80 35-80	7059 7059 7059	20-400 20-550 340-380	β+bcc β, bcc β	28-187 38-337 112-207	Up to 15% b.c.c. Phases separate or mixed	

*7059 – 7059 Corning glass

A tetragonal phase was reportedly formed from a bcc Ta film when the film was cooled quickly from high temperatures of 1200 - 1600°C [86]. Some authors reported the same tetragonal structure [127], while others [128 - 129] attributed it to an oxide phase, $Ta_{32}O_9$. From such conflicting results, it was unclear whether the residual gases or other organic contaminants could have been incorporated into the films, highlighting the problems involved with attempting accurate structure analysis of a metal as reactive as Ta. The situation is complicated further because Ta is capable of forming a large number of oxides, referred to as sub-oxides [130], as well as oxy-nitrides and carbides [131 - 132]. This raises the possibility that there is a whole range of thin film metastable Ta states, which can be formed depending on thin film deposition conditions and the nature of the substrate used to support the film. These influences on thin film structure are classified broadly into thermodynamic and kinetic factors, which are encountered in theories of thin film growth and nucleation processes. The study of metastable phases requires an appreciation of these theories, some of which are covered in Chapter 3.

CHAPTER 3

THIN FILM GROWTH AND METASTABLE PHASE FORMATION

Thin film properties depend critically on nucleation and growth processes, which in turn depend on the method used to produce the thin films and determine thin film structure and phase. Several theories have been developed to explain the mechanisms of film nucleation and growth in order to understand the processes leading to the formation of different observed thin film structures. Among the theories that have been developed to explain the nucleation of films are mean field theory [133 – 135], while the theories of film growth include the thermodynamically based models of Vollmer and Weber, van der Merwe, and Stranski and Krastanov [78]. The description of nucleation mechanisms in sections 3.1 and 3.2 is taken largely from References [66] and [67].

3.1 Nucleation Theory

Nucleation pathways are very important in determining the characteristics and structure of thin films. The crystalline size and orientation of films, which determine magnetic, resistive and dielectric properties, are influenced by nucleation mechanisms. The first attempts to correlate nucleation mechanisms with observed magnetic properties were made in the 1950s. Several models of nucleation have since been developed [136 - 138]. In one model, an atom arriving from the vapor phase is assumed to undergo three competing processes, capture by absorption at an already existing nucleus, collision with another atom to form a critical cluster, and re-evaporation into the vapor phase due to the

expiration of its lifetime on the substrate. The lifetime τ of an atom on the substrate is related to the adsorption energy E_a and vibrational frequency ν_0 by [135]

$$\tau = \frac{1}{\nu_0} \exp\left(\frac{E_a}{k_B T}\right), \tag{3.1.1}$$

with the equilibrium population N_1 of single atoms (adatoms) having lifetime τ being [66]

$$N_1 = F\tau, \tag{3.1.2}$$

where F is the deposition rate. For Cu on glass, E_d = 0.14 eV, ν_0 =10^{14} Hz and N_1 = 2 x 10^5 atoms/cm^2 [67]. The vibrational frequency ν_1 associated with the migration of atoms across a substrate surface, with surface diffusion energy of E_d and an effective speed of v sites per second is [135]

$$\nu_1 = v \exp\left(\frac{E_d}{k_B T}\right). \tag{3.1.3}$$

The adatoms move along the surface by hopping from valley to valley in the surface potential with a frequency given by ν_1. During its lifetime, the number of sites n visited by any migrating atom is [66]

$$n = v\tau = \exp\left[\frac{(E_a - E_d)}{k_B T}\right]. \tag{3.1.4}$$

The mean square displacement $\langle l^2 \rangle$ achieved by an atom hopping randomly on the substrate surface during its lifetime is given by [67]

$$\langle l^2 \rangle = Na^2 = v\tau \exp\left[\frac{(E_a - E_d)}{k_B T}\right] a^2, \tag{3.1.5}$$

where a is the jump distance and N is the number of jumps.

If the number of adsorption sites per unit area is N_0, the area occupied by n sites is n/N_0. If any particular area contains a capture site and a migrating atom is incident on the area, it will be captured and the area of capture is then described as the capture zone or catchment area [66]. All newly arriving atoms in the capture zone will be captured prior to the expiration of their lifetime. If the deposition rate is measured in atoms per unit area per second, the rate of capture in the capture zone will be equal to Fn/N_0. Denoting the capture rate for single atoms as Ω_{11},

$$\Omega_{11} = Fn/N_0 = (F/N_0)\exp\left[\frac{(E_a - E_d)}{k_B T}\right]. \tag{3.1.6}$$

An adatom may encounter other adatoms during its motion on the substrate surface, with which it may react to form a dimer cluster. If the cluster is of subcritical size and reacts with an adatom, the adatom may be irreversibly incorporated into the cluster. A stable nucleus is produced from a critical nucleus, which can consist of a single atom, by the addition of one more atom. If the number of atoms in the critical nucleus is i^*, the density of critical nuclei is N_{i^*}. The rate of formation of stable nuclei J_{i^*}, which is also the nucleation rate, is given by

$$J_{i*} = \Omega_{11*} N_{i*} = nFN_{i*}/N_0 . \tag{3.1.7}$$

The density of critical nuclei can be represented in terms of the binding energy E_{i*} of the critical nucleus as

$$N_{i*} = i* N_0 (N_1/N_0) \exp(E_{i*}/k_B T), \tag{3.1.8}$$

resulting in a nucleation rate of

$$J_{i*} = F(F/N_0 \nu_0)^{i^*} \exp\left[\frac{(E_{i*} + (i^* + 1)E_a - E_d)}{k_B T}\right]. \tag{3.1.9}$$

The size of a critical nucleus can be obtained from this equation by differentiation. This model was used to show that many elements incident on alkali halide substrates formed critical clusters of about 2Å in size [66].

The total number of sites visited by single atoms with equilibrium population of $N_1 = F\tau$ is $F\tau n$, where n is the average total number of sites visited by the atoms during their lifetime. If the total number of sites is less than the number of adsorption sites per unit area N_0, then the capture zones for the migrating atoms can be assumed not to overlap, even though a small probability of mutual coalescence still exists. Mutual coalescence or nucleation becomes possible when the number of sites exceeds N_0. Initially incomplete condensation is most likely to occur when the total number of sites $F\tau n$ is more than N_0. Nuclei are then generated by mutual coalescence processes, shifting the equilibrium in favor of single atoms separate from the capture zones of already stable nuclei. Further nucleation inside the capture zones of the stable nuclei

becomes highly improbable and as nucleation continues mutual coalescence processes approach saturation. At this point the substrate surface is covered with capture zones of established stable nuclei. The saturation density of nuclei is given by [66, 137]

$$N_s = N_0/n = N_0 \exp\left[-\frac{(E_a - E_d)}{k_B T}\right]. \tag{3.1.10}$$

Conditions of low deposition rate or high substrate temperatures, favoring a low concentration of single atoms, also favor incomplete initial condensation. The condition for initially complete condensation normally encountered is that $F\tau n > 2N_0$. This means that on average two atoms will be contained in each capture zone n/N_0. The presence of two atoms in the capture zone favors the immediate formation of stable pairs or dimers having an equilibrium density given by

$$N_s = N_1^{\,2} = FN_0\tau/n = FN_0/\nu_0 \exp(E_d/k_B T). \tag{3.1.11}$$

In a thin film deposition process, parameters such as the deposition rate F and the temperature T can be used as independent variables to control the equilibrium number of adatoms by adjusting these variables to produce a transition from initially incomplete to initially complete condensation. The transition temperature T_0 can be obtained from

$$2E_a - E_d = k_B T_0 N_0 \ln(N_0\nu_0/F), \tag{3.1.12}$$

with the transition relation being given by

$$F\tau n = N_0. \tag{3.1.13}$$

An alternative approach [67] to describing the nucleation of islands represents the island density in terms of the cluster radius r and the Gibbs free energy $\Delta G(r)$. Atoms from the vapor with source temperature T_s strike the substrate with an average kinetic energy of $3k_BT_s/2$, most of which is dissipated by the substrate within two periods of its lattice vibrations. Two lattice vibration periods corresponds to about 10^{14}Hz for most substrates. Adsorption of the incident atoms by the substrate occurs when they are trapped in the surface potential. The probability of an atom on the substrate acquiring enough energy to overcome the adsorption potential is $\exp(-E_{des}/k_BT)$, where E_{des} is the desorption energy, therefore per unit time, the fraction of atoms going back to the vapor phase is given by

$$n_d/n_a = \nu_0 \exp(-E_d/k_BT). \tag{3.1.14}$$

As long as a critical radius r^* is not exceeded, the equilibrium population of clusters with radius r is given by

$$N(r) = n_a \exp(\Delta G(r)/k_BT), \tag{3.1.15}$$

where $\Delta G(r)$ is the Gibbs free energy of formation of a cluster having radius r and is equal to $G_{solid} - G_{vapor}$. Atoms are assumed to be incorporated irreversibly into clusters with radius less than or equal to the critical radius. The formation of a cluster causes $\Delta G(r)$ to increase to the value associated with formation of a critical cluster size. When a cluster has grown to the critical size, further incorporated atoms reduce the Gibbs free

energy. A critical cluster is composed of a nucleus and an adatom. The nuclei formation rate I^* is described by the proportionality

$$I^* \propto \exp[(2E_{des} - E_d - \Delta G^*)/k_B T]. \tag{3.1.16}$$

Experimental work has shown [67] that the smallest critical nuclei must contain at least one atom and can contain a few atoms. The atomistic approach [67] was developed to better describe the formation of such nuclei. In the atomistic approach, the energy of decomposition E_i of an aggregate into its i single atoms is used to replace the Gibbs free energy of formation. The decomposition process

$$i \text{ adatoms} \xleftrightarrow{E_i} \text{cluster } i, \tag{3.1.17}$$

results in an island population having clusters i in a density n_i/N_0, where N_0 is the number of adsorption sites per unit area

$$n_i/N_0 = (n_a/N_0)^i \exp(-E_i/k_B T). \tag{3.1.18}$$

The rate of formation of clusters containing $i+1$ atoms from clusters containing i atoms is given by the following proportionalities, valid for $I_{i+1} < I_i$,

$$I_{i+1} \propto n_i n_a \nu_0 \exp(-E_d/k_B T), \tag{3.1.19}$$

and

$$I_{i+1} \propto \frac{F^{i+1}}{\nu_0{}^i N_0{}^{i-1}} \exp\{[(i+1)E_{des} - E_d - E_i]/k_B T\}. \tag{3.1.20}$$

The temperature dependence of the cluster formation rate from atomistic nucleation theory is shown in Figure 3.1. The critical atomic number is defined by the lowest nucleation rate. The discontinuous slope of the nucleation rate from Figure 3.1 demonstrates that nuclei are made up of discrete atoms.

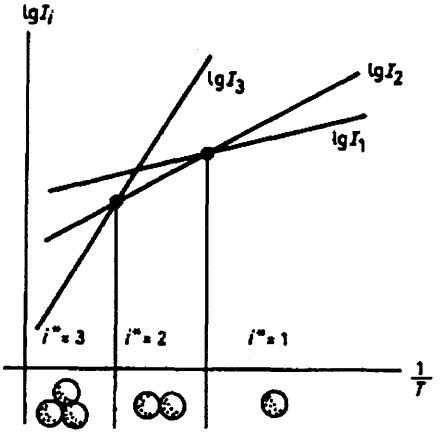


Figure 3.1 Temperature dependence of the cluster formation rate from atomistic nucleation theory showing different critical cluster size regions. I and $\lg I_i$ represent respectively the cluster formation rate and the log of the cluster formation rate (taken from Reference [67]).

The decay of supercritical clusters is usually assumed to be negligible compared to cluster growth. The maximum distance from the nucleus that an adatom can be located, and still reach the nucleus during its lifetime on the substrate, defines the edge of the nucleus capture zone. The mean square distance covered by an adatom in N steps is Na^2, and the area associated with this distance can be approximated by a circle for ease of analysis. The capture zone radius is then

$$r^2 = Na^2/\pi = 2D_C\tau/\pi \approx D_C\tau\,, \tag{3.1.21}$$

and

$$r^2 \approx \exp[(E_{des} - E_d)/k_B T] \cdot a^2, \tag{3.1.22}$$

with

$$\langle l^2 \rangle = Na^2 = 2D_C \tau, \tag{3.1.23}$$

where D_C is the diffusion coefficient. The time evolution of the number of nuclei per unit area, $N(t)$, is described by

$$\dot{N}(t) = dN/dt = I[1 - N(t)D_C \tau], \tag{3.1.24}$$

or

$$N(t) = 1/D_C \tau [1 - \exp(-ID_C \pi)], \tag{3.1.25}$$

which is valid for small times t. Therefore the time constant for the complete coverage of the substrate surface with capture zones is $1/ID_C \tau$. Nucleation ceases for times greater than this time constant.

During the nucleation process, the rate of mass increase per unit area $\dot{m}(t)$ is mostly due to atoms incident on capture zones. $\dot{m}(t)$ is given by

$$\dot{m}(t) = N(t)F\pi D_C M / N_A = FM[1 - \exp(-ID_C \pi)]/N_A, \tag{3.1.26}$$

where N_A and M are respectively, Avogadro's number and the atomic mass. $m(t)$, the mass increase as a function of time, is then given by

$$m(t) = \dot{m}_\infty \left\{ t - \left(\frac{1 - \exp(-ID_C \pi t)}{ID_C \tau} \right) \right\}, \tag{3.1.27}$$

where

$$\dot{m}_\infty = FM / N_A \,. \tag{3.1.28}$$

The onset of atomic condensation is dependent on the substrate temperature, which determines the duration of the nucleation period. The atomic mass per unit area as a function of time is shown in Figure 3.2.

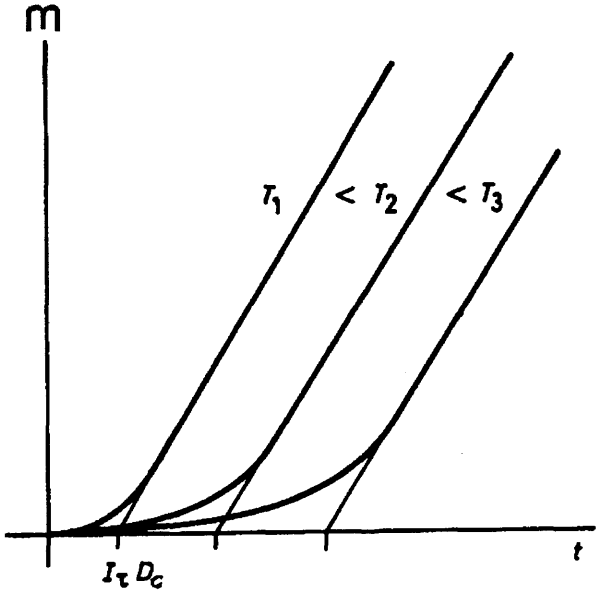


Figure 3.2 Mass per unit area as a function of time. Substrate temperature dependent onset of condensation is also shown (taken from Reference [67]).

The time lag before condensation corresponds to the nucleation period. The deposition rate F must be such that adatoms and clusters are not permitted to reach equilibrium with each other, in order for complete condensation to occur. In mathematical terms,

$$F \geq 1/a^2 \tau . \tag{3.1.29}$$

Inhibition of crystalline layer formation occurs when overgrowing layers are deposited before atomic jumps to equilibrium positions are possible. Mathematically,

$$F \geq (1/a^2)\nu_0 \exp(-E_d/k_B T). \tag{3.1.30}$$

No further nucleation occurs after the complete coverage of the substrate surface with capture zones. Islands then grow by capturing adatoms diffusing across the substrate surface, or by direct capture of vapor atoms.

It is also possible for islands to grow together. In the zone between the islands, strong capillary forces are present which result in quick coalescence. Electro microscopic studies have revealed that coalescing islands are similar in behavior to liquid droplets, even if the islands previously contained crystallographic features. The crystallographic features are observed to disappear during island coalescence and a grain boundary appears. This grain boundary is not permitted to move if the movement increases the surface area of the coalescing islands. After surface energy minimization forces have shaped the coalesced islands, the grain boundary separating them from former islands begins to move. Sections having energetically less favorable orientations are enveloped until many islands have coalesced, resulting in an extensive coverage of the substrate surface. The mazy island coverage of the substrate then evolves into a continuous film upon further capture of single atoms.

3.2 Mean Field Nucleation Theory

Mean field nucleation theory postulates that the mean free path of adatoms diffusing on a substrate equals the mean distance between islands at saturation conditions. It neglects the spatial dependence of island densities. Mean-field nucleation theory has for a long time been successfully used to extract microscopic parameters such as diffusion barriers and adsorbate-adsorbate binding energies from island density data in growth experiments. Saturation is defined as the point after which island growth becomes dominant over nucleation of new islands. The scaling laws of nucleation then apply in the saturation region i.e. in regions of higher coverage. A rate law treatment of the change in monomer and island densities in the form of differential equations describing the rate of change of the mean monomer and island densities below saturation coverage of the substrate are the main result of mean field theory. The description of the variation of monomer and island densities with time by these rate laws is consistent with the description by nucleation theory. As expected, this description is valid only up to saturation conditions. The rate law treatment does not treat coalescence of islands very well. It predicts coalescence too early and the island density does not go to zero for complete coverage as it should. A much better description of island coalescence is obtained with kinetic Monte Carlo simulations.

Scaling laws can also be used to describe island size distribution near saturation conditions. The temperature dependence of the island size distribution is used to obtain values for the energy barrier to atomic surface migration E_d, and the hop attempt frequency ν_0 for surface movement of single adatoms. Scanning Tunneling Microscopy (STM) measurements reveal the island size distribution function to be of the form

$$N_S = (\theta / \langle s \rangle^2) g(s/\langle s \rangle), \tag{3.2.1}$$

where N_S is the saturation number density of islands, θ is the coverage and is equal to Ft, where F is the deposition rate and t is the time of deposition, s is the island size, $\langle s \rangle$ is the average island size and $g(x)$ is a scaling function.

A good test of nucleation models is to measure the saturation island density as a function of temperature, whereby regions of initially incomplete and complete condensation can be distinguished. The theoretical variation of the saturation island density with inverse temperature for the initially incomplete and initially complete condensation regions is shown in Figure 3.3.

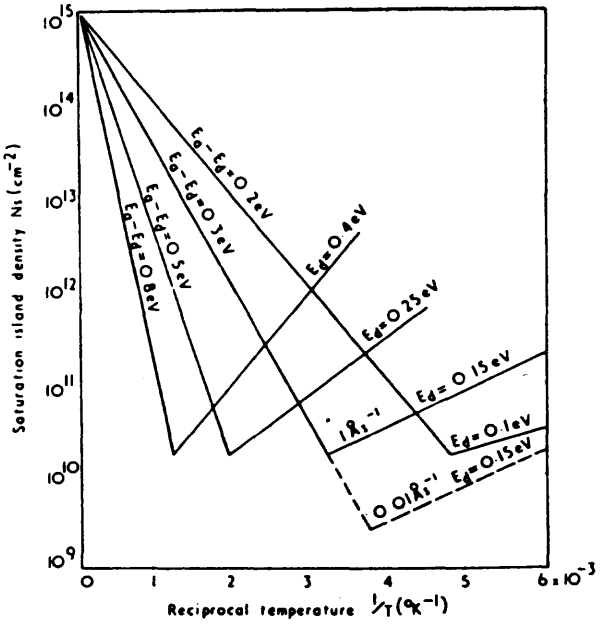


Figure 3.3 Theoretical dependence of the saturation island density N_S with temperature, adsorption energy E_a and diffusion energy E_d for an incidence rate of 1Å per second (taken from Reference [66]).

3.3 Thermodynamic Film Growth Models

During the early years of metal epitaxy, vacuum conditions and surface cleaning methods were inadequate for the reliable theoretical analysis of experimental data. Many models were proposed to explain epitaxy until they were unified in 1958 and classified on the basis of general thermodynamic criteria [66]. Epitaxial growth of a metal film on a metal substrate is therefore often described according to three standard models [78]; two-dimensional or Frank-Van der Merwe (FM) growth, three-dimensional or Volmer-Weber (VW) growth and two-dimensional growth followed by three-dimensional growth or Stranski-Krastanov (SK) growth. If quasi-equilibrium growth conditions are assumed, these three growth modes are influenced by the surface free energies, the interfacial free energy and the strain energy. They are therefore based entirely on thermodynamic considerations.

In practice quasi-equilibrium considerations are used because condensing atoms do not have much surface mobility and the equilibrium state cannot be reached because the activation energy of restructuring an already formed aggregate of atoms is too high. If σ_f is the surface free energy of the deposited film, σ_i is the interfacial energy, σ_s is the substrate free energy and σ_e is the strain energy, then the deposited film will grow in two dimensions (FM growth) if [78]

$$\sigma_f - \sigma_s + \sigma_i + \sigma_e < 0. \tag{3.3.1}$$

This condition must be fulfilled for every layer of the film ν for FM growth to occur. This can happen when the film and substrate have very similar crystallinity. In general, the SK mode of growth (3D crystals on top of a few monolayers) occurs because the

condition is only satisfied for a few monolayers. In other words SK growth is generally seen because the inequality represented by Equation (3.3.1) usually reverses with film thickness. The film grows as three-dimensional (VW growth) clusters otherwise. FM growth is desired when the goal is to produce two dimensional film structures with a certain crystallographic phase and orientation. The three idealized growth models neglect the possibility that substrate atoms can move to the film surface during film growth (substrate surface segregation). It has usually been difficult to predict the particular growth mode for two reasons; one reason is that it has not been possible to determine the interfacial energy σ_i with sufficient accuracy. This has been true even for interfaces between bulk crystalline substances. In the past, simple models such as the primitive cubic lattice with first-nearest neighbor interactions have been used to estimate σ_i [78]. The other reason is that the electronic structure of a monolayer is modified by interaction with the substrate so that values of σ_f obtained from the bulk material are not always good approximations.

Three processes, all directly proportional to the substrate temperature, determine if equilibrium growth can be expected. The three processes are surface adatom diffusion, substrate surface segregation and substrate/film interdiffusion. The surface diffusion process varies with crystal face, surface smoothness and element. It is assumed to follow an Arrhenius law, $D_0 \exp(-E_d / k_B T)$ where E_d is the activation energy for surface diffusion. Typical values of the pre-exponential D_0 are 10^{-3} cm^2 s^{-1} [78]. Measured surface diffusion barriers (activation energies) range from 0.1 - 0.9eV representing a temperature range of 40 - 350K for an adatom mobility of one hop per second. Activation energies can usually be determined experimentally. Activation energies for self diffusion

ranging from 0.28 to 0.48 eV have been determined for Cu(100) [78] . Diffusion barriers for many metal films can be estimated by scaling the activation energy of a known material to an unknown material using the cohesive energies of the two different metals. Substrate surface segregation occurs when the substrate surface free energy is lower than the surface free energy of the deposited film. Heats of solution and the elastic size mismatch energy play a role in addition to surface free energy differences, making it difficult to estimate activation energies associated with this process. Often it is desired to produce two-dimensional film structures with a certain orientation and phase. FM or two-dimensional growth is difficult to obtain when the surface free energy of the film is greater than that of the substrate. Methods have been developed to improve the morphology of deposited films in systems where the film free energy is greater than that of the substrate, one such method [78] consists of depositing the film at a temperature where thermal diffusion is minimal, then annealing the film to a temperature high enough to allow ordering, but not high enough to permit film substrate segregation. There is an upper temperature limit for two-dimensional (FM) growth to occur, above which substrate-film interdiffusion processes become important. This limit is imposed by the requirement of an abrupt film-substrate interface and has been estimated as 0.5 of the melting temperature [78]. The critical parameters in thermodynamic models of film growth are therefore the surface free energy differences between the film and substrate and the lattice mismatch between the film and substrate.

Thermodynamic models of film growth assume that film deposition is a quasi-equilibrium process, meaning that the film is in equilibrium with itself, but not with its vapor, and neglect the surface mobility of atoms and atomic diffusion between layers

except at very high temperatures. Under conditions of high deposition rates and low substrate temperatures, corresponding to high values of the ratio F/D, where D is the atomic diffusion rate on the substrate surface, supersaturation occurs, leading to strongly non-equilibrium conditions. Thermodynamic film growth models are not able to describe film growth under such conditions, and kinetic factors must be considered.

Film growth modes that result in a well-defined relationship between the structure of a film and the substrate supporting it are known as epitaxial growth modes. Epitaxial growth is also used to refer to the growth of a single crystalline layer on a single crystal surface. The most popular epitaxial method is Molecular Beam Epitaxy (MBE). In MBE the desired element, which is in the form of a molecular beam from a thermally evaporated source, is deposited onto a heated crystalline substrate to form a thin epitaxial layer. MBE can produce high-quality film layers having very abrupt interfaces and well-controlled thickness. A technique known as Chemical Beam Epitaxy (CBE), which is a combination of MBE and MOCVD (metal organic chemical vapor deposition), has also been used extensively in this regard.

Homoepitaxy and heteroepitaxy are respectively, the growth of a single crystalline film on a single crystalline substrate of the same material and growth of a single crystalline film on a single crystalline substrate of a different material. No uniform theory exists that describes these phenomena. The important features associated with heteroepitaxy are the geometrical compatibility between the substrate surface structure and a lattice plane of the film material, although some epitaxial systems containing large positive and negative misfits up to 10% have been produced [66]. A dislocation network

located near the film substrate interface or elastic stress in the deposited film usually results from such structural misfits between the film and substrate.

The substrate temperature is another important parameter in epitaxy. Usually for epitaxial systems, there is a temperature T_e above which good epitaxy is obtained [67]. This temperature depends on surface contamination and the deposition rate. T_e is higher for ultra pure substrates than for high-vacuum cleaved surfaces. Higher temperatures in general usually reduce the concentration of contaminants by desorbing them from the substrate surface. The surface mobility of adatoms is also enhanced at high temperatures, making it more possible for the atoms to reach energetically favorable sites. Recrystallization and defect annihilation is also favored at high temperatures because of the enhancement of diffusivity in the deposited film. Nuclei are also generally larger at higher temperatures. The nucleus must contain at least 3 atoms for substrates having threefold symmetry and 4 atoms for substrates having fourfold symmetry in order for a defined orientation relationship between the nucleus and the substrate to exist. When the nuclei take up the same orientation as the substrate, misfits are very unlikely to accumulate during film coalescence.

Adsorbed impurities that are epitaxially bound may inhibit epitaxy in some cases, but when they accommodate misfits, they can favor epitaxial growth. Impurities that form an amorphous layer are generally not desirable during epitaxial film growth. Electron bombardment before and during the film deposition process can stimulate the desorption of impurities and thereby clean the substrate surface. Such bombardment can also increase the number of nucleation sites, which favors epitaxial growth. Epitaxial growth

carried out under ultra-high vacuum conditions is an important technique in creating new phases of materials.

3.4 Kinetic Modeling of Film Growth Processes

Kinetic descriptions of film growth result directly from the previously described theories of nucleation. The main kinetic processes affecting film growth are adsorption processes and atomic diffusion processes. Atomic diffusion processes are of two kinds, atomic surface diffusion and atomic interlayer diffusion [78, 135]. Each of these kinetic processes has associated with it an activation energy. Surface diffusion is assumed to obey an Arrhenius relationship of the form [135, 139],

$$D = D_0 \exp(-E_d / k_B T) . \tag{3.4.1}$$

The factors affecting the relative importance of thermodynamically driven film growth processes to kinetically driven processes are the rate of atomic deposition, the substrate temperature and the nature of the film and substrate.

Conditions of high deposition rates and low substrate temperatures lead to supersaturation and strongly non-equilibrium conditions. With a lot of metal systems, supersaturation is attained below room temperature and deposition rates greater than 0.1 monolayers per second. Ordering processes due to the mobility of deposited atoms counterbalance the roughness of the surface, which is produced by the stochastic nature of the deposition process. At high deposition rates and low substrate temperatures, the ordering processes are greatly diminished, resulting in roughness of the film surface.

Kinetic processes therefore become very important in determining the final morphology of the deposited film under these conditions.

Techniques such as homoepitaxy, which consist of depositing a metallic film on a substrate made of the same metal, are used to study film growth under supersaturation conditions and have been used to gain insight into the atomic scale kinetic processes affecting film growth. The reasoning behind the use of homoepitaxy is that the relatively large contribution to the thermodynamic energy of the system (the interfacial energy σ_e) due to the free energy difference between the film and substrate when they consist of different materials is eliminated. The study of kinetic processes is important because film interlayer diffusion does occur and does have a significant effect on film morphology. A much studied system in this regard is Pt/Pt(111) [141]. The usual source of Pt atoms for these studies is electron bombardment. In the Pt/Pt(111) system, for 100K < T < 340K, the observed growth of platinum atoms is quasi two-dimensional. For 340K < T < 450K, the observed growth is three-dimensional. For T > 450K, the observed film growth is two-dimensional. The switch from quasi two-dimensional growth to three-dimensional and then back to two-dimensional is termed re-entrant growth. The re-entrant growth observed for Pt on Pt(111) has not been adequately explained. A frequently studied heteroepitaxial system is Au/Ag (110) [141 - 142]. This system satisfies the thermodynamic criteria for two-dimensional (FM) film growth. However, interlayer diffusion and clustering indicative of three-dimensional growth are still observed, attesting to significant kinetic influence on the film growth in this system.

A much-used aid to understanding the kinetics of film growth is molecular dynamics simulation [139 - 141, 143]. Molecular dynamics calculations have been used

to model the effect of an incident Au atom on interlayer diffusion on an Ag (110) surface [141]. Three features of the substrate surface were studied; the edge of an ascending step, the edge of a descending step, and a terrace. The authors found that the effect of the incident atom depends on its point of impact on the substrate surface and that the probability that a surface atom will be caused to diffuse into the film is highest when the point of impact of the incident atom is at the edge of the ascending step feature of the substrate surface. The probabilities were determined by giving the incident atom a certain energy and angle of incidence and then following the atomic trajectory for a fixed period of time. In the sputtering process, the kinetic energy of an atom impinging on an adatom (adsorbed atom) helps to overcome the energy barrier to interlayer diffusion. Interlayer diffusion is therefore very likely to occur in sputtering processes, even when sputtering at low temperatures. Sputter deposition processes are therefore more suited to kinetic modeling because thermodynamic models of film growth neglect interlayer diffusion at low temperatures.

It has been determined from the growth of Pb on Cu (001) at low temperatures that island nucleation occurs early while the coverage is very low [135]. The number density of islands is believed to remain constant, while the average island size increases with coverage. Most atoms arriving on the substrate attach to already-formed islands and do not nucleate new islands. Characterization of heteroepitaxial film growth by theoretical modeling has supported these conclusions about growth modes in such systems. One theoretical model (sometimes referred to as the continuum model) [144 – 145] is based on the proposition that the interface is expected to become rougher as deposition time increases. Continuum equations of motion are related to thermodynamic

quantities and provide a long-term view of growth. Evolution of surface roughness is described in terms of scaling laws. The scaling laws do not contain atomic scale detail. A mean square width and lateral roughness are defined. The kinetic roughening of a surface is described in terms of a dynamic scaling hypothesis for the mean square width σ of the surface height $h(x, y)$. The mean square width σ (or variance) is given by [139]

$$\sigma = \left\langle \left[h(x,y) - \langle h(x,y) \rangle \right]^2 \right\rangle^2, \tag{3.4.2}$$

where $h(x, y)$ is the height function perpendicular to the substrate and the angular brackets denote the spatial average over the surface. The lateral roughness as characterized by a height-height correlation function $g(R)$ given by [146]

$$g(R) = \left\langle \left[z(x,y) - z(x',y') \right]^2 \right\rangle, \tag{3.4.3}$$

is a measure of the jaggedness of short length scale profiles parallel to the substrate surface, where

$$R = \sqrt{(x-x')^2 + (y-y')^2}\,. \tag{3.4.4}$$

The case where the lateral roughness is of the form [146]

$$g(R) \sim R^{2\alpha}, \tag{3.4.5}$$

corresponds to a profile that is the same at all length scales. α is the roughness exponent and determines the rate of change of the surface profile over lengths parallel to the

surface. The roughness of the surface is inversely proportional to the magnitude of α. The roughness $g(R)$ is small only when there are many peaks and valleys in the profile parallel to the surface. The surface profile is allowed to evolve in time corresponding to the continuous deposition of atoms and then the spatial and temporal change of the roughness is investigated. The concept of dynamical scaling is used to describe the evolution of interfaces in film growth processes. The scaling of the interface width $\sigma(L,t)$, where L is the length scale and t is the time, is of the form [147 - 148]

$$\sigma(L,t) \sim L^{\alpha} f(t / L^{\alpha/\beta}). \tag{3.4.6}$$

The function f is defined by

$$f(x) \sim x^{\beta} \text{ for } x << 1, \tag{3.4.7}$$

and

$$f(x) \sim \text{ constant for } x >> 1, \tag{3.4.8}$$

where β is the growth exponent. The interface width $\sigma(L,t)$ is a measure of vertical roughness. When $L^{\alpha/\beta} >> t$, the interface width grows with time according to a power law $\sigma \sim t^{\beta}$, while for $L^{\alpha/\beta} << t$, $\sigma \sim L^{\alpha}$ [147 - 148]. The lateral roughness correlation length ξ is a measure of lateral coarsening and grows with time according to the power law $\xi \sim t^{\beta/\alpha}$ [147 – 148]. The exponents α and β depend on the dimensions and growth mechanism of the substrate-film interface and have been measured [147] for several systems. Representative values are shown in Table 3.1 for some film-substrate systems.

Table 3.1 Interface Width σ, Lateral Correlation Length ξ, and Average Terrace Width η for Several Film-Substrate Systems

System	σ (layers)	ξ (Å)	η (Å)
Fe/Fe(001) (room temp)	~ 1.1	~ 16	~ 14
Cu/Cu(001) (low temp)	~ 1.2	~ 22	~ 15
Pb/Cu(001) (low temp)	~ 1.4	~ 46	~ 33

Source: Gianfranco Vidali and Hong Zeng, "Recent advances in island and multilayer growth of metals on metals far from equilibrium" Applied Surface Science, Vol. 92, 11, (1996).

The growth mode of Pt on Pt(111) has been observed [141] to change from step flow growth at high substrate temperatures to disordered growth at room temperature and to quasi layer-by-layer growth at low temperatures. Re-entrant wetting is the term used to describe the unusual reappearance of ordered growth at low temperatures. The reappearance of ordered growth was explained by realizing that the mobilities of adatoms are reduced at low substrate temperatures, leading to the formation of islands with narrow extensions. The narrow extensions are very unlikely to capture and hold depositing atoms, which subsequently slide off and in this process are somewhat mobile. Interlayer diffusion of atoms has been substantially reduced in an experiment where the diffusion length on the upper terrace was made larger than that on the lower terrace. This was done with the system Ag on Ag(111) [149] which usually produces three dimensional clusters under ordinary deposition conditions. The diffusion length on the lower terrace was reduced relative to that on the upper terrace by increasing the density of nucleation centers. The low density of nucleation centers on the lower terrace was achieved by depositing silver atoms on a low temperature substrate, which resulted in a drastic reduction in their mobility. The deposited silver atoms then acted as nucleation centers

during deposition and reduced the diffusion length on the lower terrace. The same effect has been obtained [150] by pre-sputtering the substrate for a short time to create a low concentration of isolated vacancies and adsorbed atoms, which can be used as nucleation centers.

The Ehrlich-Schwoebel (ES) barrier [151 - 152] is another factor that controls metal film growth. This extra energy barrier encountered by atoms when they step down onto a lower terrace leads to rough growth when it is comparable to the activation energy associated with atomic diffusion on terraces. The growth is affected because the interlayer mass transport is significantly reduced. The ES barrier has been determined by using an analytic model coupled with kinetic Monte Carlo simulations, which take as input STM images at various growth stages. The ES barrier causes the film profile to show a definite length scale.

Diffractive methods such as high resolution low energy electron diffraction and atom beam scattering have been used to probe metal thin film growth on the atomic to nanometer scale [66, 78]. A beam of electrons in the case of LEED and neutral helium atoms in the case of atom beam scattering is focused on the sample and the elastically scattered specular peak is monitored. The structure factor is obtained after deconvolution with the corresponding instrument response functions. The peak profiles obtained are analyzed in this way under varying experimental initial conditions. The helium specularly reflected beam line shape and intensity have been used to investigate the quality of substrate surfaces by obtaining information about the average terrace width of the surface. Helium beam scattering, reflection high-energy electron diffraction (RHEED) and glancing incidence x-ray diffraction have been used to monitor film growth during

deposition. Helium beam scattering is more sensitive than reflection high-energy electron diffraction (RHEED) or glancing incidence x-ray diffraction and analysis is simpler. STM has been used to study growth in far from equilibrium conditions, although the films can only be examined after growth.

3.5 Metastable Phases in Thin Films

Thin films can contain very small crystalline sizes, leading to high surface energies and susceptibility to impurity contamination effects, which in turn lead to the stabilization of a much wider range of structures than are observed in bulk form. For example, metals such as tantalum, tungsten, molybdenum and germanium, which normally exist as bcc structures in bulk form, can grow with an fcc structure in very thin films grown epitaxially as well as by other methods [43]. Often it is not certain which of these structures form due to impurity stabilization effects and which are due to stress stabilization.

The metastable crystal structures are material phases with lattice parameters which result in a local minimum in the total energy per unit cell, and which are separated from the structure representing the absolute energy minimum by an activation energy. These new phases are stabilized and are a special case of depositionally stabilized structures. They are usually formed most easily as thin films and have been observed to contain a significant amount of defects and non-crystallinity. A metastable phase of a metal is by definition a phase that is not stable over the same temperature or pressure range as the stable phase. Metastable phases are usually studied by a wide array of diffraction techniques such as X-ray and electron diffraction and are often difficult to

characterize. As will be shown for tantalum, sometimes there is controversy about the exact crystallographic structure of a metastable phase. Recently, high-resolution instruments such as the Atomic Force Microscope (AFM) [64] and the Scanning Tunneling Microscope (STM) [153] have been used to observe and manipulate thin film crystalline structures at very high levels of resolution and detail. These instruments have been used to describe the morphology of a variety of novel material phases. The morphologies of metastable phases are however very sensitive to preparation details and can therefore vary considerably from sample to sample. For example, samples prepared by techniques such as sputter deposition can contain varying densities of vacancies and foreign interstitial atoms, depending on sputtering variables such as chamber pressure and impurity gas concentration.

The attempted classification of some metastable phases by an examination of their physical morphology quickly becomes an exercise in frustration. In a typical laboratory thin film preparation environment, gaseous impurity levels are almost always considerable and can cause prepared thin film samples to incorporate large amounts of these impurities. Since diffraction based techniques rely on scattering from periodic structures, deviations from periodicity of an experimental sample can make reliable analysis of diffraction data very difficult and in many cases impossible. In such cases it is often useful to assume periodicity and treat the structural problem theoretically with electronic structure calculations, the results of which are compared to experimental data in order to gain insight into the significance of non-periodic effects.

Important metastable systems include metastable β-Ta_2O_5 thin films and ferromagnetic Fe and Co films on Cu substrates. There is an extensive literature on these

metastable phases, which have varying degrees of technological importance. This work is concerned with metastable phases of tantalum and neighboring bcc metals. Metastable phases in elemental form of vanadium [154 - 155], chromium [156], molybdenum [157], tantalum [106, 114, 116], and tungsten [157, 159 - 160] have been reported in the literature. These metastable phases are usually formed most easily as thin films and have been observed to contain a significant amount of defects and non-crystallinity. They also exhibit a number of interesting electronic phenomena, for example, the superconducting transition temperatures in metastable phases of tungsten and tantalum are believed to differ appreciably relative to the bulk bcc values [95, 161]. These metastable phases are discussed in the following sections.

3.5.1 Metastable Phase of Tungsten

Tungsten films are technologically important because they are used as x-ray masks, metallization layers and interconnects on silicon-based devices. Several authors [159 - 160, 162] have reported a metastable form of tungsten with a lattice constant of approximately 0.504 nm, which crystallizes in the A-15 structure and is referred to as β tungsten. The A-15 β-W phase has been prepared by sputter deposition [159, 163 - 164], evaporation [161], and reduction of tungsten oxides [165 - 166]. This metastable phase has been reported to transform to the stable cubic A-2 α-W upon heating, although the reported transition temperature ranges vary widely. The oxygen content of the A-15 β-W phase is believed to influence the transition temperature range, with higher oxygen content films requiring longer transition times and higher heating temperatures. The A-2 α-W structure formed from A-15 β-W in this manner has the same oxygen content and XPS binding energy as its precursor, but much lower resistivity, on the order of 80% less

[160, 162, 167]. The decrease in resistivity has been ascribed to the elimination of the β-W defect structure and an increase in particle size. The A-15 β-W phase has been reported [163] to systematically evolve to the A-2 α-W phase with decreasing oxygen concentration and lattice parameter. The stability of the A-15 β-W was also determined to decrease with decreasing oxygen concentration, implying that oxygen impurities stabilize the A-15 β-W phase.

3.5.2 Metastable Phase of Vanadium

Tian and coworkers [154 - 155] have used the FP LAPW method in the DFT formalism to carry out total energy calculations for vanadium as a function of the c/a and V/V_0 ratios, where V_0 is the equilibrium volume. They predicted a metastable phase of vanadium with c/a and V/V_0 ratios of 1.78 and 1.02 respectively, and then proceeded to prepare metastable phases of vanadium on different substrates, which they identified as strained states of the theoretically predicted metastable phase. In the first of those experiments [154], vanadium was deposited on Ni{001} with an a lattice parameter 0.8% smaller than the theoretically predicted square side (2.51Å) of tetragonal vanadium. The vanadium films could be grown only up to thicknesses of about 20 Å before defects were observed. The total energy per unit cell of vanadium as a function of the V/V_0 and c/a ratios is shown in Figure 3.4.

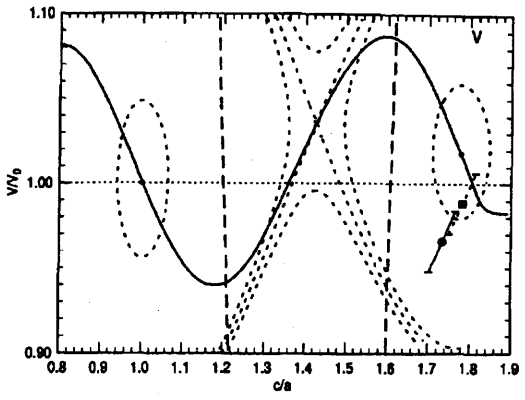


Figure 3.4 Total energy per unit cell of vanadium as a function of the V/V_0 and c/a ratios (taken from Reference [155]).

Since the lattice parameter of Ni{001} is smaller than that of the predicted tetragonal phase of vanadium, the vanadium film was believed by the authors to be compressively stressed. Making use of the Bain path [168] concept, the authors conjectured that using a substrate with a lattice parameter slightly larger (by 2%) than that of tetragonal vanadium would result in more ordered vanadium thin films undergoing tensile stress, and chose Cu{001}(a =2.56 Å) as a substrate. Quantitative Low Energy Electron Diffraction (QLEED) analysis was subsequently used to determine that very thin films of tetragonal vanadium were formed on the Cu{001} substrate, with structures derived from strain on the theoretically predicted metastable tetragonal phase of vanadium. The films had many defects and poor crystallinity and eventually collapsed into {110} bcc vanadium. The high density of disorder and defects in the thin film of tetragonal vanadium was attributed to the small energy barrier between the strained metastable state and the bcc ground state of vanadium. As far as is known, no similar study has been carried out for tantalum.

3.5.3 Metastable Phases of Tantalum

The formation of tantalum metastable phases on a variety of substrates has been reported in a large number of sputtering experiments [81, 91, 93, 106 – 108, 110 – 116, 158]. The metastable phase, referred to as the β phase, is sometimes formed alongside the stable body centered cubic (bcc) phase, referred to as the α phase. The β phase of tantalum occurs mostly as a thin film and is not easily formed as bulk material. It is harder, more brittle, and less ductile than the bcc phase. It is known to have a 45% smaller average grain size than bcc tantalum, and to frequently contain a high concentration of impurities and defects, in addition to having a resistivity (170 - 210 μΩ cm) much higher than that (15 – 60 μΩ cm) of the stable bcc phase [81, 91, 93, 95, 107 – 108, 111, 152]. The β phase of tantalum has been reported to nucleate preferentially on certain kinds of substrates [91, 110 - 113] and may be stabilized by oxygen impurities [91, 108, 110, 113 - 115].

3.6 Thin Film Deposition Methods

Thin film deposition methods can be divided into chemical and physical methods. Chemical methods of thin film deposition include electroplating, electroless plating, chemical vapor deposition (CVD), anodization, hydrophily and thermal growth [67]. CVD, anodization, hydrophily and thermal growth are mostly used for the production of insulating and semiconducting films. Among the chemical methods, electro and electroless plating are much more relevant to metallic films and will therefore be discussed. Physical methods of thin film deposition include evaporation and sputter deposition. The most prevalent thin film deposition method in the microelectronics

industry is sputter deposition. The high ion energies involved in sputtering processes means that modifications of thin film structure and properties occur very frequently in sputtered films. High ion energies also correspond to large values of the F/D ratio, so a high ratio of F/D in a deposition process means that the process is likely to result in the formation of thin films with modified structures. The substrate temperature and the F/D ratio are deposition parameters that can be directly related to nucleation theory and which can be optimized to produce films having desired characteristics. Careful control of these parameters can provide a detailed understanding of the dependence of thin film structural properties on film growth parameters.

Studies of thin film growth phenomena generally require well-controlled vacuum environments in which film growth resulting in a well-defined relationship between the structure of a film and the substrate supporting it can be achieved. Such a well-controlled environment is afforded by vacuum evaporation systems. The basic understanding of film growth mechanisms and the relationships between the structure and physical properties of films are based mainly on data obtained from evaporated films [66]. Several thin film deposition methods, including sputtering and evaporation, are described in this section, mostly following References [66] and [67].

3.6.1 Electroplating

This method is used to deposit metals and metallic alloys on conducting substrates. The electrolyte is an aqueous solution of an ionic compound, which contains positive ions of the film material. The mass of deposited material and the quantity of ions discharged at the cathode can be obtained from Faraday's law [67],

$$\frac{m}{A} = \frac{Mj\alpha t}{nF}, \qquad (3.6.1.1)$$

where m/A represents the mass per unit area, M the molecular weight, j the current density, α the current efficiency, t the time, n the valency and F the Faraday constant which is equal to 96490As/g.

There are 33 metals that can be electrodeposited [67], but only 14, consisting of Al, Ag, Au, Cd, Co, Cu, Cr, Fe, Ni, Pb, Pt, Rh, Sn, Zn are deposited regularly. All ions are accelerated towards electrodes bearing charges opposite to the ions. The ions screen the bulk of the electrolyte from most of the electric field by forming a double layer. There is a voltage drop in the double layer, resulting in relatively high magnetic field strengths reaching 10^7V cm $^{-1}$. The reactions undergone by positive ions in aqueous solution include dehydration, discharge, surface diffusion, nucleation and crystallization. The process of electroplating usually results in polycrystalline deposited films; however, in the case of Ni on Cu for example, epitaxial deposition on a single crystal cathode can be achieved. High growth rates can be achieved in electro-deposition, for example films can be deposited at a rate of 1 monolayer/μs at a current density of 1A cm $^{-2}$.

3.6.2 Electroless Plating

The electrochemical reactions in electroless plating processes occur without the presence of an external field, for example using formaldehyde or sugars as a reducing agent, silvering of mirrors by $AgNO_3$ solution can be achieved. The deposition of Ag occurs on any surface placed in such a bath. Some electroless reactions are more specific in that they will only occur on certain surfaces, such as the reaction NiCl + sodium hypophosphite will deposit Ni metal only on Ni, Co or Fe surfaces [67]. Sometimes parts

of surfaces can be activated with special substances such as $PdCl_2$ solution, used for activating nonmetallic surfaces for Ni or Cu deposits. Electroless plating is highly specific and difficult to control. It works very well for nonconducting substances and is useful for deposition in hard-to-reach areas [67].

3.6.3 Vacuum Evaporation

Physical vacuum evaporation is one of the most widely used methods for the deposition of thin films. Data obtained from evaporated films has contributed greatly to the basic understanding of film growth mechanisms and the relationships between the structure and physical properties of films. Evaporation is carried out by heating the source material in a variety of ways including use of electric currents to heat a metallic boat made of the source material, flash evaporation, arc evaporation, and electron guns.

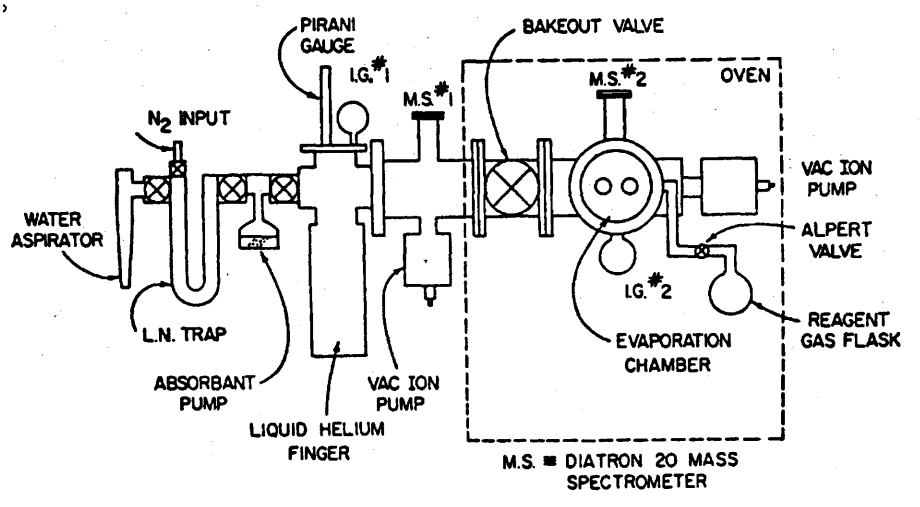


Figure 3.5 Schematic of a typical ultra-high vacuum evaporation system (taken from Reference [66]).

For evaporation to occur, a net flux of atoms out of the condensed phase must be produced, and this occurs when the partial pressure at the solid-vapor phase boundary of the source material is lower than the vapor pressure. A very good vacuum system is essential for this process because the mean free path of particles decreases with increasing pressure, so that higher pressures would cause the evaporated atoms to recondense near the source material. The most universally applicable heating source for evaporation is the electron gun. The most efficient electron guns are capable of a 270° magnetic electron beam deflection, and therefore can use the whole half space for evaporation, while simultaneously avoiding the exposure of the electron optics and filament to the vapor. Reaction of the source material with the containing crucible is also avoided because only a small part of the source is melted at any time.

According to the kinetic theory of gases, the amount of gaseous atoms striking a surface per unit area and per unit time at a pressure P is given by [67]

$$\frac{dN}{dAdt} = \frac{P}{\sqrt{2\pi m k_B T}}, \tag{3.6.3.1}$$

where T is the absolute temperature, m is the mass of the particles, and k_B is the Boltzmann constant. In a homogeneous system consisting of a condensed phase in equilibrium with its vapor at saturation vapor pressure, the number of evaporating atoms is the same as the number condensing, which is the quantity given by the equation above. The equation then also describes the number of evaporating atoms under equilibrium conditions and the equation is assumed to hold also under non-equilibrium conditions. The vapor pressure is then related to the temperature according to the Clausius-Clapeyron equation [67],

$$\frac{dP^*}{dT} = \frac{\Delta H(T)}{T(V_g(T) - V_c(T))}, \tag{3.6.3.2}$$

where $\Delta H(T)$ is the molar heat of evaporation or sublimation, and $V_g(T)$ and $V_c(T)$ are the molar volumes of the gaseous and condensed phases respectively. Assuming ideal gas behavior of the vapor, the above expression can be integrated to yield the following expression for the pressure

$$P^*(T) = P_0 e^{-\frac{\Delta H / N_A}{k_B T}}. \tag{3.6.3.3}$$

Physical evaporation can then be considered a thermally activated process, with an atomic activation energy of $\Delta H/N_A$, where N_A is Avogadro's number. The exponential represents the probability of an atom receiving energy from the thermal reservoir to overcome the potential barrier $\Delta H/N_A$. The number of evaporated particles is therefore strongly dependent on the binding energy of a surface atom, which is the heat of sublimation. A pressure $P^* \approx 0.1$ Torr is usually necessary for a useful film growth rate [67]. Higher temperatures are also necessary to obtain an appreciable number of evaporated atoms. For $P^* = 0.1$ Torr, the required minimum temperatures for a few metals are [67], Fe = 1920K, Cu = 1790K, Al = 1640K, Sb = 885K.

Evaporation of metals results in surface evaporation only because of the high thermal conductivity of metals, which makes it difficult to sustain a temperature gradient necessary for vapor formation below metal surfaces. The vapor density as a function of direction relative to the surface normal must be known in order to achieve optimal substrate positioning. At equilibrium, the number densities of condensing and

evaporating atoms must be equal; therefore the angular atomic velocity distribution is isotropic. The number of condensing atoms per unit area and per unit time is therefore proportional to the number of atoms in a cylinder of volume $n\upsilon\cos\theta dA$, where n is the number of atoms per unit volume and υ their velocity. Using Knudsen's law together with these observations, the following expression [67]

$$\frac{dN}{dAdtd\Omega} = \frac{P^*}{\sqrt{2\pi mk_BT}}\cos\theta, \tag{3.6.3.4}$$

develops from which the thermal velocities of evaporated atoms can be obtained. The Maxwell velocity distribution is [67]

$$\frac{dn}{d\upsilon} = \frac{4n}{\sqrt{\pi}}\left(\frac{m}{2k_BT}\right)^{3/2}\upsilon^2 e^{-m\upsilon^2/2k_BT}, \tag{3.6.3.5}$$

with the average energy $\langle E\rangle$ given by [67],

$$\langle E\rangle = m\langle\upsilon^2\rangle/2 = 3k_BT/2, \tag{3.6.3.6}$$

and

$$\sqrt{\langle\upsilon^2\rangle}:\langle|\upsilon|\rangle:\overline{\upsilon} = \sqrt{\frac{3}{2}}:\sqrt{\frac{4}{\pi}}:1, \tag{3.6.3.7}$$

where $\overline{\upsilon}$ is the velocity corresponding to maximum $dn/d\upsilon$.

The number of gas atoms per unit volume at pressure P is,

$$n = P/k_B T \,. \tag{3.6.3.8}$$

At atmospheric pressure, and room temperature, there are 2.5 x 10^{20} atoms present in a volume of 1 cm^3. When the pressure is reduced to 10^{-8} Torr, there are 3.2 x 10^8 atoms per cm^3. An atom of diameter can strike gaseous atoms of the same size, giving a mean free path (MFP) of $1/nd^2\pi$ [67]. The number of migrating atoms still collision-free after a distance z is [67]

$$N(z) = N(z=0)\exp(-z/MFP)\,. \tag{3.6.3.9}$$

For a distance between source and substrate of z = 20 cm and evaporation at 10^{-5} Torr, almost all of the evaporated atoms reach the substrate without undergoing any collisions. Straight trajectories are therefore obtained which is very important in masking applications in microelectronics [67]. The number of atoms striking the substrate in an evaporation process is then [67],

$$\frac{dN}{dA_s dt} = \frac{P^* A_{src}}{4r_0^{\,2}\sqrt{2\pi m k_B T}}\,, \tag{3.6.3.10}$$

where A_{src} is the surface area of the source material and r_0 is the atomic radius.

3.6.4 Sputter Deposition

The first reports of sputter deposited films occurred in 1852, when Sir W. R. Grove investigated a glow discharge and discovered that surface coatings were formed. The use of the term 'sputtering' to describe this phenomenon was initiated by Sir W. Thomson [67]. Sputtering is the phenomenon whereby atoms or ions having energies greater than

30 eV strike a surface, referred to as the target, and eject material from that surface [27, 66 - 67]. The process of sputtering is the result of momentum transfer from the incident particle to the atoms of the target substance. The sputtering yield depends on the energy as well as the mass of the incident particles. The number of atoms of material ejected per incident atom or ion is termed the sputtering yield. The sputtering yield Y can be expressed as [27, 67]

$$Y = \frac{n_s}{n_0} = \frac{\langle E_{kin} \rangle}{\Delta H^{sub}} \frac{m_0 m_s}{m_0 + m_s}, \tag{3.6.4.1}$$

with n_0 and n_s being the numbers of incident and sputtered off particles respectively, with respective masses m_0 and m_s, $\langle E_{kin} \rangle$ is the average kinetic energy of the incident particles, and ΔH^{sub} is the heat of sublimation of the source material.

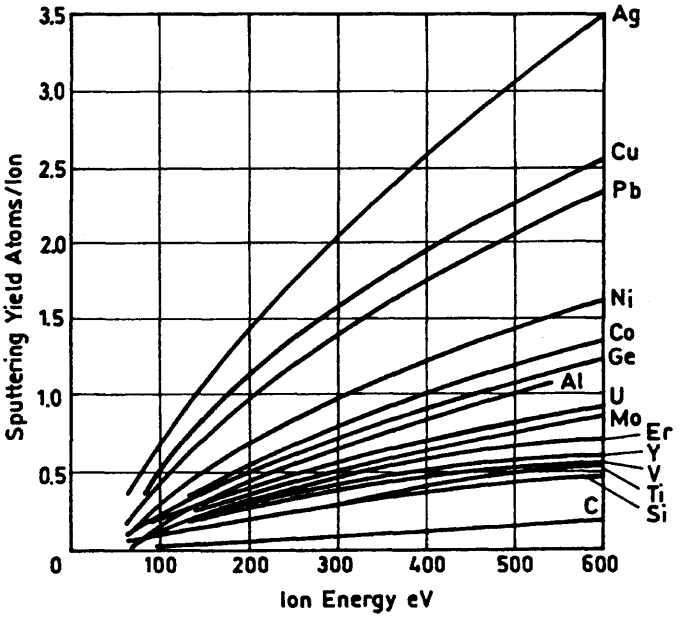


Figure 3.6 Sputtering yield for various elements as a function of Ar+ energy (taken from Reference [67]).

The angular distribution of particles sputtered off a surface depends on the angle of incidence of the incident particles. Particles that arrive at oblique angles of incidence result in higher sputtering yields than those that arrive normal to the target surface [66]. This effect is useful for the sputter etching of surfaces. Particles sputtered from single crystal targets exhibit preferred directions when they leave the target. Usually approximately 95% of the kinetic energy possessed by the incident ions goes into initiating lattice vibrations in the sputtering target, which causes a large increase in its temperature [67]. The remaining 5% of the kinetic energy is transferred to the sputtered particles. Sputtering with ions enables the ions to be accelerated in a controlled manner by an electric field.

In a triode sputtering system [27, 66], a source of electrons from for example, a heated tungsten filament is accelerated in a gas at low pressure to a strategically placed anode. The electrons ionize the gaseous atoms on their way to the anode to create a plasma. The target is placed in the plasma and negatively biased with respect to it, while a substrate is placed in the appropriate position if it is desired to collect the sputtered material. Positive ions from the plasma are attracted to the negatively biased target where they form a space charge due to their low mobility relative to the electrons [66]. The space charge region is also referred to as the Langmuir dark space [27]. An overwhelming majority of the voltage difference between the plasma and the target is dropped across the Langmuir dark space. A DC (direct current) triode sputtering system is shown in Figure 3.7. The potential of the plasma is defined by the most positive electrode in the assembly. The substrate holder is usually also biased negatively with respect to the plasma and therefore can be subject to some sputtering itself. The negative

bias on the substrate holder serves to deflect thermionically emitted electrons from it and avoid distortions in the plasma uniformity.

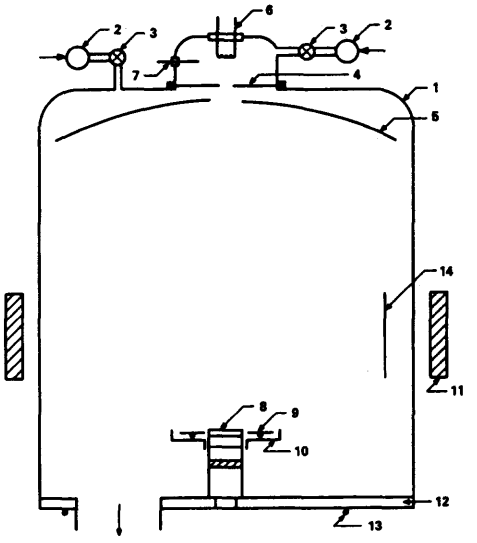
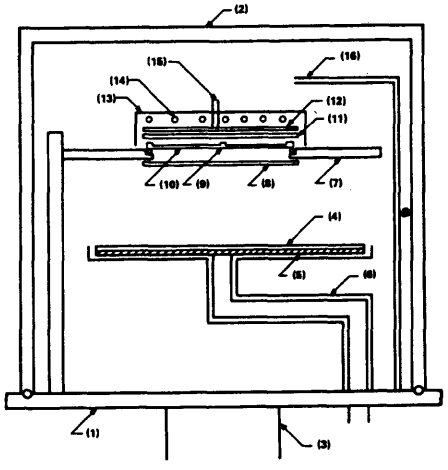


Figure 3.7 Schematic representation of a DC triode sputtering system (taken from Reference [83]).

To increase the probability of the emitted electrons ionizing the gas in the sputtering chamber for a given traveled distance, a magnetic field is usually applied around the plasma, which causes the electrons to spiral around the magnetic lines of force [27, 66]. The triode discharge does not depend on secondary emission to sustain itself and can therefore operate at relatively low pressures. The ion current to the target can also be modified without changing the gas pressure or the voltage at the target material.

3.6.4.1 DC Glow Discharge Sputtering. A DC glow discharge sputtering system is ignited when the critical number of auxiliary charge carriers for the system is exceeded [27, 66]. This is achieved by the ionization of neutral atoms by primary ions generated by some other means such as externally applied radiation. A typical DC diode sputtering apparatus is shown in Figure 3.8. Positive ions accumulate in front of the cathode where they form a dark space referred to as the Crookes dark space [27, 66]. When accelerated across the Crookes dark space, electrons make relatively few collisions with gaseous atoms until they arrive at the negative glow region. The electrons begin ionization of the gas in the negative glow region, losing energy with each collision. The radiation emitted by excited atoms causes a visible glow. The Faraday dark space is formed [27] when electrons that originated at the cathode lose most of their energy and are unable to cause further ionization of gaseous atoms.



Schematic diagram of a dc diode sputtering system.
(1) baseplate (2) bell-jar (3) pump (4) cathode (5) water-cooled copper electrode (6) dark space shield (7) anode (8) shutter (9) substrate holder (10) substrate (11) pyrex plate (12) metal plate (13) heater box (14) Ta resistance heater (15) thermocouple (16) gas inlet.

Figure 3.8 Schematic diagram of a DC diode sputtering apparatus (taken from Reference [83]).

The electrons are able to move by diffusive means only at this stage. The electrons acquire increased speed after traversing the dark space and gain sufficient energy from the anodic attraction to cause more ionization. The positive column is the region between the end of the Faraday dark space region and the anode. The anode of the glow discharge as well as the substrate are usually placed in the negative glow region before the occurrence of the Faraday dark space, so that the Faraday dark space is not usually visible in operational sputtering systems [66].

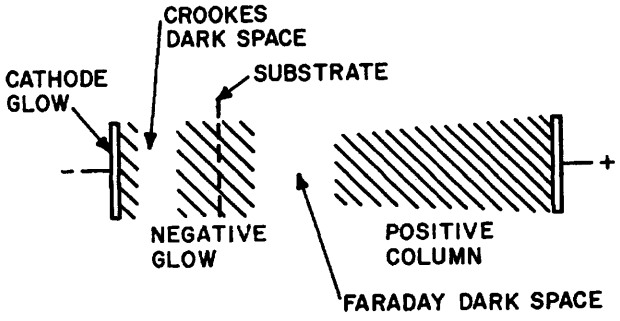


Figure 3.9 Cross sectional view of a glow discharge (taken from Reference [66]).

The substrate is often used as the anode. In order to protect the back of the cathode from being sputtered, a grounded shield is usually placed near the back of the cathode. An inert gas, usually argon, is used for sputtering to avoid chemical reactions between the sputtering gas and the depositing film. A discharge cannot be generated between any two electrodes that are separated by a distance less than the Crooke's dark space, because electrons do not undergo a significant amount of ionizing collisions until they have traversed this region. The extension of the Crooke's dark space is described by [67]

$$d = \frac{1}{P}\left(a + \frac{b}{V - c}\right), \qquad (3.6.4.1.1)$$

where d is the distance, P is the pressure, V is the voltage, and a, b and c are constants. Often it is desired to confine the plasma close to the cathode to increase the efficiency of the sputtering process. Magnetron sputtering consists of using magnets to confine the plasma to a small space around the cathode, minimizing particle dispersion and enhancing the number of particles available to bombard the target [66].

DC sputtering cannot be used to sputter insulator targets because nothing would happen with the application of voltage. Assuming that a plasma could be created by some other means, any positive ions which manage to strike the target would not be neutralized and would accumulate a positive charge on the target surface, making further positive ion bombardment impossible. There would be no way to impart sufficient energy to plasma ions in order to cause sputtering. The use of an AC power source would be able to transmit energy to the plasma, but a low frequency AC source would lead to the ions arriving at the insulator target not having enough energy to cause sputtering because most of the applied voltage would be dropped across the insulator due to the good conductivity of the plasma. A high frequency (> 100 kHz) AC source would cause the heavier ions to lag behind the electrons in following the periodic electric field changes and being less likely to reach the target in a single cycle, the ions would begin to accumulate in front of the target to form a space charge region, similar to what would occur in a DC sputtering system [66]. Radio frequency sputtering overcomes these limitations with regard to the sputtering of insulator targets.

3.6.4.2 Radio Frequency (RF) Sputtering. RF sputtering operates by using frequencies above 100 kHz to feed energy into the plasma and to impart energy to the ions in sufficient amounts to effect the sputtering process [66]. The RF discharge improves over the DC discharge in efficiency of ion and electron generation within the sputtering gas. The electrons can oscillate in the RF field over several cycles and thereby acquire enough energy to cause ionization. An RF discharge operating at frequencies over 100 kHz would be self-sustained down to pressures of 1 millitorr [66]. RF sputtering, unlike DC sputtering, can be used to sputter insulators. RF sputtering is usually carried out at a frequency of 13.66 mHz because of difficulties associated with operating at higher frequencies. The power is admitted into an RF sputtering system by an impedance matching network, otherwise most of the RF generator power will be reflected back and will not reach the discharge.

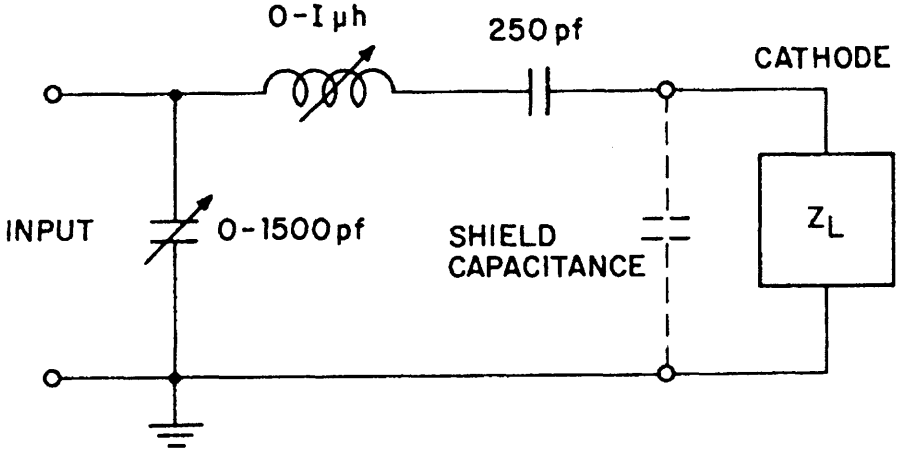


Figure 3.10 Impedance matching network for RF sputtering (taken from Reference [66]).

The impedance of the RF generator is usually 50, much more than that of the discharge. A capacitor in series with the variable inductor is usually used [27, 66] in the

network to increase the inductor when necessary and to enable the accumulation of self-bias voltage. This capacitor, called a blocking capacitor, is essential if metals are to be sputtered with the RF system. There is also usually a shield capacitance [27, 66], which is reduced by placing the cathode-shield as far away from the RF electrode as possible without exceeding the Crookes dark space. In an RF system, sputtering of the substrate holder occurs to a much smaller extent than sputtering of the target because the capacitance between the RF plasma and the grounded surfaces is much greater than that between the plasma and the target, resulting in most of the plasma potential being discharged at the target surface. To eliminate re-sputtering at the substrate, symmetrical RF sputtering systems have been designed [66] containing two targets, and which eliminate the problem of re-sputtering to a large extent.

Sputter deposited materials usually achieve better adhesion to the substrate than evaporated materials [66 – 67]. It is also possible to clean the substrate to be used inside the sputtering apparatus by ion bombardment. Uniform thicknesses over large substrate areas are also easy to achieve with sputter deposition. The thickness of sputtered films is proportional to the deposition time, making it straightforward to control deposited film thicknesses. However, sputter deposition rates are usually less than 40 $Å s^{-1}$ [67], the target material to be used must be available in sheet form, and due to the high kinetic energy of the impinging ions, the substrate must be cooled frequently when sputtering is carried out for long periods.

From work such as that of Tian and coworkers [154 - 155], and Marcus and coworkers [169], who have used total energy calculations to show that the bulk structure of a thin Ti film on an Al{001} substrate is strained fcc Ti, it is evident that the nature of

the substrate on which a thin film is deposited can determine the kind of stress prevailing in the deposited film, and therefore the final structural form of the film. As previously noted, β-Ta has been observed to form preferentially on certain substrates, so the nature of the substrate upon which thin tantalum films are deposited appears to play an important role in the subsequent formation of the β phase. It is not known however whether metastable tantalum structures are formed at any stage of the creation of β tantalum in a substrate-strain mediated process analogous to what has been related above for vanadium.

In sputter-deposition experiments, transport properties are often used as an in-process monitor of the type of tantalum film being deposited because these properties are sensitive to the structure of the film, as well as the amount of impurities incorporated into the film. In Chapter 6, an attempt is made to predict the metastable tantalum structure that would result from a substrate-strain mediated process, calculate its transport properties and then compare them to the observed transport properties of β-Ta. One objective of this work then is to investigate whether tetragonal distortions of the stable bcc structure of tantalum can form a metastable phase that can account for the large resistivity of β–Ta relative to α bcc Ta. This is accomplished by calculating, for tantalum as well as for the rest of the bcc metals in groups 5B – 6B and periods 4 - 6 of the Periodic Table, the total energy per unit cell as a function of the unit cell volume and the c/a ratio. The electrical resistivities are then calculated for all the bcc metals and their predicted bct phases using the Lowest Order Variational Approximation (LOVA) formalism for the solution of the Bloch-Boltzmann transport equation. Different alloy combinations of the metals in this study have been prepared as metastable phases before [170 – 172]. In addition,

metastable fcc phases of Ta [96], Cr [150], Mo and W [151] have been observed experimentally, but as far as is known, except for V, no metastable bct phases of these metals which do not have the fcc or bcc structure have been prepared in elemental form.

An important factor that determines the structure of sputter-deposited films is the presence of impurities in the sputtering environment. In the initial stages of metallic atom deposition onto a substrate to form a film, the concentration of impurities is comparable to that of incident metal atoms. Under these conditions, impurities such as nitrogen and oxygen could act as electron sinks by charge transfer from the metallic atoms to reduce the energy of interaction of the incident atoms with the substrate. The resulting reduced electron concentration in the metallic atoms could mean that the magnitude of the overall charge density is altered from the amount necessary to bind the metallic ion cores into the customary bcc packing arrangement, causing the atomic arrangement in the first few layers of film to deviate from bcc. This could explain the previously discussed observations by a number of authors [91, 108, 110, 113 - 115] that the presence of impurities seems to result in the nucleation of β-Ta. The presence of impurities in the sputtering environment and their influence on film properties will be covered in more detail in Chapter 7.

In this Chapter, nucleation theory and various models of film growth, as well as thin film deposition methods that sometimes result in metastable phase growth have been discussed. In the following Chapter, the computational method used to study metastable phase formation is discussed in detail.

CHAPTER 4

DENSITY FUNCTIONAL THEORY

4.1 The Interacting Many-Body Problem

A condensed phase or solid is made up of nuclei, which are relatively heavy positively charged particles, and electrons, which are relatively lighter negatively charged particles. Such systems represent a many-bodied quantum mechanical problem so that if N nuclei are present, the system contains $N + ZN$ particles, all interacting with one another. The non-relativistic Hamiltonian operator for such a system, following the treatment in References [173] and [174], is

$$\hat{H} = -\frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\vec{R}_i}}{M_i} - \frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\vec{r}_i}}{m_e}$$

$$-\frac{1}{4\pi\varepsilon_0}\sum_{i,j}\frac{e^2 Z_i}{\left|\vec{R}_i - \vec{r}_j\right|} + \frac{1}{8\pi\varepsilon_0}\sum_{i\neq j}\frac{e^2}{\left|\vec{r}_i - \vec{r}_j\right|} + \frac{1}{8\pi\varepsilon_0}\sum_{i\neq j}\frac{e^2 Z_i Z_j}{\left|\vec{R}_i - \vec{R}_j\right|}, \tag{4.1.1}$$

where M_i is the nuclear mass at position $\vec{R}_i$ and m_e is the mass of each electron at position $\vec{r}_i$. The Hamiltonian for the system can be re-written as

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{N-e} + \hat{V}_{e-e} + \hat{V}_{N-N}, \tag{4.1.2}$$

where $\hat{T}_N$, $\hat{T}_e$, $\hat{V}_{N-e}$, $\hat{V}_{e-e}$ and $\hat{V}_{N-N}$ represent respectively the kinetic energy operator for the nuclei, the kinetic energy operator for the electrons, the operator for the Coulombic

interaction energy between the nuclei and the electrons, the operator for the Coulombic interaction energy between the electrons, and the operator for the Coulombic interaction energy between the nuclei. It is not possible to obtain an exact solution for the problem posed by this system of interacting particles as represented by the Hamiltonian operator above. Three approximations at varying levels of conceptual complexity are required to obtain acceptable eigenfunctions and eigenvalues for the system [173]. The first necessary approximation is the Born-Oppenheimer approximation, which neglects the motion of the nuclei, regarding them as stationary since they are so much heavier than the electrons and move much more slowly. This approximation enables the electrons to be fixed, while the electrons are assumed to be moving in the field of the stationary nuclei and in instantaneous equilibrium with them. The nuclei are thereby reduced to a fixed source of positive charge, and the system now consists only of ZN negative particles, interacting with one another and moving in the external and constant potential due to the nuclei. The Hamiltonian in Equation (4.1.1) is then reduced to

$$\hat{H} = \hat{T}_e + \hat{V}_{N-e} + \hat{V}_{e-e}, \tag{4.1.3}$$

because since the nuclei are now assumed to be stationary, $\hat{T}_N$ becomes zero, $\hat{V}_{N-N}$ is reduced to a constant and $\hat{V}_{N-e}$ is now the potential energy of the electrons due to the external potential of the fixed nuclei, and may be referred to as $\hat{V}_{ext}$ [173]. The kinetic and potential energy terms depend only on the fact that a many-electron system (as opposed to a proton-neutron system for instance) is being studied and does not depend on

the particular kind of many-electron system. All system specific information is contained in the external potential term $\hat{V}_{ext}$.

After making the Born-Oppenheimer approximation, the new form of the quantum mechanical many-body problem is much simpler than the original form, but still far too difficult to solve exactly. The next approximation to be made in order to make the problem more tractable can take a variety of forms such as the Hartree-Fock method, which has worked very well for atoms and molecules, but not for solids, and Density Functional Theory (DFT). Following two theorems by Hohenberg and Kohn in 1964 [173 - 174], DFT was established as the dominant problem-solving methodology in the quantum mechanics of solid-state physics. It is a general method used to solve the quantum-mechanical many-body problem. Not only can it be used to solve problems consisting of an interacting all-electron gas, as has been developed so far, but it can be applied to a proton-neutron gas in order to construct models of nuclear particle interaction, or to the gas consisting of interacting nuclei and electrons without using the Born-Oppenheimer approximation in order to describe solids containing light elements.

4.2 The Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem states that there is a one-to-one correspondence between the ground state density $\rho(\vec{r})$ of a many electron system, which may be an atom, a molecule or a solid, and the external potential $\hat{V}_{ext}$. An immediate consequence of this correspondence is that the expectation value of any observable $\hat{O}$ in the ground state is a unique functional of the exact ground-state electron density [173],

$$\langle \Psi | \hat{O} | \Psi \rangle = O[\rho(\vec{r})]. \tag{4.2.1}$$

The second Hohenberg-Kohn theorem states that when the observable $\hat{O}$ is the Hamiltonian $\hat{H}$, the ground state total energy functional $H[\rho(\vec{r})]$, which equals $E_{V_{ext}}[\rho(\vec{r})]$, has the form [173]

$$E_{V_{ext}}[\rho(\vec{r})] = \langle \Psi | \hat{T}_{e-e} + \hat{V}_{e-e} | \Psi \rangle + \langle \Psi | \hat{V}_{ext} | \Psi \rangle, \tag{4.2.2}$$

which is equivalent to

$$E_{V_{ext}}[\rho(\vec{r})] = F_{HK}[\rho(\vec{r})] + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r}, \tag{4.2.3}$$

where

$$F_{HK}[\rho(\vec{r})] = \langle \Psi | \hat{T}_{e-e} + \hat{V}_{e-e} | \Psi \rangle. \tag{4.2.4}$$

The Hohenberg-Kohn density functional $F_{HK}[\rho(\vec{r})]$ is universal and in principle should apply to any many-electron system. The ground state total energy functional $E_{V_{ext}}[\rho(\vec{r})]$ achieves its minimum value at the ground state electron density corresponding to the external potential $\hat{V}_{ext}$, which is unique for any given many-electron system. All the observable quantities of a system can be obtained in a unique manner (can be represented as unique functionals of the density); all that is needed is the density. If the ground state density is known, the contribution of the external potential to the total energy can be exactly calculated.

The fact that the Hohenberg-Kohn density functional $F_{HK}[\rho(\vec{r})]$ is universal for any and all many-electron systems means that in principle an expression for $F_{HK}[\rho(\vec{r})]$ exists which can be used to describe any possible atom, molecule or solid, but an explicit expression for it is not known [173]. The second Hohenberg-Kohn theorem makes it possible to use the Rayleigh-Ritz variational principle to find the ground state density of any many-electron system if an approximation to $F_{HK}[\rho(\vec{r})]$ is known. The density, out of the infinite number of possible densities, that minimizes $E_{V_{ext}}[\rho(\vec{r})]$ is the ground-state density corresponding to $\hat{V}_{ext}$. When $E_{V_{ext}}[\rho(\vec{r})]$ is evaluated for the ground-state density corresponding to the $\hat{V}_{ext}$ unique to a particular system, the ground state energy of the system results.

4.3 The Kohn-Sham Equations

The Kohn-Sham equations enable a practical procedure in order to obtain the ground-state density of a many-electron system. Denoting the total energy functionals corresponding to the exact Hamiltonian and the Hartree-Fock Hamiltonian respectively as $E_e[\rho(\vec{r})]$ and $E_{HF}[\rho(\vec{r})]$, they can be represented as [173]

$$E_e[\rho(\vec{r})] = T_{e-e}[\rho(\vec{r})] + V_{e-e}[\rho(\vec{r})], \tag{4.3.1}$$

and

$$E_{HF}[\rho(\vec{r})] = T^0_{e-e}[\rho(\vec{r})] + V_H[\rho(\vec{r})] + V_{ex}[\rho(\vec{r})], \tag{4.3.2}$$

where

$$V_{e-e}[\rho(\vec{r})] = V_H[\rho(\vec{r})] + V_{ex}[\rho(\vec{r})], \tag{4.3.3}$$

and $T_{e-e}[\rho(\vec{r})]$ and $V_{e-e}[\rho(\vec{r})]$ are respectively, the exact kinetic and electron-electron potential energy functionals, $T^0_{e-e}[\rho(\vec{r})]$ is the functional representing the kinetic energy of a non-interacting electron gas, $V_H[\rho(\vec{r})]$ represents the Hartree potential functional, and $V_{ex}[\rho(\vec{r})]$ represents the exchange functional [173].

The electronic wavefunction must be anti-symmetric with respect to interchange of any two electrons. The anti-symmetric property of the wavefunctions produces a spatial separation between electrons with identical spin coordinates, which reduces the Coulombic energy of interaction of the electronic system. This reduction in energy due to the anti-symmetry of the wave function to exchange is called the exchange energy. The Coulombic energy of the electronic system can be less than its Hartree-Fock value if electrons that have opposite spins are also spatially separated. However, separating electrons with opposite spins increases the total electronic kinetic energy.

The contribution of electronic correlation to the total energy is [173]

$$V_c[\rho(\vec{r})] = T_{e-e}[\rho(\vec{r})] - T^0_{e-e}[\rho(\vec{r})]. \tag{4.3.4}$$

The contribution of the exchange to the total energy functional $V_{ex}[\rho(\vec{r})]$ is defined as the part of the total energy functional, which is absent from the Hartree solution, but present in the Hartree-Fock solution.

$$V_{ex}[\rho(\vec{r})] = V_{e-e}[\rho(\vec{r})] - V_H[\rho(\vec{r})], \tag{4.3.5}$$

where the Hartree functional is defined as [173]

$$E_H[\rho(\vec{r})] = T^0_{e-e}[\rho(\vec{r})] + V_H[\rho(\vec{r})]. \tag{4.3.6}$$

The Hohenberg-Kohn functional $F_{HK}[\rho(\vec{r})]$ can then be represented as

$$F_{HK}[\rho(\vec{r})] = T^0_{e-e}[\rho(\vec{r})] + V_H[\rho(\vec{r})] + V_{xc}[\rho(\vec{r})], \tag{4.3.7}$$

with $V_{xc}[\rho(\vec{r})]$ equaling

$$V_{xc}[\rho(\vec{r})] = V_H[\rho(\vec{r})] + V_{xc}[\rho(\vec{r})]. \tag{4.3.8}$$

The total energy functional is then

$$E_{V_{ext}}[\rho(\vec{r})] = T^0_{e-e}[\rho(\vec{r})] + V_H[\rho(\vec{r})] + V_{xc}[\rho(\vec{r})] + V_{ext}[\rho(\vec{r})]. \tag{4.3.9}$$

This functional can be interpreted as the energy functional of a non-interacting classical electron gas, subject to two external potentials, namely one due to the positive charge of the nuclei and the other due to the effects of electronic correlation and exchange. The Hamiltonian operator corresponding to this functional is called the Kohn-Sham Hamiltonian $\hat{H}_{KS}$ and is given by [173]

$$\hat{H}_{KS} = -\frac{\hbar^2}{2m_e}\vec{\nabla}_i^2 + \frac{e^2}{4\pi\varepsilon_0}\int\frac{\rho(\vec{r})}{|\vec{r}-\vec{r}'|}d\vec{r} + \hat{V}_{xc} + \hat{V}_{ext}, \tag{4.3.10}$$

with the exchange-correlation potential equal to

$$\hat{V}_{xc} = \frac{\delta V_{xc}[\rho(\vec{r})]}{\delta\rho(\vec{r})}. \tag{4.3.11}$$

These equations represent the reduction of the interacting many-electron system to a system of independent electrons moving in an effective potential due to the other electrons. The Kohn-Sham eigenvalues represent the derivatives of the total energy with respect to the occupation numbers of the single-particle states. The Kohn-Sham theorem can then be re-formulated as the following: The exact ground-state density $\rho(\vec{r})$ of an N electron system is given by [173 – 174]

$$\rho(\vec{r}) = \sum_{n=1}^{N} \phi_n^*(\vec{r})\phi_n(\vec{r}), \tag{4.3.12}$$

where the single particle wave functions $\phi_n(\vec{r})$ are the N lowest-energy solutions of the Kohn-Sham equation [173]

$$\left[-\frac{\hbar^2}{2m_e}\vec{\nabla}_n^2 + \frac{e^2}{4\pi\varepsilon_0}\int\frac{\rho(\vec{r})}{|\vec{r}-\vec{r}'|}d\vec{r} + \hat{V}_{xc} + \hat{V}_{ext}\right]\phi_n(\vec{r}) = \varepsilon_n\phi_n(\vec{r}). \tag{4.3.13}$$

The far more difficult regular problem represented by the single Schrödinger equation has now been reduced to a set of Schrödinger-like non-interacting single particle equations. These equations must be solved self-consistently because the potentials $\hat{V}_H$ and $\hat{V}_{ext}$ depend on the density $\rho(\vec{r})$, which depends on the wavefunctions, which in turn depend on the potentials $\hat{V}_H$ and $\hat{V}_{ext}$. The iterative method [173 – 174] used to solve this problem is to guess some starting density $\rho_0(\vec{r})$ from which a Kohn-Sham Hamiltonian

$\hat{H}_{KS}^{0}$ is constructed. The Kohn-Sham equations are solved for the eigenvalues ε_n^{1} and eigenfunctions $\phi_n^{1}(\vec{r})$ from which a new density $\rho_1(\vec{r})$ is created according to Equation (4.3.12). If $\rho_1(\vec{r})$ differs from $\rho_0(\vec{r})$, $\rho_1(\vec{r})$ is used to construct yet another new Hamiltonian $\hat{H}_{KS}^{1}$, from which yet another density $\rho_2(\vec{r})$ is extracted. This iterative process is continued until a density $\rho_f(\vec{r})$ is obtained from which a Hamiltonian $\hat{H}_{KS}^{f}$ results, which in turn leads to eigenfunctions $\phi_n^{f}(\vec{r})$, which reproduce the density $\rho_f(\vec{r})$. This procedure is shown in Figure 4.1.

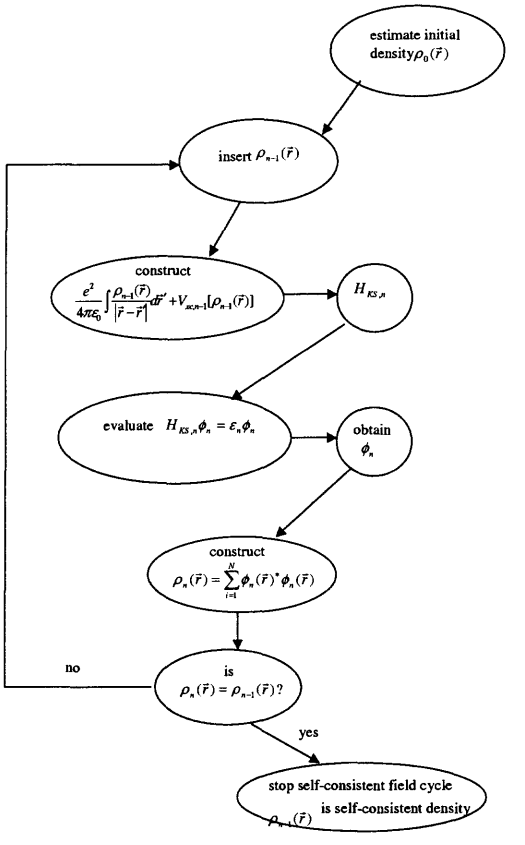


Figure 4.1 The self-consistent cycle involved in solving the Kohn-Sham equations.

The last approximation to be made concerns the exchange-correlation functional, which is not known exactly. The expression above for $V_{XC}(r)$ will calculate $V_{XC}(r)$ exactly if the exchange-correlation functional $V_{xc}[\rho(\vec{r})]$ was known exactly. Usually only an approximate expression for $V_{xc}[\rho(\vec{r})]$ can be obtained. The simplest way to obtain this approximate expression is to use the local density approximation (LDA) in which the $V_{xc}[\rho(\vec{r})]$ per electron at a point r in the electron cloud is set equal to the $V_{xc}[\rho(\vec{r})]$ per electron of a uniform electron cloud of the same electron density as the original electron cloud at the point r. $V_{xc}[\rho(\vec{r})]$ is therefore written [173] in the local density approximation as

$$V_{xc}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon^{u}{}_{xc}(\rho\uparrow(\vec{r}), \rho\downarrow(\vec{r}))d\vec{r}\,, \tag{4.3.14}$$

where $\varepsilon^{u}{}_{xc}$ is the exchange-correlation energy per electron of a uniform electron gas having spin up and spin down densities. The LDA therefore ignores corrections to $V_{xc}[\rho(\vec{r})]$ at point r due to nearby non-uniformity in the electron density. The ground state exchange energy can be expanded in the gradient [175]

$$E[\rho(r)] = A_1 \int \rho^{4/3} dr + A_2 \int \rho^{4/3} |\nabla\rho|^2 dr + \text{higher terms in } |\nabla\rho|\,. \tag{4.3.15}$$

If the density of the exchange hole at $r+R$ about an electron at r is $\rho_{ex}(r, r+R)$, then it follows that [175]

$$\rho_{ex}(r,r) = \frac{-\rho(r)}{2},\ \rho_{ex}(r, r+R) \le 0\,, \tag{4.3.16}$$

and

$$\int \rho_{ex}(r, r+R)dr = -1. \tag{4.3.17}$$

The local density approximation (LDA) corresponds to keeping only the first term in the expansion represented by Equation (4.3.15) [175]. The gradient expansion approximation (GEA), which keeps two terms in the gradient expansion, violates the above conditions. The generalized gradient approximation (GGA) parameterizes a functional of the density gradient in such a way that the conditions on the electron hole are satisfied by the gradient expansion form of the exchange energy and it is possible to take functional derivatives to obtain the exchange potential. The GGA, unlike the LDA, reflects the fact that the exchange hole around an electron is not spherically symmetric around the electron [175].

The motivation for developing alternative approximations to the exchange-correlation energy such as GGAs came partly from the fact that solid cohesive energies predicted by the LDA were consistently greater than those determined experimentally [176 - 177]. The LDA also tends to underestimate equilibrium unit cell volumes while the GGA tends to overestimate them. The reverse is true for bulk moduli, they are overestimated by the LDA and underestimated by the GGA [176 - 177]. The LDA however performs better in bulk solid calculations than some GGA functionals obtained by fitting to atomic and molecular data. Exchange energies for atoms calculated with the GGA are within 1% of the Hartree-Fock values, better than any other form of exchange energy especially the LDA form, which results in an average error of 14% [173]. The GGA tends to improve over the LDA when computing atomic and molecular properties

and other quantities such as total energies and structural energy differences [176]. The GGA expresses the exchange-correlation energy $\varepsilon^{u}{}_{xc}$ in the form [175 - 176],

$$V_{xc}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon_{xc}^{GGA}(\rho^{up}(\vec{r}), \rho^{dn}(\vec{r}), \nabla\rho^{up}(\vec{r}), \nabla\rho^{dn}(\vec{r}))d\vec{r}\,, \tag{4.3.18}$$

with $\rho^{up}(\vec{r})$ and $\rho^{dn}(\vec{r})$ representing respectively, the spin-up and spin-down electron density. ε_{xc}^{GGA} is not uniquely defined and there are several existing forms in the literature. Yet another approximation to the exchange-correlation energy is the meta–GGA (MGGA) which is of the form [175 - 176],

$$V_{xc}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon_{xc}^{GGA}(up, dn, \nabla up, \nabla dn, \nabla^2 up, \nabla^2 dn)d\vec{r}\,, \tag{4.3.19}$$

where *up* and *dn* represent $\rho^{up}(\vec{r})$ and $\rho^{dn}(\vec{r})$ respectively.

Any sort of approximation to the true exchange-correlation functional still results [173] in an infinite set of one-electron equations of the form

$$\left[-\frac{\hbar^2}{2m_e}\vec{\nabla}_n^2 + \frac{e^2}{4\pi\varepsilon_0}\int\frac{\rho(\vec{r})}{|\vec{r}-\vec{r}'|}d\vec{r} + \hat{V}_\lambda + \hat{V}_{ext}\right]\phi_n(\vec{r}) = \varepsilon_n\phi_n(\vec{r})\,, \tag{4.3.20}$$

where $\left[-\frac{\hbar^2}{2m_e}\vec{\nabla}_n^2 + \frac{e^2}{4\pi\varepsilon_0}\int\frac{\rho(\vec{r})}{|\vec{r}-\vec{r}'|}d\vec{r} + \hat{V}_\lambda + \hat{V}_{ext}\right]$ is the single particle Hamiltonian . If the Hartree-Fock approximation is in effect, the eigenfunctions $\phi_n(\vec{r})$ represent one-electron orbitals, correlation effects are not accounted for and $\hat{V}_\lambda$ represents the exchange operator, which treats exchange effects exactly. When the DFT approximation is in

effect, the eigenfunctions $\phi_n(\vec{r})$ are single-particle orbitals, $\hat{V}_\lambda$ becomes the exchange-correlation operator in other approximations within DFT such as the LSDA, GGA, or enhanced forms of these approximations. In the DFT formalism, exchange and correlation effects are treated in an approximate manner.

In order to solve the infinite set of one-electron equations for the eigenvalues and eigenfunctions, a suitable set of coefficients c_j^i with which to expand the eigenfunctions $\phi_i(\vec{r})$ in a suitable basis set $\psi_j(\vec{r})$ must be found in order to express the eigenfunctions in the form [173 – 174]

$$|\phi_i\rangle = \sum_{j=1}^{\infty} c_j^i |\psi_j\rangle . \tag{4.3.21}$$

Substituting Equation (4.3.21) for $\phi_i(\vec{r})$ into Equation (4.3.20) for each i, left multiplying with each basis function and then applying the variation theorem results in the set of equations [173]

$$\begin{bmatrix} \langle\psi_1|\hat{H}_{sp}|\psi_1\rangle - \varepsilon_1 S_{11} & \cdot & \cdot & \cdot & \langle\psi_1|\hat{H}_{sp}|\psi_J\rangle - \varepsilon_1 S_{1J} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \langle\psi_i|\hat{H}_{sp}|\psi_1\rangle - \varepsilon_i S_{i1} & \cdot & \langle\psi_i|\hat{H}_{sp}|\psi_j\rangle - \varepsilon_i S_{ij} & \cdot & \langle\psi_i|\hat{H}_{sp}|\psi_J\rangle - \varepsilon_i S_{iJ} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \langle\psi_J|\hat{H}_{sp}|\psi_1\rangle - \varepsilon_J S_{J1} & \cdot & \cdot & \cdot & \langle\psi_J|\hat{H}_{sp}|\psi_J\rangle - \varepsilon_J S_{JJ} \end{bmatrix} \begin{bmatrix} c_1^i \\ \cdot \\ c_j^i \\ \cdot \\ c_J^i \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \\ \cdot \\ 0 \end{bmatrix}, \tag{4.3.22}$$

where the overlap integral or overlap matrix element $\langle\psi_i|\psi_j\rangle$, represented by S_{ij}, is unity for $i = j$ and zero for $i \neq j$ if the basis set is orthogonal. The matrix equation

above represents a system of linear homogenized equations. For a solution to this system of equations to exist, the determinant of coefficients must vanish *i.e.*

$$\begin{vmatrix} \langle \psi_1 | \hat{H}_{sp} | \psi_1 \rangle - \varepsilon_1 S_{11} & . & . & . & \langle \psi_1 | \hat{H}_{sp} | \psi_J \rangle - \varepsilon_1 S_{1J} \\ . & . & . & . & . \\ \langle \psi_i | \hat{H}_{sp} | \psi_1 \rangle - \varepsilon_i S_{i1} & . & \langle \psi_i | \hat{H}_{sp} | \psi_j \rangle - \varepsilon_i S_{ij} & . & \langle \psi_i | \hat{H}_{sp} | \psi_J \rangle - \varepsilon_i S_{iJ} \\ . & . & . & . & . \\ \langle \psi_J | \hat{H}_{sp} | \psi_1 \rangle - \varepsilon_J S_{J1} & . & . & . & \langle \psi_J | \hat{H}_{sp} | \psi_J \rangle - \varepsilon_J S_{JJ} \end{vmatrix} = 0 . \tag{4.3.23}$$

The diagonalization of this determinant (Hamiltonian matrix) will result in J eigenvalues and J eigenfunctions, as well as J sets of coefficients c^i that are used to expand each eigenfunction $|\phi_i\rangle$. A higher value of J represents a larger basis set and a better approximation to $|\phi_i\rangle$, but at the cost of more Hamiltonian matrix diagonalization time.

When choosing an appropriate basis set, there are two tradeoffs involved. An efficient basis set needs only a few functions to describe the eigenfunctions with sufficient accuracy and leads to a small Hamiltonian matrix. However these small basis sets are highly problem-specific and cannot be used to obtain sufficiently accurate eigenfunctions for other problems. In cases where a larger basis set is required, limiting J would lead to unacceptably biased approximate wavefunctions, which would exhibit many basis function properties. The ideal basis set should be unbiased, meaning that it should not force predetermined solutions of the eigenvalue problem, and it should be efficient enough that matrix diagonalization times are reasonable. Further requirements are mathematical simplicity and tractability. Large matrices can be avoided by using the

pseudo-potential approximation in which only the valence electrons are used in the calculation.

4.4 Potential Energy Formulations in DFT

The pseudo-potential approximation represents one of many basis set formulations that were developed in the past for computational convenience. The most widely used ones in recent years are described [178] below.

4.4.1 The Pseudo-Potential Method

A simple and unbiased basis set is the plane wave basis set, in which any eigenfunction $\left|\psi_{n,\vec{k}}\right\rangle$ of a Hamiltonian periodic in the lattice with lattice vectors $\left\{\vec{K}_i\right\}$ can be expanded with an infinite set of coefficients $c_{\vec{K}}^{n,\vec{k}}$,

$$\left|\psi_{n,\vec{k}}\right\rangle = \sum_{\vec{K}} c_{\vec{K}}^{n,\vec{k}} e^{i(\vec{k}+\vec{K})\cdot\vec{r}} \,. \tag{4.4.1.1}$$

In practice, it is necessary to limit the plane wave basis set size, and this is achieved by restricting the basis functions to vectors $\vec{K}$ having $K \le K_{\max}$. This is equivalent to restricting the reciprocal lattice vectors of the basis functions to lie inside a sphere of radius $K_{\max}$ centered at the origin of reciprocal space. Equivalently, the free electron energy $E_{cut-off}$ corresponding to $K_{\max}$ can be specified to limit the basis set size, this energy, referred to as the cut-off energy, is given by

$$E_{cut-off} = \hbar^2 K_{\max}^2 / 2m_e \,. \tag{4.4.1.2}$$

The number of plane waves needed in the basis set is determined by the length scale magnitudes over which the wavefunction varies appreciably. For example [173] to describe wavefunction variations within 0.01Å, plane waves with a period an order of magnitude less than this distance, say 0.001Å, are needed, corresponding to a K_{max} order of magnitude of $K_{max} = 2\pi/10^{-13} = 6.3\times10^{13}$ meters. The volume of a sphere of this radius divided by the volume of the first Brillouin zone is about 4×10^{10} for a cubic lattice with a lattice constant of 3Å. About 10^{10} plane waves will be needed when it is realized that the first Brillouin zone contains one $\vec{K}$, leading to a $10^{10}\times10^{10}$ matrix.

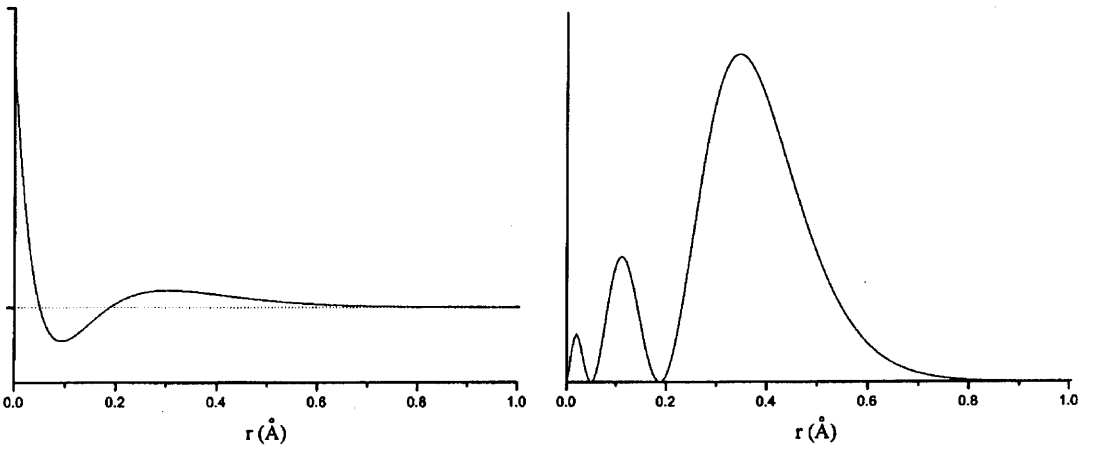


Figure 4.2 The radial part (left panel) and radial probability distribution (right panel) of a Ca 3*s* wave function (*y* -axis units are arbitrary, taken from Reference [173]).

In the pseudo-potential method, the valence wave function is written as a linear combination of orthogonal plane waves (OPWs) [178],

$$\left|\psi_k{}^v\right\rangle = \sum_K c_{\vec{K}}^{\vec{k}} e^{i(\vec{k}+\vec{K})\cdot\vec{r}} - \sum_{core}\left\langle\psi_k{}^c\middle|\psi_k{}^v\right\rangle\left|\psi_k{}^c\right\rangle. \qquad (4.4.1.3)$$

The pseudo-potential is the sum of the periodic potential V and U_R, where

$$U_R|\psi\rangle = \sum (e_k^{\,v} - e_c)\langle \psi_k^{\,c} | \psi \rangle | \psi_k^{\,c} \rangle, \tag{4.4.1.4}$$

and is part of an effective Hamiltonian

$$H + U_R = \frac{-\hbar^2}{2m}\nabla^2 + U_{pseudo}. \tag{4.4.1.5}$$

When using the pseudo-potential method, it is hoped that U_{pseudo} is small enough that nearly free electron theory can be used to find the valence levels of the effective Hamiltonian. Wavefunctions oscillate most rapidly near the nuclei. The electrons in the regions near the nuclei, or the core regions are shielded from the valence regions of the atoms, where chemical combinations occur, and essentially behave like electrons in free isolated atoms. In the pseudo-potential method the true valence wave functions in the core region are approximated by wave functions with no radial nodes in the core region. The true ionic potential in the core is approximated by a pseudo-potential which is chosen such that the phase shifts imposed on the pseudo valence wave functions by the pseudo-potential are identical to the phase shifts imposed on the true valence wave functions by the true ionic potential. In the outer regions of the atoms, the pseudo-potential gradually morphs into the true potential. This approximation makes it possible to represent the valence wave functions with a finite linear combination of plane waves. It is then possible to use a cut-off energy $E_{cut-off}$ of 20Ry or 272eV corresponding to a $K_{\max}$ of 4.5 bohr^{-1}, which for the example previously considered, allows for a much more manageable basis set of approximately 270 plane waves.

However because the core electrons have been removed from the problem in the pseudo-potential approximation, the total energy is not being computed. Only energy differences between different ionic configurations in a solid are meaningful when using the pseudo-potential approach. Operating on the wave function with the potential energy operator in the core region results in the unique potential energy eigenvalue V_l acting on a particular angular momentum component of the wave function. The valence wave functions in the core region are therefore eigenfunctions of the potential energy operator in that region.

The pseudo-potential approximation corresponds to using the same potential energy eigenvalue for each angular momentum component of the wave functions in the core region. Pseudo-potentials are termed soft and ultra-soft when they lead to only a few and even fewer plane waves respectively in the basis set. A soft pseudo-potential will probably work very well only for a specific problems, whereas transferable pseudo-potentials can be used with very good results on a wide variety of problems representing molecules, clusters, surfaces, insulators, conductors and other kinds of systems.

4.4.2 The Tight Binding Method

The tight binding method is the other extreme compared to the free electron approximation. Near each lattice point of a crystal, the atomic Hamiltonian, H_{atom}, at that point, approximates the total periodic crystal Hamiltonian. The crystal Hamiltonian can then be expressed as $H_{atom} + \Delta V$ with ΔV going to zero near an atomic site in the crystal lattice. The atomic wave functions, which are linear combinations of Bloch functions, are also assumed to be localized around the atom with a range close to the lattice constant of the crystal. The wave function $|\phi\rangle$, which is introduced to correct for

the difference between the atomic Hamiltonian and the full periodic crystal Hamiltonian far from atomic sites is then equal to

$$|\phi\rangle = \sum_i b_i |\psi_i\rangle, \tag{4.4.2.1}$$

where the $|\psi_i\rangle$ are bound atomic levels. The crystal Schrödinger equation is

$$H|\breve{\psi}\rangle = (H_{atom} + \Delta V)|\breve{\psi}\rangle. \tag{4.4.2.2}$$

The wavefunctions that solve the Schrödinger equation for the crystal (the crystal eigenfunctions) are then of the form

$$|\breve{\psi}\rangle = \psi(r) = \sum_R e^{ik\cdot R}\phi(r-R), \tag{4.4.2.3}$$

with eigenvalues $\varepsilon(k)$. If the eigenvalues of the atomic Hamiltonian H_{atom} are labeled as E^i_{atom}, then

$$(\varepsilon(k) - E^i_{atom})\langle\psi_j|\breve{\psi}\rangle = \langle\psi_j|\Delta V|\breve{\psi}\rangle. \tag{4.4.2.4}$$

Substituting into this the expanded forms of $|\breve{\psi}\rangle$ and $|\phi\rangle$, the expression

$$(\varepsilon(\vec{k}) - E^i{}_{atom})b_j = -(\varepsilon(\vec{k}) - E^i_{atom})\times$$

$$\sum_i b_i\left(\sum_R\langle\psi_j|\psi_i(r-R)|e^{ik\cdot R}\rangle\right) + \sum_i b_i\langle\psi_j|\Delta V|\psi_i\rangle + \sum_i b_i\left(\sum_R\langle\psi_j|\Delta V|\psi_i(r-R)e^{ik\cdot R}\rangle\right), \tag{4.4.2.5}$$

is obtained. The matrix element $\langle \psi_j | O | \psi_i \rangle$ is zero for any operator O when $|\psi_i\rangle$ and $|\psi_j\rangle$ are different eigenfunctions of the same atom (from orthogonality) and 1 when they are one and the same eigenfunction. This matrix element is very small when $|\psi_i\rangle$ and $|\psi_j\rangle$ are eigenfunctions of different atoms (because of the extremely small overlap between wavefunctions corresponding to tightly bound levels centered on different atoms). $E^i{}_{atom} - \varepsilon(k) \approx 0$ when b_i is large. The above expression is then summed over the levels i that are close to the energy to get the band energy. The width of the energy band depends on the overlap integral representing the degree of overlap between different atomic wavefunctions. The bandwidth decreases as the degree of overlap decreases and in the case of isolated atoms, there is no bandwidth leading to N-fold degeneracy where N is the number of atoms. As long as there is any overlap, an electron in an atomic level has a nonzero velocity and is not tied to any particular atom. This means that $\varepsilon(k)$ is not constant in the zone and that the electron can move throughout the cells of the crystal with a velocity $\partial(\varepsilon(k))/\hbar\partial k$.

Wavefunctions corresponding to higher energy atomic levels have a greater range in space and therefore lead to wider bands. Lower lying atomic levels result in narrow bands and higher lying levels lead to broader bands. Since the binding energy decreases as the energy of the levels increase, the tight binding approximation is less accurate for higher-lying energy levels. The free electron method also does not work for the higher lying (valence) levels because the potential experienced by these levels is much larger than the weak periodic potential assumed in the free electron treatment. Wavefunctions of core levels oscillate rapidly near ionic sites because of the high kinetic energy of the

electrons near the ions and also the high potential there. The wavefunctions go to zero in the interstitial regions between ions because being core (lower-lying) levels, they are of the tight binding type. The valence wavefunctions however, lying higher in energy than the core wavefunctions, oscillate even more rapidly near the ions (core region) because of their higher kinetic energy. The valence wave functions also have a larger range in space and therefore do not go to zero in the interstitial region.

4.4.3 The Muffin-Tin Potential Method

The muffin-tin potential was introduced to address the fact that the potential used in the cellular method does not have a continuous derivative at the cell boundaries and also in order to avoid the difficulties involved in numerically satisfying the boundary conditions of continuity of the wavefunctions and the first derivatives of the wavefunctions over the usually complex Wigner-Seitz primitive cell surface. The muffin-tin potential is [178]

$$\begin{aligned} V(r) &= U(|r-R|) \qquad |r-R| < r_c \\ &= U(r_c) = 0 \qquad |r-R| > r_c, \end{aligned} \tag{4.4.3.1}$$

where r_c is the radius of the core region. The methods used to compute band structure in a muffin-tin potential are the augmented plane wave (APW) method and the Korringa, Kohn and Rostock (KKR) method.

4.4.4 The Orthogonalized Plane Wave (OPW) Method

This method allows some freedom in choosing a potential because it does not require a muffin-tin potential. The core wave functions are localized around atomic sites and the valence wave functions have a finite probability of being found in the interstitial regions.

The OPW method approximates the valence wave functions in the interstitial region by a finite number of plane waves and reproduces the rapid oscillations of the valence wave functions in the core region by requiring them to be orthogonal to the core wave functions. An OPW has the form [178]

$$|\phi_k\rangle = e^{ik\cdot r} + \sum_{core} a_i |\psi_k{}^i\rangle, \tag{4.4.4.1}$$

where the $|\psi^i{}_k\rangle$ are the core levels. Since the OPWs are required to be orthogonal to each of the core levels,

$$\langle \psi_k{}^i | \phi_k \rangle = \langle \psi_k{}^i | e^{ik\cdot r} + a_i = 0, \tag{4.4.4.2}$$

and

$$a_i = -\langle \psi_k{}^i | e^{ik\cdot r}. \tag{4.4.4.3}$$

The eigenvalues of the crystal Hamiltonian can be expanded in a linear combination of OPWs

$$\psi_k = \sum_K c_K |\phi_{k+K}\rangle. \tag{4.4.4.4}$$

The coefficients c_K are calculated by the variational principle. The matrix elements $\langle \phi_{k+K} | V(r) | \phi_{k+K} \rangle$, where $V(r)$ is the crystal potential, are smaller in magnitude than the corresponding plane wave matrix elements ensuring quick convergence. The OPW method can be used by carrying out a first principles calculation and working with a secular problem large enough to ensure adequate convergence or by fitting the basis set

expansion coefficients $u_{\vec{K}}$ to experimental data and then carrying out a nearly free electron calculation using OPW basis functions.

4.4.5 The Augmented Plane Wave (APW) Method

In the augmented plane wave (APW) method [173 – 174], the basis set is composed of plane waves in regions far away from the nuclei, and is composed of atomic-like functions in the core regions near the nuclei. The space is now divided into two regions, around each atom α, a sphere, called the muffin-tin sphere, with radius R_α is created, inside which the basis functions are atomic-like, and outside the atomic spheres where the eigenfunctions are more slowly varying, the basis functions are expanded in plane waves. The part of space occupied by the spheres is referred to as the muffin-tin region (region I in Figure 4.3) while the rest of the space outside the spheres is called the interstitial region I (region II in Figure 4.3).

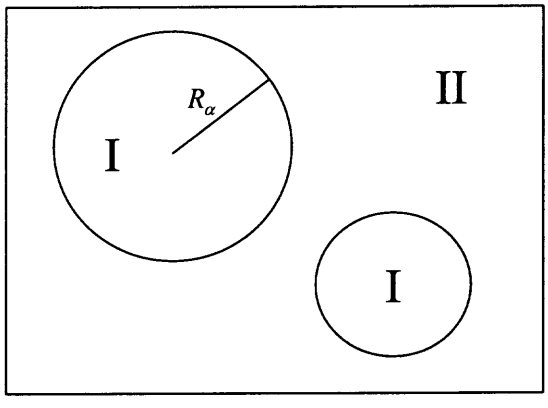


Figure 4.3 Division of the APW unit cell into atomic sphere regions (I) and the interstitial region (II).

An augmented plane wave (APW) $\left|\phi_{\vec{K}}^{n,\vec{k}}\right\rangle$ is not an eigenfunction of the crystal Hamiltonian, but in the interstitial region the eigenfunctions of the crystal Hamiltonian are a linear combination of these APWs, which are plane waves of the form $e^{i\vec{k}\cdot r}$. The APW but not its derivative, is continuous from the interstitial region to the boundary of the atomic core region at R. Inside the core region, the APW is an eigenfunction of the atomic Hamiltonian [173 – 174]

$$\frac{-\hbar^2}{2m}\nabla^2\left|\phi_{\vec{K}}^{n,\vec{k}}\right\rangle + U(|r-R|)\left|\phi_{\vec{K}}^{n,\vec{k}}\right\rangle = \varepsilon\left|\phi_{\vec{K}}^{n,\vec{k}}\right\rangle \qquad |r-R| < r_c . \tag{4.4.5.1}$$

The expansion of the crystal eigenfunctions in terms of APWs of the same energy is [173 – 174]

$$\left|\psi_k\right\rangle = \sum_K c_K \left|\phi_{\vec{K},E}^{n,\vec{k}}\right\rangle, \tag{4.4.5.2}$$

where the APWs are of the form [173 – 174]

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r},E) = \frac{1}{\sqrt{V}} e^{i(\vec{k}+\vec{K})\cdot\vec{r}} \qquad \vec{r}\in I , \tag{4.4.5.3}$$

and

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r},E) = \sum_{l,m} A_{l,m}^{\alpha,\vec{k}+\vec{K}} u_l^{\alpha}(r',E) Y_{l,m}(\theta',\phi') \qquad \vec{r}\in S_\alpha , \tag{4.4.5.4}$$

and V is the unit cell volume, $\vec{r}'$ is the position vector of the sphere center, measured from the origin of the direct lattice, $r' = |\vec{r}'| = |\vec{r} - \vec{r}_\alpha|$ is the distance inside the spheres

with respect to the center of each sphere, the angles θ' and ϕ' specify the direction of $\vec{r}'$ in spherical coordinates and the $Y_{l,m}$ are spherical harmonics. E and $A_{l,m}^{\alpha,\vec{k}+\vec{K}}$ are undetermined parameters, while the functions u_l^{α} are solutions to the radial atomic Schrödinger equation at the energy E. The plane wave basis functions in the interstitial region must be matched to the basis functions in the spheres to avoid discontinuities in the eigenfunctions. The matching of the functions over the entire surface of the atomic spheres occurs by first expanding the plane wave in spherical harmonics about the origin of the atomic sphere α [173 – 174],

$$\frac{1}{\sqrt{V}}e^{i(\vec{k}+\vec{K})\cdot\vec{r}} = \frac{4\pi}{\sqrt{V}}e^{i(\vec{k}+\vec{K})\cdot\vec{r}_{\alpha}}\sum_{l,m} i^l j_l\left(\left|\vec{k}+\vec{K}\right|\left|\vec{r}'\right|\right)Y_{l,m}^{*}\left(\theta_{\vec{k}+\vec{K}},\phi_{\vec{k}+\vec{K}}\right)Y_{l,m}\left(\theta_{\vec{r}},\phi_{\vec{r}}\right). \tag{4.4.5.5}$$

Recognizing that $j_l(x)$ is the Bessel function of order l and requiring this Bessel function to be equal to the (l,m) dependent part of Equation (4.4.5.4) on the atomic sphere surface results in the following expression,

$$A_{l,m}^{\alpha,\vec{k}+\vec{K}} = \frac{4\pi i^l e^{i(\vec{k}+\vec{K})\cdot\vec{r}_{\alpha}}}{\sqrt{V}u_l^{\alpha}(\vec{R}_{\alpha},E)}j_l\left(\left|\vec{k}+\vec{K}\right|R_{\alpha}\right)Y_{l,m}^{*}\left(\theta_{\vec{k}+\vec{K}},\phi_{\vec{k}+\vec{K}}\right). \tag{4.4.5.6}$$

The expansion of plane waves in spherical harmonics has to be truncated at some maximum l value $l_{\max}$ for practical computational reasons. $l_{\max}$ is determined by realizing that for any given $l_{\max}$, $Y_{l_{\max},m}(\theta,\phi)$ can have at most $2l_{\max}$ nodes along any great circle created by fixing the angle ϕ arbitrarily and allowing the other angle θ to range from 0 to 2π [173]. The corresponding number of nodes per unit length is then

given by $2l_{max}/2\pi R_\alpha$ or $l_{max}/\pi R_\alpha$. A plane wave that would match satisfactorily with the spherical harmonics on the sphere surface would have to have a comparable number of nodes per unit length. Being that the shortest period plane wave has K_{max}/π nodes per unit length, the condition $R_\alpha K_{max} = l_{max}$ arises for acceptable matching of the plane waves to the spherical harmonic functions on the sphere boundaries. The cutoff in the number of APWs in the basis set is determined by the degree of change of the $\varepsilon(k)$ when more APWs are added to the basis.

Another parameter E is needed in order to use the APW basis set. The APW basis describes an eigenfunction best when the parameter E is equal to the eigenvalue, which is not known *a priori*, of the eigenfunction. Therefore in order to use APW basis sets, the eigenvalue is guessed and used as the initial value of E, the APWs are determined, and the Hamiltonian matrix elements as well as the overlap matrix elements are constructed, using the APW expansion of the eigenfunction ψ in the variational method to obtain the secular equation. Setting the secular determinant equal to zero results in equations from whose roots the energies $\varepsilon(\vec{k})$ can be obtained. The secular determinant is the determinant of the coefficients of the equations obtained by the condition that $E(\psi)$ is independent of the coefficients c_K. The secular equation can then be solved to check if the guessed E is contained among its roots. If not, the whole process is repeated again until the guessed E is among the roots of the secular equation, at which point the problem is considered solved. The procedure is equivalent to making a functional [173, 178]

$$E(\psi) = \frac{\int \left(\frac{\hbar^2}{2m} |\nabla \psi|^2 + V(r) |\psi|^2 \right) dr}{\int |\psi|^2 \, dr}, \tag{4.4.5.7}$$

constant with respect to ψ where ψ is differentiable and satisfies the Bloch condition with wave vector $\vec{k}$. The constant value of the functional is then equal to $\varepsilon(\vec{k})$.

It is customary to insert an expansion using a basis set of APWs with a certain energy into the energy functional and then the values of k for which the secular determinant is zero are sought to map out constant energy surfaces in k-space. Usually a K_{max} value of 3.5 bohr^{-1} is required for good results, compared to K_{max} values of 5.5 for plane waves and pseudo-potentials, so that the basis set size J is about 130 for the APW method, and about 270 for plane waves [173]. The time of calculation is mostly spent in matrix diagonalization, and scales as the basis set size cubed suggesting that the APW method would be a factor of 10 faster than the pseudo-potential method. However, with a plane wave basis set, J eigenvalues are found with each diagonalization step, while with the APW method, a separate diagonalization step is required for each eigenvalue, so that the APW method is much slower than the pseudo-potential method.

4.4.6 The Linearized Augmented Plane Wave (LAPW) Method

In the APW method, the functions $u_l^{\alpha}(r', E)$ are calculated with estimated parameters E. The linearized augmented plane wave (LAPW) method enables immediate calculation of the $u_l^{\alpha}(r', E)$ from known quantities. $u_l^{\alpha}(r', E)$ at an energy E is calculated in terms of $u_l^{\alpha}(r', E_0)$ which has already been calculated at an energy E_0 close to E using a Taylor expansion about E_0 [173 –174]

$$u_l^\alpha(r',\varepsilon_{n,\vec{k}}) = u_l^\alpha(r',E_0) + \left.\frac{\partial u_l^\alpha(r',E)}{\partial E}\right|_{E=E_0} (E_0 - \varepsilon_{n,\vec{k}}) + O(E_0 - \varepsilon_{n,\vec{k}})^2, \tag{4.4.6.1}$$

with

$$\left.\frac{\partial u_l^\alpha(r',E)}{\partial E}\right|_{E=E_0} = \dot{u}_l^\alpha(r',E_0). \tag{4.4.6.2}$$

For a fixed E_0, substituting the first two terms in the Taylor expansion in the APW basis function results in the LAPW basis function. The energy difference $(E_0 - \varepsilon_{n,\vec{k}})$ is however unknown, which introduces a new coefficient into the calculation. The LAPW basis functions are defined as [173 – 174]

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r}) = \frac{1}{\sqrt{V}} e^{i(\vec{k}+\vec{K})\cdot\vec{r}} \qquad \vec{r} \in I, \tag{4.4.6.3}$$

and

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r}) = \sum_{l,m} \left(A_{l,m}^{\alpha,\vec{k}+\vec{K}} u_l^\alpha(r',E_0) + B_{l,m}^{\alpha,\vec{k}+\vec{K}} \dot{u}_l^\alpha(r',E_0)\right) Y_{l,m}(\theta',\phi') \qquad \vec{r} \in S_\alpha. \tag{4.4.6.4}$$

The LAPW basis functions therefore have more variational freedom inside the spheres than the APW basis functions. The coefficients $A_{l,m}^{\alpha,\vec{k}+\vec{K}}$ and $B_{l,m}^{\alpha,\vec{k}+\vec{K}}$ are determined by requiring that the LAPW basis functions used inside the spheres match the plane waves used in the interstitial region in both value and slope at the sphere boundaries. When it is desired to describe an eigenstate having mostly p character for example [173], the

$A_{l=1,m}^{\alpha,\vec{k}+\vec{K}}$ coefficients will be large relative to the other coefficients. Choosing E_0 near the center of the p-band ensures that the neglected terms $O(E_0 - \varepsilon_{n,\vec{k}})^2$ in the Taylor expansion remain small. This procedure is necessary for other important states up to $l = 3$ as well as for every inequivalent atom. A set of parameters E_l^α with l up to 3 are therefore usually chosen in the LAPW method, rather than a single parameter E_0. The final form of the LAPW basis functions inside the atomic spheres is then [173 – 174]

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r}) = \sum_{l,m} \left(A_{l,m}^{\alpha,\vec{k}+\vec{K}} u_l^\alpha(r', E_l^\alpha) + B_{l,m}^{\alpha,\vec{k}+\vec{K}} u_l^\alpha(r', E_l^\alpha) \right) Y_{l,m}(\theta', \phi') \qquad \vec{r} \in S_\alpha . \tag{4.4.6.5}$$

In the LAPW method, accurate energy bands at a given point can be obtained with a single matrix diagonalization step, while a diagonalization step is required for each energy band in the APW method [173]. The product of the smallest muffin-tin radius $R_\alpha^{\min}$ and $K_{\max}$, $R_\alpha^{\min} K_{\max}$ is used to determine the accuracy of the LAPW basis set because if $R_\alpha^{\min}$ is increased, then $K_{\max}$ should be reduced because it determines the size of the basis set. The increase in $R_\alpha^{\min}$ increases the closest distance that a plane wave can approach the nucleus. The part of the wave function closest to the nucleus exhibits the greatest oscillations and increasing $R_\alpha^{\min}$ removes the necessity of describing this highly oscillatory part with plane waves and so $K_{\max}$ can be reduced accordingly because less plane waves will be needed to describe the remaining smoother parts of the wavefunction. The increase of $R_\alpha^{\min}$ therefore reduces matrix diagonalization time. The LAPW basis set is much smaller than the plane wave pseudo-potential basis set because

$R_\alpha^{\min} K_{\max}$ values needed for reasonable accuracy range from 7.5 – 9, giving a value of approximately 4 bohr^{-1} for $K_{\max}$ and resulting in a basis set size of approximately 195 compared to approximately 270 for the plane wave pseudo-potential method [173]. The non-orthogonal LAPW basis set slows down LAPW calculations compared to an all-plane wave basis set calculation so that in the final analysis, the LAPW method is comparable in speed to the plane wave pseudo-potential method [173].

In the LAPW method, core states that are well bound to the nucleus are contained entirely in the muffin-tin sphere and do not participate directly in chemical bonding with other atoms. States that leak from the muffin-tin sphere are called valence states and are treated by LAPW while core states are treated like free atom states which are subject to the potential from the valence states. Sometimes states having the same angular momentum quantum number l but different values of the principal quantum number n are both valence states, and it is not clear how the E_l^α parameter should be chosen. For example [173] bcc Fe has a significant amount of 4 p -character in its valence states about 0.2 Ry below the Fermi level and also has 3 p states that lie 4.3 Ry below the Fermi level which leak from the core region. States such as these, which lie far below the Fermi energy and leak from the core, are called semi-core states. Another sort of basis function, called a local orbital is employed to solve this problem. A local orbital (LO) is defined for a particular (l,m) combination and for a particular atom α' as [173 – 174]

$$\phi_\alpha^{l,m}(\vec{r}) = 0 \qquad \vec{r} \notin S_{\alpha'}, \tag{4.4.6.6}$$

and

$$\phi_{\alpha}^{l,m}(\vec{r}) = \left(A_{l,m}^{\alpha',LO} u_l^{\alpha'}(r', E_{1,l}^{\alpha'}) + B_{l,m}^{\alpha',LO} \dot{u}_l^{\alpha'}(r', E_{1,l}^{\alpha'}) + C_{l,m}^{\alpha',LO} u_l^{\alpha'}(r', E_{2,l}^{\alpha'})\right) Y_{l,m}(\theta', \phi')$$

$$\vec{r} \in S_{\alpha'}.$$

(4.4.6.7)

The local orbitals are zero in the interstitial region and in the muffin-tin spheres around other atoms. A linearization energy $E_{1,l}^{\alpha'}$ [173 – 174] suitable for the higher of the two valence states is used, along with the same functions $u_l^{\alpha'}(r', E_{1,l}^{\alpha'})$ and $\dot{u}_l^{\alpha'}(r', E_{1,l}^{\alpha'})$ used in the LAPW basis set. The lower valence state, which will have more free atomic character, will be sharply peaked at an energy $E_{2,l}^{\alpha'}$ making it amenable to being described by a single radial function $\dot{u}_l^{\alpha'}(r', E_{2,l}^{\alpha'})$ at the peak energy $E_{2,l}^{\alpha'}$. Local orbitals, having no connection to plane waves in the interstitial region, have no $\vec{k}$ or $\vec{K}$ dependence. The coefficients $A_{l,m}^{\alpha',LO}$, $B_{l,m}^{\alpha',LO}$ and $C_{l,m}^{\alpha',LO}$ are determined by requiring the local orbital to have zero value and slope at the muffin-tin sphere boundary, and also to be normalized [173 – 174]. Local orbitals result in much better calculational accuracy of eigenstates in materials having semi-core states but their addition slightly increases the LAPW basis set size. For example, if d local orbitals are added for each atom in a calculation, the basis set will increase by $(2l+1)$ or 5 functions, which is only a small fraction of the typical LAPW basis set size of a few hundred functions. The LAPW method does not handle valence d and f states easily. The method also has some difficulty calculating atomic states when the atom has a muffin-tin sphere radius much smaller than that of the spheres corresponding to the other atoms in the unit cell.

4.4.7 The APW+lo Method

In order to avoid energy dependence of the basis set and still maintain small basis set sizes in the APW method, the best features of the APW and the LAPW+LO methodologies are combined to obtain the APW+lo method, in which the basis set contains two kinds of functions [173]

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r},E)=\frac{1}{\sqrt{V}}e^{i(\vec{k}+\vec{K})\cdot\vec{r}} \qquad \vec{r}\in I\,, \tag{4.4.7.1}$$

and

$$\phi_{\vec{K}}^{n,\vec{k}}(\vec{r},E)=\sum_{l,m}A_{l,m}^{\alpha,\vec{k}+\vec{K}}u_l^{\alpha}(r',E_l^{\alpha})Y_{l,m}(\theta',\phi') \qquad \vec{r}\in S_{\alpha}\,, \tag{4.4.7.2}$$

and a different kind of local orbital to those used in the LAPW method

$$\phi_{l,m}^{\alpha',lo}(\vec{r},E)=0 \qquad \vec{r}\notin S_{\alpha'}\,, \tag{4.4.7.3}$$

and

$$\phi_{l,m}^{\alpha',lo}(\vec{r},E)=\left(A_{l,m}^{\alpha',lo}u_l^{\alpha'}(r',E_l^{\alpha'})+B_{l,m}^{\alpha',lo}\dot{u}_l^{\alpha'}(r',E_l^{\alpha'})\right)Y_{l,m}(\theta',\phi')\,. \qquad \vec{r}\in S_{\alpha'}\,. \tag{4.4.7.4}$$

The two coefficients $A_{l,m}^{\alpha',lo}$ and $B_{l,m}^{\alpha',lo}$ are determined by normalization and by requiring the value of the local orbitals to be zero at the muffin-tin sphere boundary to be zero, where the APWs and the local orbitals are continuous, but their first derivatives are not.

Just as in the LAPW method, eigenvalues are obtained from a single diagonalization step in the APW+lo method, while the basis set size is comparable in size to that of the APW method, and smaller than that of the LAPW+LO method [173].

The APW+lo method is very useful for treating valence d and f states, which are not easily handled by the LAPW method, and for calculating atomic states where the atom has a muffin-tin sphere radius much smaller than that of the other spheres in the unit cell. The most accurate calculations therefore use a mixed LAPW and APW+lo basis set.

4.5 Force Minimization in the LAPW Formalism

According to the Hellman-Feynman (HF) theorem [179 – 180], the force on an atom is equivalent to the electrostatic force experienced by its nucleus. An electronic-structure calculation can be used to accurately calculate the force by means of the change in total energy due to atomic movements. The force on atom i due to displacement by a small amount is given by [179]

$$F_i = -\frac{\delta E}{\delta \tau_i}, \tag{4.5.1}$$

which decomposes to

$$F_i = F_i^{HF} - \frac{1}{\delta \tau_i}\left(\sum f_n \delta \varepsilon_n - \int \rho(\vec{r}) \delta V^{eff}(\vec{r}) d\vec{r}\right), \tag{4.5.2}$$

where the first term in Equation (4.5.2) is the HF force [179], represented as

$$F_i^{HF} = Z_i \frac{d}{d\tau_i}\left(\int \frac{\rho(\vec{r})}{\tau_i - \vec{r}} d\vec{r} - \sum_j \sum_{\vec{R}} \frac{Z_j}{\left|\tau_i - \tau_j + \vec{R}\right|} \right). \tag{4.5.3}$$

The variation in the eigenvalues due to atomic movement depends on the method by which the Kohn-Sham equation is solved. The derivations have been worked out for the all-electron method in cases where the radial augmentation functions are frozen and unfrozen. After evaluating these quantities, the total force on atom i is [179 - 180]

$$F_i = F_i^{HF} + F_i^{core} + F_i^{IBS}, \tag{4.5.4}$$

where [179 - 180]

$$F_i^{core} = -\frac{1}{\delta\tau_i}\left(\sum f_n \delta\varepsilon_n - \int \rho_i^{core}(\vec{r}) \delta V^{eff}(\vec{r}) d\vec{r} \right), \tag{4.5.5}$$

and [179 - 180]

$$F_i^{IBS} = -\sum_n f_n \left[\langle \delta\psi_n / \delta\tau_i | (H - \varepsilon_n) | \psi_n \rangle + \langle \psi_n | (H - \varepsilon_n) | \delta\psi_n / \delta\tau_i \rangle + D_n \right], \tag{4.5.6}$$

with

$$D_n = \oint \left(\psi_n^*(\vec{r}) \overline{T} \psi_n(\vec{r}) \Big|_{MT} - \psi_n^*(\vec{r}) \overline{T} \psi_n(\vec{r}) \Big|_I \right) dS_n. \tag{4.5.7}$$

The basis functions $\phi_n(r)$ used in the LAPW method have discontinuous second derivatives (and therefore discontinuous single particle kinetic energy) [173 - 174, 180] across the muffin-tin sphere boundaries, which results in a large surface contribution to the incomplete basis set corrections to the Hellmann-Feynman force. Basically the matrix

elements $\langle \delta\varphi_n / \delta\tau | (H - \varepsilon_n) | \varphi_n \rangle$ and $\langle \varphi_n | \delta\overline{T} / \delta\tau_i | \varphi_n \rangle$ must be accounted for in the total correction to the Hellmann-Feynman force. The F_i^{IBS} term in the expression for the total force on atom i can be decomposed [179 – 180] into Pulay contributions from semi-core and valence electrons

$$F_i^{HF} = Z_i \sum_{m=-l}^{l} \lim_{\tau_i \to 0} \frac{V_{lm,i}(\tau_i)}{\tau_i} \frac{\partial(\tau_i Y_{lm}(\tau_i'))}{\partial \tau_i}, \tag{4.5.8}$$

$$F_i^{core} = -\int d\vec{r} \rho^{core}(\vec{r}) \frac{\partial V^{eff}}{\partial \vec{r}}, \tag{4.5.9}$$

and

$$F_i^{val} = \int d\vec{r} V^{eff} \frac{\partial \rho^{val}(r)}{\partial r}$$
$$+ \sum_{\vec{k},n} f_{k,n} \sum_{\vec{K},\vec{K}'} c_n^*(\vec{K}') c_n(\vec{K}) \left\{ (K^2 - \varepsilon_n) \int dS \phi_{\vec{K}'}^*(\vec{r}) \phi_{\vec{K}}(\vec{r}) - n(\vec{K} - \vec{K}') \langle \phi_{\vec{K}'} | H - \varepsilon_n | \phi_{\vec{K}} \rangle_i \right\}. \tag{4.5.10}$$

The incomplete basis set corrections to the Hellmann-Feynman force arise because of the practical limitations of having to use an incomplete basis set in any electronic structure calculation. The LAPW basis functions are centered at the positions of the nuclei and therefore they move with the atoms. The corrections vanish if the basis functions are chosen to be atomic position independent. The formulation has been applied successfully to frozen-phonon calculations in semiconductor, transition metal and noble

metal systems [179]. The forces calculated in this manner are used in a minimization scheme to optimize the geometry of systems containing many atoms [180].

4.6 Minimization of Atomic Forces

The total energy $E_{tot}(\tau)$ of a system with N independent atoms is minimized with respect to a position vector $\tau = (\tau_1, \tau_2, ..., \tau_N)$ which is $3N$ dimensional using the directly calculated forces on the atoms [179 – 180]

$$F = -\frac{\delta E}{\delta \tau}. \tag{4.6.1}$$

The previous force minimization history is taken into account in order to minimize oscillations in the computed forces. The various formalisms for carrying out force minimizations on a system of atoms are described in the following sections.

4.6.1 The Variable Metric Method

The variable metric or quasi-Newton method minimizes forces efficiently if the total energy surface close to a certain atomic geometry can be represented by an expression with quadratic terms as the highest order terms, meaning that the energy surface is close to an energy minimum [180]

$$E_{Tot}(r) = E_{Tot}(r') - F(r')(r - r') + \frac{1}{2}(r - r')\mathrm{H}(r')(r - r'), \tag{4.6.1.1}$$

where $\mathrm{H}(r')$ is the Hessian matrix at the geometry r'. The force is found by taking the derivative of the expression for the total energy [180]

$$F(r) = F(r') - \mathrm{H}(r')(r - r') . \tag{4.6.1.2}$$

This force is zero at the minimum energy, therefore

$$\Delta r' = r - r' = \mathrm{H}^{-1}(r')F(r') . \tag{4.6.1.3}$$

In the event that the inverse Hessian $\mathrm{H}^{-1}(r')$ and the quadratic approximation to the total energy $E_{Tot}(r)$ are exact, then $\Delta r'$ points in the direction of the minimum energy. However, this is usually not the case, therefore an exact inverse Hessian is not available and if higher order terms are significant in the expression for the energy surface, then the quadratic approximation is no longer valid and $\Delta r'$ will not necessarily point in the direction of steepest descent. An algorithm due to Broyden, Fletcher, Goldfarb and Shanno (BFGS) [181 – 183] can be used to overcome these difficulties. In this method, which has been implemented as a FORTRAN program [184], an approximation to $\mathrm{H}^{-1}(r')$ is iteratively built up using computed forces from previous structure optimization steps. The matrix is constrained to be positive definite and symmetric throughout, so that the total energy $E_{Tot}(r)$ will initially decrease as the direction $\Delta r'$ is approached. If the attempted step causes an increase in $E_{Tot}(r)$, meaning that $\Delta r'$ is too large, backtracking along the same direction using smaller steps is carried out in order to reduce the energy. When all atomic forces are below a certain predetermined value, the minimization process stops.

4.6.2 Damped Newton Dynamics

If the total energy of a system is not close to an energy minimum, then a quadratic approximation to the energy surface is not justified. In these kinds of cases, a damped Newton dynamics based algorithm is more efficient. The time evolution of an atomic coordinate R_m is described by the expression [180],

$$R_m{}^{\tau+1} = R_m{}^{\tau} + \eta_m (R_m{}^{\tau} - R_m{}^{\tau+1}) + \delta_m F_m{}^{\tau}, \tag{4.6.2.1}$$

where $R_m{}^{\tau}$ and $F_m{}^{\tau}$ are the coordinate and force respectively at time τ. The two parameters η_m and δ_m control the damping and speed of motion. When these parameters are chosen optimally, a fast movement towards the nearest local minimum on the Born-Oppenheimer surface is performed, with a minimum of oscillations about this minimum. A small damping factor η_m improves the stability of the atomic relaxation [180]. A large damping factor makes it possible to overcome energy barriers and escape from local minima. The structure relaxation continues until all components of the force are below a certain predetermined value. If η_m is one, there is no damping and this approach becomes equivalent to an ab-initio molecular dynamics method.

4.7 Force Calculation Method

The objective of the force minimization calculation is to optimize the structural parameters of the beta tantalum unit cell by calculating the forces on the atomic nuclei, which will move accordingly in order to obtain the equilibrium positions of each atom in the unit cell. Either of the damped Newton dynamics [180] or Broyden-Fletcher-

Goldfarb-Shanno (BFGS) [181 – 184] variable metric methods previously described may be used. The force minimization calculations therefore consist of moving the atoms with free structural parameters by amounts specified in input and initialization files. After each such movement of the atomic positions, a self-consistent solution of the Schrödinger equation for that particular structural configuration resulting from the movements must be carried out. This step of the calculation is referred to as a self-consistent field (SCF) cycle.

The SCF cycle calculation is complete when the final potential produced by the calculation, when used as input in solving the Schrödinger equation for the system, produces wave functions, which generate the charges, which in turn generate the potential used as input according to

$$\rho(r) = \sum_{n=1}^{\infty} f_n \psi_n^*(r) \psi_n(r) . \tag{4.7.1}$$

A typical criterion used to terminate the SCF cycle is that the charge densities computed as shown above differ by less than 10^{-4} electrons/a_0^3 between iterations inside the cycle. The forces on the atoms in an atomic arrangement are computed when the SCF cycle has converged for that particular atomic arrangement. The force minimization calculation is terminated when the forces calculated between iterations differ by amounts less than certain predetermined values. A sample input file is shown in Figure 4.4. Each line in the file in Figure 4.4 after the first line corresponds to a unique atomic symmetry position.

```
 NEWT 2.0              #(NOSE, NEWT, BFGS, MOLD, tolf (a4,f5.2))
0   0   0   0.8     # Atom1   (NOSE, MOLD:Masse, delta t, T,nose
1.5 1.5 1.5 0.8     # Atom2   (NEWT: 1,2,3:delta, 4:eta(1=MOLD))
1.5 1.5 1.5 0.8     # Atom3   (NEWT: 1,2,3:delta, 4:eta(1=MOLD))
1.5 1.5 1.5 0.8     # Atom4   (NEWT: 1,2,3:delta, 4:eta(1=MOLD))
1.5 1.5 1.5 0.8     # Atom5   (NEWT: 1,2,3:delta, 4:eta(1=MOLD))
```

Figure 4.4 Sample input file for force minimization calculation.

The first line indicates that the Newton force minimization scheme is being used and that the force minimization will stop when the differences between all calculated forces are below 2 mRy/a.u between iterations. The line after the first indicates that the atomic position represented there is fixed by symmetry and has no adjustable structural parameters. The parameter δ_m, which controls the speed of motion of the atoms, is set to 1.5 and the damping parameter η_m is set to 0.8.

4.8 Spin-Orbit Coupling in the LAPW Formalism

Relativistic corrections become important in electronic structure calculations when (a) the electronic velocities are of the same order as the speed of light and (b) phenomena are observed applying the corrections that would not be observed otherwise. Relativistic corrections are important in the atomic spheres but are not important in the interstitial region of the unit cell because the energy cutoff limits the electron velocity there. They are important when effects such as magnetocrystalline anisotropy in which the dominant contributing effect is the combined effects of spin-orbit interactions and the crystal field. The following description of spin-orbit coupling is taken entirely from Reference [185].

The Dirac equation includes all relativistic effects and is given by

$$H_D = \vec{\alpha}\vec{p}c + (\beta - 1)mc^2 + V(\vec{r}). \tag{4.8.1}$$

$\vec{\alpha}$ and β are 4×4 matrices consisting of

$$\vec{\alpha} = \begin{pmatrix} 0 & \vec{\sigma} \\ \vec{\sigma} & 0 \end{pmatrix}; \ \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{4.8.2}$$

and the components of σ, σ_x, σ_y and σ_z represent the Pauli spin matrices. The eigenfunctions Ψ of the Dirac Hamiltonian are functions of four variables, and so they are written as functions of two two-variable functions Φ and χ

$$\Psi = \begin{pmatrix} \Phi \\ \chi \end{pmatrix}. \tag{4.8.3}$$

When the particles are electrons, Φ and χ are called respectively, the large and small components of the wavefunction [185]. The two-variable functions are coupled by the following equations

$$c(\vec{\sigma}\vec{p})\chi = (\varepsilon - V)\Phi, \tag{4.8.4}$$

and

$$c(\vec{\sigma}\vec{p})\Phi = (\varepsilon - V + 2mc^2)\chi, \tag{4.8.5}$$

giving for the large component

$$\frac{1}{2m}(\vec{\sigma}\vec{p})\frac{1}{1+(\varepsilon-V)/2mc^2}(\vec{\sigma}\vec{p})\Phi+V\Phi=\varepsilon\Phi\,, \tag{4.8.6}$$

which can be re-written as

$$\left[\left(1-\frac{\varepsilon-V}{2mc^2}\right)\frac{p^2}{2m}+V\right]\Phi-\frac{\hbar^2}{4m^2c^2}(\vec{\nabla}V\vec{\nabla}\Phi)+\frac{\hbar^2}{4m^2c^2}(\vec{\sigma}[\vec{\nabla}V,\vec{p}]\Phi)=\varepsilon\Phi\,, \tag{4.8.7}$$

using the approximation

$$\frac{1}{1+(\varepsilon-V)/2mc^2}=1-\frac{\varepsilon-V}{2mc^2}\,, \tag{4.8.8}$$

and the relations

$$\vec{p}V=V\vec{p}-i\hbar\vec{\nabla}V\,, \tag{4.8.9}$$

and

$$(\vec{\sigma}\vec{\nabla}V)(\vec{\sigma}\vec{p})=(\vec{\nabla}V\vec{p})+i\vec{\sigma}[\vec{\nabla},\vec{p}]\,. \tag{4.8.10}$$

If the potential is spherically symmetric Equation (4.8.7) becomes

$$\left[\frac{p^2}{2m}+V-\frac{p^4}{8m^3c^2}-\frac{\hbar^2}{4m^2c^2}\frac{dV}{dr}\frac{\partial}{\partial\vec{r}}+\frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}(\vec{l}\vec{s})\right]\Phi=\varepsilon\Phi\,. \tag{4.8.11}$$

The non-relativistic Schrödinger equation is given by the first and second terms alone on the left hand side (LHS) and the right hand side (RHS). The third and fourth terms on the

LHS represent respectively the mass and Darwin correction, and the last LHS term represents the spin-orbit coupling. The good quantum numbers are j, j_z and κ where

$$\vec{j} = \vec{l} + \vec{s}\,, \tag{4.8.12}$$

and

$$\kappa = \pm(j + 1/2)\,. \tag{4.8.13}$$

Ψ is not an eigenfunction of the spin or orbital moment operators because of the spin-orbit coupling. The $\hbar\kappa$ are eigenvalues of the operator

$$\mathrm{K} = \begin{pmatrix} \vec{\sigma}\vec{l} + \hbar & 0 \\ 0 & -\vec{\sigma}\vec{l} - \hbar \end{pmatrix}. \tag{4.8.14}$$

The four variable eigenfunction can now be represented as

$$\Psi = \begin{pmatrix} \Phi \\ \chi \end{pmatrix} = \begin{pmatrix} g(r)\Pi_{j,l}^{j_z} \\ if(r)\Pi_{j,l'}^{j_z} \end{pmatrix}, \tag{4.8.15}$$

where $g(r)$ and $f(r)$ are radial functions, $\Pi_{j,l}^{j_z}$ is the r-independent eigenfunction of j^2, j_z, l^2 and s^2 formed by combining the Pauli spinor with the spherical harmonics. The equations coupling $g(r)$ and $f(r)$ are

$$\hbar c\left[\frac{df(r)}{dr} + \frac{1-\kappa}{r} f(r)\right] + (\varepsilon - V)g(r) = 0\,, \tag{4.8.16}$$

and

$$\hbar c\left[\frac{dg(r)}{dr}+\frac{1+\kappa}{r}g(r)\right]-(\varepsilon-V+2mc^2)f(r)=0, \tag{4.8.17}$$

using the equality

$$\kappa(\kappa+1)=l(l+1), \tag{4.8.18}$$

and the relativistically enhanced mass

$$M=m+\frac{\varepsilon-V}{2c^2}, \tag{4.8.19}$$

the equation

$$-\frac{\hbar^2}{2Mr^2}\frac{d}{dr}\left(r^2\frac{dg(r)}{dr}\right)+\left[V+\frac{\hbar^2}{2Mr^2}\frac{l(l+1)}{r^2}\right]g(r)-\frac{\hbar^2}{4M^2c^2}\frac{dV}{dr}\frac{dg(r)}{dr}$$
$$-\frac{\hbar^2}{4M^2c^2}\frac{dV}{dr}\frac{1+\kappa}{r}g(r)=\varepsilon g(r), \tag{4.8.20}$$

results, with the function $f(r)$ given by

$$f(r)=\frac{\hbar^2}{2Mc}\left(\frac{dg(r)}{dr}+\frac{1+\kappa}{r}g(r)\right). \tag{4.8.21}$$

In order to obtain the scalar relativistic approximation, the terms which depend on κ are omitted in Equations (4.8.20) and (4.8.21). The scalar relativistic approximation results in l and s being good quantum numbers, which is very useful in spin-polarized

calculations. If the scalar relativistic approximations to $g(r)$ and $f(r)$ are denoted by $\tilde{g}(r)$ and $\tilde{f}(r)$ respectively, Equation (4.8.20) can be rewritten as

$$-\frac{\hbar^2}{2Mr^2}\frac{d}{dr}\left(r^2\frac{d\tilde{g}(r)}{dr}\right)+\left[V+\frac{\hbar^2}{2Mr^2}\frac{l(l+1)}{r^2}\right]\tilde{g}(r)-\frac{\hbar^2}{4M^2c^2}\frac{dV}{dr}\frac{d\tilde{g}(r)}{dr}=\varepsilon\tilde{g}(r), \tag{4.8.22}$$

where $\tilde{g}(r)$ and $\tilde{f}(r)$ obey the equations

$$\tilde{f}(r)=\frac{\hbar}{2Mc}\frac{d\tilde{g}(r)}{dr}, \tag{4.8.23}$$

and

$$\tilde{g}(r)=-\frac{\hbar c}{\varepsilon-V}\frac{d\tilde{f}(r)}{dr}, \tag{4.8.24}$$

with

$$\tilde{\Psi}=\begin{pmatrix}\tilde{\Phi}\\ \tilde{\chi}\end{pmatrix}, \tag{4.8.25}$$

where $\tilde{\Phi}$ is a pure spin state given by

$$\tilde{\Phi}=\tilde{g}Y_{l,m}\chi_s\,. \tag{4.8.26}$$

$\tilde{\chi}$ contains a mixture of spin-up and spin-down functions and is represented by

$$\tilde{\chi} = i\frac{\vec{\sigma r}}{r}\left(-\tilde{f}(r) + \frac{g(r)}{2Mcr}\vec{\sigma l}\right)Y_{l,m}\chi_s . \tag{4.8.27}$$

The functions $\tilde{\Psi}$ are not eigenfunctions of the Dirac Hamiltonian. The degree of deviation of the $\tilde{\Psi}$ from being eigenfunctions is used to define the spin-orbit Hamiltonian H_{so}

$$H\tilde{\psi} = \varepsilon\tilde{\psi} + H_{so}\tilde{\psi} . \tag{4.8.28}$$

H_{so} has the following form when equations (4.8.22) and (4.8.23) are used as basis functions

$$H_{so} = \frac{\hbar}{2Mc^2}\frac{1}{r}\frac{dV}{dr}\begin{pmatrix}\vec{\sigma l} & 0\\ 0 & 0\end{pmatrix}, \tag{4.8.29}$$

and operates only on the large component of the wave function.

When carrying out relativistic calculations, the core states are taken to be fully occupied, making fully relativistic calculation possible. The valence and semi-core orbitals are treated in the scalar relativistic approximation inside the atomic muffin-tin spheres. The coupled equations represented by Equations (4.8.22) and (4.8.23) are first solved in order to obtain the LAPW basis set. Since the spin quantum number is a good quantum number, spin-up and spin-down states can be considered separately as follows,

$$\phi_{l,m}^{\alpha,\uparrow} = \begin{pmatrix}\tilde{g}_l^{\uparrow}(r)Y_{l,m}\\ -i\tilde{f}_l^{\uparrow}(r)Y_{l,m}\end{pmatrix}\chi^{\uparrow}; \phi_{l,m}^{\alpha,\downarrow} = \begin{pmatrix}\tilde{g}_l^{\downarrow}(r)Y_{l,m}\\ -i\tilde{f}_l^{\downarrow}(r)Y_{l,m}\end{pmatrix}\chi^{\downarrow}; \chi^{\uparrow} = \begin{pmatrix}1\\0\end{pmatrix}; \chi^{\downarrow} = \begin{pmatrix}0\\1\end{pmatrix}. \tag{4.8.30}$$

The LAPW basis sets for spin-up or spin-down states are constructed by combining these $\phi_{l,m}^{\alpha,\uparrow}$ and $\phi_{l,m}^{\alpha,\downarrow}$ basis functions respectively. To include spin-orbit coupling, which will have non-zero matrix elements between spin-up and spin-down basis functions, the second variational method is used [185]. According to this procedure, the eigenvalue equation is solved with the regular non-relativistic Hamiltonian in the usual way separately for spin-up states and eigenvalues $\psi_{n,\vec{k}}^{\uparrow}$, $E_{n,\vec{k}}^{\uparrow}$ and the spin-down states and eigenvalues $\psi_{n,\vec{k}}^{\downarrow}$, $E_{n,\vec{k}}^{\downarrow}$. In the second variational step, a new eigenvalue problem is solved with the completely relativistic Hamiltonian that includes H_{so} using the $\psi_{n,\vec{k}}^{\uparrow}$ and $\psi_{n,\vec{k}}^{\downarrow}$ obtained in the first variational step as basis functions. In this new step, the contributions to the eigenvalues due to the scalar relativistic Hamiltonian occur as the already calculated $E_{n,\vec{k}}^{\uparrow}$ and $E_{n,\vec{k}}^{\downarrow}$ in the diagonal matrix elements, making it straightforward to calculate the matrix elements due to H_{so}. In most cases the basis used in the second variational step is much smaller than the basis used in the first step, so that the second eigenvalue problem results in much smaller matrix sizes. The option exists also to choose H_{so} matrix elements between states according to the magnitude of their energy differences.

In the following chapters, the computational methods described in this chapter as implemented in the WIEN97 code package [186] will be used to search for metastable phases of Ta and the neighboring bcc metals in the Periodic Table. They will also be used to calculate the mechanical properties and transport properties of the normal state and metastable phases of these metals.

CHAPTER 5

METASTABLE PHASES OF GROUP VB – VIB TRANSITION METALS

Elastic properties such as the bulk modulus, shear modulus, Young's modulus and the Poisson ratio control the response of a crystal structure to applied external forces. These properties provide insight into the nature of bonding and hence structural stability. The nature of the chemical bonds formed between the atoms in a solid is dictated by the valence electron charge density distribution. Ductile materials such as for example Al have highly uniform valence electron charge density distributions, which is reflected in the nearly spherical Al Fermi surface.

In the bcc transition metals, the valence band is significantly influenced by unfilled d bands derived from the atomic d states. The individual d states are highly localized and directional, rendering their ability to form bonds quite sensitive to the nearest-neighbor environment. Cohesion and bonding properties are determined by the $e_g - t_{2g}$ splitting of the d band and hybridization resulting in highly directional charge accumulation and highly non-uniform charge density distribution. These band structure properties would be expected to play an important role in the determination of the structure of thin films of transition metals deposited on substrates by any of the film deposition methods discussed in Chapter 3. Two of the most important factors determining the crystal structure of deposited films are the difference in surface energy between the film and substrate material and the energy barrier to atomic diffusion. The degree of atomic diffusion on substrates is controlled by the energy barrier to diffusion and is very important in determining the morphology of the resulting film. The energy barrier associated with atomic rearrangements in atomic clusters of a material is related to

the shear constant of the material, which is an elastic property. The tendency towards the formation of metastable phases can therefore be studied for Ta and the rest of the bcc metals in Groups VB-VIB of the Periodic table by studying their elastic properties. The discussion of strain analysis following in Section 5.1 is necessary for an appreciation of these properties. This discussion is taken entirely from Reference [187].

5.1 Strain Analysis

In general, the strain on a crystalline material due to an applied mechanical stress is not isotropic. Any anisotropic physical property can be represented by a tensor. A tensor of second rank, represented by T_{ij} is symmetrical if $T_{ij} = T_{ji}$ and anti-symmetrical if $T_{ij} = -T_{ji}$. When the physical property in question is a relation between two vectors, the tensor representing it will be of second rank since a vector is a tensor of first rank. If a vector $\vec{p}$ with components p_1, p_2 and p_3 depends on another vector $\vec{q}$ with components q_1, q_2 and q_3 then

$$p_1 = \alpha_{11}q_1 + \alpha_{12}q_2 + \alpha_{13}q_3, \tag{5.1.1}$$

$$p_2 = \alpha_{21}q_1 + \alpha_{22}q_2 + \alpha_{23}q_3, \tag{5.1.2}$$

and

$$p_3 = \alpha_{31}q_1 + \alpha_{32}q_2 + \alpha_{33}q_3. \tag{5.1.3}$$

The matrix of coefficients

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{3} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}, \tag{5.1.4}$$

represents a tensor [187]. A symmetric tensor has diagonal elements equal to zero and can be represented geometrically by a quadric. The general equation of the quadric S is

$$\sum_i \sum_j S_{ij} X_i X_j = 1, \tag{5.1.5}$$

where i and j are summed up to 3. In matrix form the quadric can be represented as

$$S_{ij} = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix}. \tag{5.1.6}$$

A set of principal axes consists of any three mutually perpendicular directions that stay perpendicular to each other during deformation. The quadric is reduced to three components with reference to its principal axes. The physical properties of crystals, which are represented by tensors, must possess at least all the elements of the point group symmetry of the crystal. The symmetry of the representation quadric used to represent a physical property of a crystal contains every symmetry element of the crystal. Stress acting on a crystal produces a strain in the crystal. In general the tensor representing the stress and the strain do not have to conform to the symmetry of the crystal. However when the stress on a crystal is imposed by another crystalline material, then for small strains, the strain in the crystal can be governed by the symmetry of the crystal class to which the metal substrate belongs.

The matrix of coefficients of a tensor depends on which set of axes are being used as a reference. Changing from one orthogonal set of axes (x_1, x_2, x_3) to another set (x_1', x_2', x_3'), a second rank tensor T transforms as

$$T_{ij}' = \sum_k \sum_l b_{ik} T_{kl} b_{jl} , \tag{5.1.7}$$

where the b_{ij} are the direction cosines for the transformation of axes. There will therefore be eighty-one possible terms for the transformation of a tensor by change of reference axes. In comparison, a vector v transforms as

$$v_i' = \sum_j b_{ij} v_j . \tag{5.1.8}$$

The following relations between the two sets of axes hold for the transformation

$$x_1' = b_{11} x_1 + b_{12} x_2 + b_{13} x_3 , \tag{5.1.9}$$

$$x_2' = b_{21} x_1 + b_{22} x_2 + b_{23} x_3 , \tag{5.1.10}$$

and

$$x_3' = b_{31} x_1 + b_{32} x_2 + b_{33} x_3 . \tag{5.1.11}$$

The matrix of direction cosines for the transformation is

$$b_{ij} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}. \tag{5.1.12}$$

The relations between the direction cosines are

$$b_{11}{}^2 + b_{12}{}^2 + b_{13}{}^2 = 1, \tag{5.1.13}$$

$$b_{21}{}^2 + b_{22}{}^2 + b_{23}{}^2 = 1, \tag{5.1.14}$$

and

$$b_{31}{}^2 + b_{32}{}^2 + b_{33}{}^2 = 1, \tag{5.1.15}$$

and also

$$b_{21}b_{31} + b_{22}b_{32} + b_{23}b_{33} = 0, \tag{5.1.16}$$

$$b_{31}b_{11} + b_{32}b_{12} + b_{33}b_{13} = 0, \tag{5.1.17}$$

and

$$b_{11}b_{21} + b_{12}b_{22} + b_{13}b_{23} = 0. \tag{5.1.18}$$

The stress σ is defined by the equation

$$F = \sigma dA, \tag{5.1.19}$$

which relates the force and unit area vectors. In general the stress is anisotropic so that

$$F_i = \sigma_{ij} dA_j, \tag{5.1.20}$$

and the stress can be represented as a tensor where σ_{ij} means that the stress is applied in the i direction on the surface with normal in the j direction. The strain e at a point is defined as the rate of change of elongation l between two points separated by a distance Δx with length as Δx goes to zero.

$$e_{ij} = \frac{\partial l_i}{\partial x_j}. \tag{5.1.21}$$

The strain is a tensor with nine components. This tensor consists of a symmetrical and an anti-symmetrical part so that it is written as

$$e_{ij} = \frac{1}{2}(e_{ij} + e_{ji}) + \frac{1}{2}(e_{ij} - e_{ji}), \tag{5.1.22}$$

or

$$e_{ij} = e_{ij}^{sym} + e_{ij}^{asym}. \tag{5.1.23}$$

The symmetric part of e_{ij} disappears when there is no distortion. It corresponds to a pure strain. The anti-symmetric part of e_{ij} corresponds to a rigid body rotation without any

distortions and is therefore finite when no distortions are imposed. The symmetric part of e_{ij} is the strain tensor ε_{ij} given by

$$\varepsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji}). \tag{5.1.24}$$

The displacement due to a homogeneous distortion is then equal to

$$l_i = (l_0)_i + e_{ij}^{\ asym} x_j + e_{ij}^{\ sym} x_j, \tag{5.1.25}$$

where the displacement of the point at the origin is $(l_0)_i$. The displacement due to a rigid body translation and rotation is $(l_0)_i + e_{ij}^{\ asym} x_j$, while the displacement due to strain is $e_{ij}^{\ sym} x_j$. Since the strain tensor

$$\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}, \tag{5.1.26}$$

is symmetrical, it can be referenced to its principal axes to produce

$$\begin{pmatrix} \varepsilon_{11} & 0 & 0 \\ 0 & \varepsilon_{22} & 0 \\ 0 & 0 & \varepsilon_{33} \end{pmatrix}, \tag{5.1.27}$$

where the quantities ε_{11}, ε_{22} and ε_{33} are the principal strains.

Applying a stress σ to a material produces strain in that material according to the equation

$$\varepsilon = s\sigma, \tag{5.1.28}$$

where s is the compliance of the material and ε is the strain tensor. The stiffness c of the material is defined as

$$c = \frac{1}{s}. \tag{5.1.29}$$

The stress can then be written in terms of the strain and the stiffness as

$$\sigma = c\varepsilon. \tag{5.1.30}$$

Since the stiffness relates the stress and strain tensors, it is a fourth rank tensor and the relation can be written as

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl}. \tag{5.1.31}$$

This means that there are nine possible stress and strain components and eighty-one possible stiffness components. Also

$$\varepsilon_{ij} = s_{ijkl}\sigma_{kl}, \tag{5.1.32}$$

and the compliance is also a fourth rank tensor with eighty-one components. The stresses σ_{ij} and σ_{ji} always occur together in pairs, giving

$$s_{ijkl} = s_{ijlk}, \; s_{ijkl} = s_{jikl}, \tag{5.1.33}$$

and

$$c_{ijkl} = c_{ijlk}\,,\ c_{ijkl} = c_{jikl}\,. \tag{5.1.34}$$

When the reference axes are changed, the compliances and stiffnesses transform as

$$s'_{ijkl} = \sum_m \sum_n \sum_o \sum_p b_{im} b_{jn} b_{ko} b_{lp} s_{mnop}\,, \tag{5.1.35}$$

and

$$c'_{ijkl} = \sum_m \sum_n \sum_o \sum_p b_{im} b_{jn} b_{ko} b_{lp} c_{mnop}\,. \tag{5.1.36}$$

A change of notation is usually effected [187] in order to reduce the number of subscripts on the compliances and stiffnesses from four to two. The change of notation is as follows; 11, 22 and 33 subscripts become 1, 2 and 3 respectively, and 12, 13 and 23 (equivalent to 21, 31 and 32 respectively) become 6, 5 and 4 respectively.

$$s_{ijkl} = s_{op}\,, \tag{5.1.37}$$

when o and p are both equal to 1, 2 or 3,

$$2s_{ijkl} = s_{op}\,, \tag{5.1.38}$$

when either o or p are 4, 5 or 6, and

$$4s_{ijkl} = s_{op}\,, \tag{5.1.39}$$

when o and p are both equal to 4, 5 or 6. For example

$$\varepsilon_{11} = s_{1111}\sigma_{11} + s_{1112}\sigma_{12} + s_{1113}\sigma_{13} + s_{1121}\sigma_{21} + s_{1122}\sigma_{22} + s_{1123}\sigma_{23}$$
$$+ s_{1131}\sigma_{31} + s_{1132}\sigma_{32} + s_{1133}\sigma_{33}\,, \tag{5.1.40}$$

is equivalent to

$$\varepsilon_1 = s_{11}\sigma_1 + s_{16}\sigma_6 + s_{15}\sigma_5 + s_{16}\sigma_6 + s_{12}\sigma_2 + s_{14}\sigma_4 + s_{15}\sigma_5 + s_{14}\sigma_4 + s_{13}\sigma_3\,, \tag{5.1.41}$$

or

$$\varepsilon_1 = s_{11}\sigma_1 + 2s_{16}\sigma_6 + 2s_{15}\sigma_5 + s_{12}\sigma_2 + 2s_{14}\sigma_4 + s_{13}\sigma_3\,, \tag{5.1.42}$$

without this convention and becomes

$$\varepsilon_1 = s_{11}\sigma_1 + s_{16}\sigma_6 + s_{15}\sigma_5 + s_{12}\sigma_2 + s_{14}\sigma_4 + s_{13}\sigma_3\,, \tag{5.1.43}$$

with the convention given above. The components of the strain tensor can then be represented as

$$\varepsilon_o = \sum_{p=1}^{6} s_{op}\sigma_p\,. \tag{5.1.44}$$

The other ε_{ij} are similarly transformed when changing to the ε_o notation.

The symmetry elements of the point group to which a crystal belongs can be applied to the crystal referenced to a certain set of axes to determine how the axes

transform under the symmetry operations about the symmetry elements. The way in which the axes transform determine further relations between the compliances and elastic constants and since the compliance and stiffness tensors are symmetrical, this can be used as a constraint to determine which of the components of the compliance and stiffness tensors vanish for any given crystal belonging to a crystal system since if $T_{ij} = -T_{ji}$ for a symmetrical tensor, then $T_{ij} = -T_{ji} = 0$.

5.2 Elastic Properties and the Equation of State

The equation of state of a material describes the volume and pressure dependence of its total energy. If E_{tot} is the total energy of a strained crystalline structure, it can be expressed [188] as

$$E_{tot} = E_{tot}^{0} + P(V - V_0) + E_{elast}, \tag{5.2.1}$$

where E_{tot}^{0} and E_{elast} are respectively, the unstrained crystalline total energy and the strain or elastic energy, V_0 and V are respectively, the equilibrium and strained crystal lattice volumes and P is the pressure given by [188]

$$P = -\left(\frac{\partial E_{tot}^{0}}{\partial V}\right)_{V_0}. \tag{5.2.2}$$

The strained lattice vectors $\vec{a}$ are related to the unstrained lattice vectors $\vec{a}_0$ to first order by [188]

$$\vec{a} = \vec{a}_0(\mathrm{I} + \bar{\varepsilon}), \tag{5.2.3}$$

where I is the identity matrix and $\bar{\varepsilon}$ is the strain tensor. There are two cases of interest, namely cases of cubic symmetry and tetragonal symmetry, in which it is desired to calculate the elastic constants, which determine the behavior of crystals under strain.

5.2.1 Cubic Symmetry

The work done on a unit cubic crystal by the stress components σ_{ij} which change a small homogeneous strain with components ε_{ij} to one with components $\varepsilon_i + d\varepsilon_i$ is [187 – 188]

$$dW = \sum_i \sum_j \sigma_{ij} d\varepsilon_{ij} . \tag{5.2.1.1}$$

For an isothermal and reversible deformation, the work done is equal to the Gibbs free energy or the elastic energy $d\phi_{elast}$ so that [187 – 188]

$$d\phi_{elast} = \sum_i \sum_j \sigma_{ij} d\varepsilon_{ij} , \tag{5.2.1.2}$$

and

$$d\phi_{elast} = \sum_i \sum_j c_{ij} \varepsilon_j d\varepsilon_i , \tag{5.2.1.3}$$

by Hooke's law. Integrating to obtain the elastic energy, which is the work needed to produce the strain ε_i,

$$\phi_{elast} = \frac{1}{2}\sum_i \sum_j c_{ij}\varepsilon_i\varepsilon_j \,. \tag{5.2.1.4}$$

The strain or elastic energy of a crystalline structure per unit volume, ϕ_{elast}, is then

$$\phi_{elast} = \frac{1}{2}\sum_i \sum_j c_{ij}\varepsilon_i\varepsilon_j \,, \tag{5.2.1.5}$$

and the total elastic energy E_{elast} is [187]

$$E_{elast} = \frac{V}{2}\sum_i \sum_j c_{ij}\varepsilon_i\varepsilon_j \,. \tag{5.2.1.6}$$

The c_{ij} are the so-called elastic constants. The elastic energy must be positive for the crystalline structure to be stable with respect to an applied mechanical stress. This requirement imposes restrictions on the elastic constants c_{ij}. For cubic crystals, the restrictions are [187 – 188]

$$c_{44} > 0, \quad c_{11} > |c_{12}|, \quad c_{11} + 2c_{12} > 0. \tag{5.2.1.7}$$

The volume term in Equation (5.2.1.6), just as in Equation (5.2.1), is linear with the strain, suggesting that the elastic constants can be obtained from [188]

$$c_{ij} = \frac{1}{V_0}\left(\frac{\partial^2 E_{tot}}{\partial\varepsilon_i\varepsilon_j}\right)_{V_0} \,. \tag{5.2.1.8}$$

The three elastic constants c_{11}, c_{12} and c_{44} completely describe the elastic behavior of a cubic crystal under strain. In order to calculate these three constants, three kinds of strains must be applied to the cubic crystal to obtain the required three equations. The first kind of applied strain consists of keeping the crystal c/a ratio constant while calculating the total energy E_{tot} for different values of the volume V. The resulting energy values can then be fitted to an equation of state such as the Murnaghan [189] equation of state

$$E_{tot} = E_{tot}^{0} + \frac{BV}{B'}\left[\frac{1}{B'-1}\left(\frac{V_0}{V}\right)^{B'} + 1\right], \tag{5.2.1.9}$$

or the Birch-Murnaghan [190] equation of state

$$E_{tot} = a + bV^{-2/3} + cV^{-1/3} + dV^{-2}, \tag{5.2.1.10}$$

where B and B' are respectively, the bulk modulus and its first derivative, and a, b, c and d are constants. The bulk modulus is the resistance of a structure to deformation by a uniform hydrostatic pressure and is defined as [187 – 188]

$$B = V_0\left(\frac{\partial^2 E_{tot}}{\partial V^2}\right)_{V_0} = \frac{1}{3}(c_{11} + 2c_{12}). \tag{5.2.1.11}$$

Values for B can then be calculated accordingly from the fit of calculated total energies as a function of volume to the Murnaghan equation. Such a fit is shown in Figure 5.1 for cubic Ta.

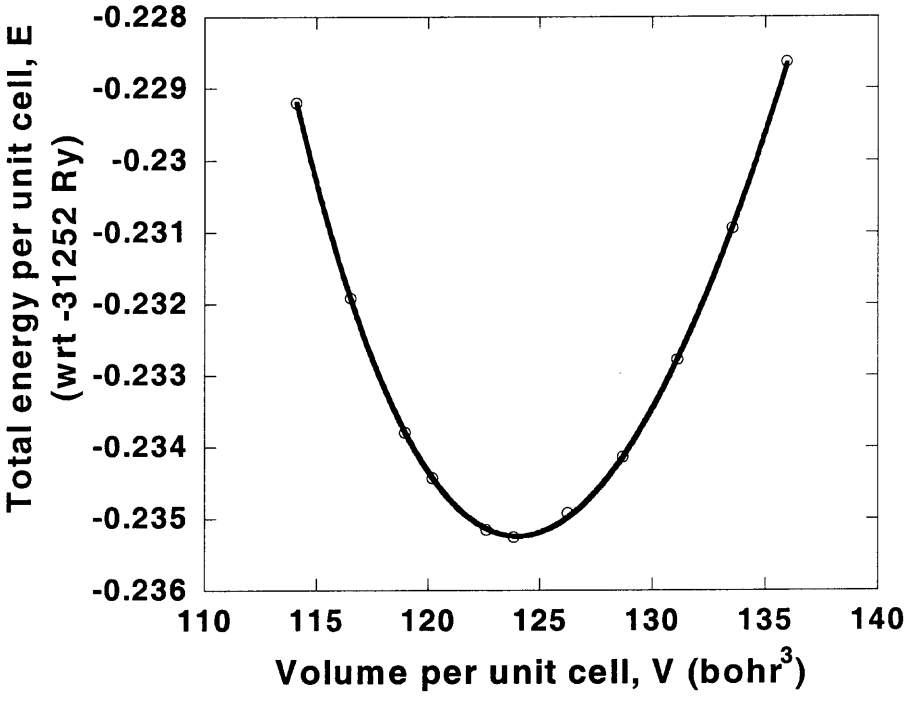


Figure 5.1 Energy vs. volume curve for cubic Ta obtained from a fit of LAPW calculated total energies to the Murnaghan equation of state. The fit results in values for Ta of V_0=124 bohr^3, and B = 195.6 GPa.

The second kind of symmetrical structure deformation involves volume-conserving changes in the c/a ratio of the form [188]

$$c/a = 1 + e, \tag{5.2.1.12}$$

for several values of e, resulting in the strain tensor

$$\bar{\delta} = \begin{bmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_1 & 0 \\ 0 & 0 & \dfrac{1}{(1+\delta_1)^2} - 1 \end{bmatrix}, \tag{5.2.1.13}$$

with δ_1 given by [188]

$$\delta_1 = (1+e)^{-1/3} - 1. \tag{5.2.1.14}$$

The following expression involving the elastic constants c_{11} and c_{12} is then obtained [188] for the elastic energy

$$E_{elast} = 3V_0(c_{11} - c_{12})\delta_1^{\,2} + \mathrm{O}(\delta_1^{\,3}). \tag{5.2.1.15}$$

The total energy can be fitted to a polynomial p of degree N, where N is determined by the number of different structures formed by the deformations in Equation (5.2.1.12). Taking the second derivative of the resulting fitted polynomial with respect to δ_1 yields a simple expression for the elastic constants c_{11} and c_{12},

$$p''(\delta_1 = 0) = 6V_0(c_{11} - c_{12}). \tag{5.2.1.16}$$

At this point two equations for the two unknown elastic constants have been obtained and another important elastic property, the shear constant, can be calculated.

The shear constant G [187 – 188] is the resistance to shear deformation across the (110) plane in the $[1\,\bar{1}\,0]$ direction and is defined as

$$G = (c_{11} - c_{12})/2. \tag{5.2.1.17}$$

The last kind of deformation required in order to determine all the cubic elastic constants is a rhombohedral distortion, which varies the great diagonal length of a cubic cell [188]. The rhombohedral unit cell vectors $\vec{a}_i^r$ are given in terms of the cubic unit cell vectors $\vec{a}_j^0$ in the following manner,

$$\begin{pmatrix} \vec{a}_1^r \\ \vec{a}_2^r \\ \vec{a}_3^r \end{pmatrix} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{pmatrix} \vec{a}_1^0 \\ \vec{a}_2^0 \\ \vec{a}_3^0 \end{pmatrix}, \tag{5.2.1.18}$$

for fcc lattice structures, and given by

$$\begin{pmatrix} \vec{a}_1^r \\ \vec{a}_2^r \\ \vec{a}_3^r \end{pmatrix} = \begin{bmatrix} -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{pmatrix} \vec{a}_1^0 \\ \vec{a}_2^0 \\ \vec{a}_3^0 \end{pmatrix}, \tag{5.2.1.19}$$

for bcc lattice structures and

$$\vec{a}_i^r = \vec{a}_i^0, \tag{5.2.1.20}$$

for simple cubic lattice structures. This defines a rhombohedral lattice with hexagonal unit cell vectors $\vec{a}_i^H$ in terms of the rhombohedral lattice vectors

$$\begin{pmatrix} \vec{a}_1^H \\ \vec{a}_2^H \\ \vec{a}_3^H \end{pmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} \vec{a}_1^r \\ \vec{a}_2^r \\ \vec{a}_3^r \end{pmatrix}, \tag{5.2.1.21}$$

varying $\vec{a}_3^H$ according to [188]

$$\vec{a}_3^H = \vec{a}_3^{H,0}(1+e), \tag{5.2.1.22}$$

for several values of e, resulting in the tensor

$$\bar{\delta} = \frac{e}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \tag{5.2.1.23}$$

and the following expression [188] for the elastic energy

$$E_{elast} = \frac{V_0 e^2}{3}\left(\frac{c_{11}+2c_{12}}{2}+2c_{44}\right)+\mathrm{O}(e^3). \tag{5.2.1.24}$$

Fitting this expression to a polynomial p and taking the second derivative of p results in

$$p''(e=0) = \frac{1}{3}(c_{11}+2c_{12}+4c_{44}). \tag{5.2.1.25}$$

This third equation can now be used to calculate the elastic constant c_{44}, which is the resistance to shear deformation across the (100) plane in the [010] direction.

5.2.2 Tetragonal Symmetry

The tetragonal elastic constants are coefficients in the expression for the total energy of a crystal having tetragonal symmetry. Expanding the total energy in a Taylor series about the point (0, 0) [169, 191],

$$E_{tot} = E(a_0, c_0) + \left(\frac{\partial E}{\partial a}\right)_c a + \left(\frac{\partial E}{\partial c}\right)_a c + \left(\frac{\partial^2 E}{\partial a \partial c}\right) ac$$

$$+ \frac{1}{2}\left(\frac{\partial^2 E}{\partial a \partial c}\right) a^2 + \frac{1}{2}\left(\frac{\partial^2 E}{\partial a^2}\right) c^2 + \cdots . \qquad (5.2.2.1)$$

Taking the first differential of this expression for the energy and neglecting terms of order higher than two gives

$$\delta E = \left(\frac{\partial E}{\partial a}\right)_c \delta a + \left(\frac{\partial E}{\partial c}\right)_a \delta c + \left(\frac{\partial^2 E}{\partial a \partial c}\right) c\delta a + \left(\frac{\partial^2 E}{\partial a \partial c}\right) a\delta c$$

$$+ \left(\frac{\partial^2 E}{\partial a^2}\right) a\delta a + \left(\frac{\partial^2 E}{\partial c^2}\right) c\delta c . \qquad (5.2.2.2)$$

Taking the second differential results in

$$\delta^2 E = \left(\frac{\partial E}{\partial a}\right)_c \delta^2 a + \left(\frac{\partial E}{\partial c}\right)_a \delta^2 c + 2\left(\frac{\partial^2 E}{\partial a \partial c}\right) \delta c \delta a + \left(\frac{\partial^2 E}{\partial a \partial c}\right) c\delta^2 a + \left(\frac{\partial^2 E}{\partial a \partial c}\right) a\delta^2 c$$

$$+ \left(\frac{\partial^2 E}{\partial a^2}\right) a\delta^2 a + \left(\frac{\partial^2 E}{\partial a^2}\right) (\delta a)^2 + \left(\frac{\partial^2 E}{\partial c^2}\right) c\delta^2 c + \left(\frac{\partial^2 E}{\partial c^2}\right) (\delta c)^2 . \qquad (5.2.2.3)$$

Neglecting the second order differentials of the lattice constants, the second differential of the Taylor series representation of the energy can be written in the form

$$\delta^2 E = 2\left(\frac{\partial^2 E}{\partial a \partial c}\right) \delta c \delta a + \left(\frac{\partial^2 E}{\partial a^2}\right) (\delta a)^2 + \left(\frac{\partial^2 E}{\partial c^2}\right) (\delta c)^2 . \qquad (5.2.2.4)$$

Defining the tetragonal elastic stiffness coefficients [191] as

$$\bar{c}_{11} = \frac{a^2}{V}\frac{\partial^2 E}{\partial a^2}, \tag{5.2.2.5}$$

$$\bar{c}_{13} = \frac{ac}{V}\frac{\partial^2 E}{\partial a \partial c}, \tag{5.2.2.6}$$

and

$$\bar{c}_{33} = \frac{c^2}{V}\frac{\partial^2 E}{\partial c^2}. \tag{5.2.2.7}$$

These are the coefficients of the second order differential of E, $\delta^2 E$, as has been previously shown by the expansion of the energy in a Taylor series. The relation is [169, 191]

$$\frac{\delta^2 E}{V} = \bar{c}_{13}\left(\frac{\delta a}{a}\right)\left(\frac{\delta c}{c}\right) + \frac{\bar{c}_{11}}{2}\left(\frac{\delta a}{a}\right)^2 + \frac{\bar{c}_{33}}{2}\left(\frac{\delta c}{c}\right)^2. \tag{5.2.2.8}$$

The epitaxial line is produced by the value of the c lattice parameter that results in the minimum total energy at each value of the a lattice parameter. The isotropic in plane stress σ along the epitaxial line can be obtained from [169]

$$\sigma = \frac{a}{2V}\left(\frac{\partial E}{\partial a}\right)_c. \tag{5.2.2.9}$$

In order to identify any deposited film as a strained sample of a particular equilibrium phase and also to determine the maximum compressive and tensile strain that

the equilibrium phase and any other metastable phase can accept before they become unstable, the volume, energy and stiffness coefficient dependence on the c/a ratio can be examined to identify the regions of instability according to the criterion [169, 191]

$$D = \bar{c}_{11}\bar{c}_{33} - \bar{c}_{13}^{\,2} > 0 . \tag{5.2.2.10}$$

The above stability criterion arises from the requirement that $\delta^2 G$ is positive definite. When D is less than zero, the tetragonal structure under the strains defined by $\bar{c}_{11}$, $\bar{c}_{33}$ and $\bar{c}_{13}$ is unstable and cannot be stabilized [169, 191]. It will therefore not be possible to grow more than a few layers of this phase.

5.3 The Epitaxial Bain Path

A particular tetragonal phase of a metallic film can be formed when the tetragonal phase in question is a local minimum of energy with respect to tetragonal stresses. The tetragonal structure will be metastable or stable depending on whether it corresponds to a local or global minimum of energy and whether certain conditions are satisfied by its elastic constants. The Bain path concept [154 – 155, 168 – 169, 191] can provide predictive power in terms of being able to determine by first principles calculations which strain derived structure of an equilibrium metallic phase can be formed on any given substrate. If a cubic structured material with lattice constant a_0 is deposited on a substrate with square mesh size a_1, the lattice mismatch between the substrate with square mesh of size a_1 and the film material with lattice constant a_0 will result in the film being

strained. This strain is mostly parallel to the substrate surface and will affect the interlayer spacing c in the perpendicular direction, leading to a body centered tetragonal structure.

The Epitaxial Bain Path (EBP) describes the Bain [168] transformation, which can take a bcc structure to a bct structure, forming an fcc phase as an intermediate phase. The bcc and fcc phases occur at c/a ratios of 1 and $\sqrt{2}$ respectively. The Bain transformation involves uniaxial strain along the [001] direction and the shear of bcc (101) planes in the $[\bar{1}01]$ direction so that the (111) fcc planes are parallel to the (101) bcc planes. The EBP is defined [154 – 155, 169, 191] for tetragonal states as the curve in the $(a,\ c)$ plane such that each c value corresponds to the lowest energy of the tetragonal structure corresponding to a substrate square mesh size of a. Given a value of a, a total energy calculation is done to find the value of c that minimizes the energy. The total energy calculated from density functional theory can be fitted to a polynomial of the form [191]

$$E = \sum_i \sum_j A_{ij} (c/a)^i V^j, \tag{5.3.1}$$

to obtain an analytical expression for the total energy. Elastic constants for crystals undergoing any kind of stress or strain can then be calculated in a straightforward way by simply taking derivatives of the analytical expression for the total energy. This approach enables the straightforward study of stresses on the film surface. The EBP for a metal can be obtained by setting the quantity $(\partial E/\partial c)_a$ equal to zero and solving for c [191]. The resulting c value will then be the c value that minimizes the total energy for that

particular constant value of the a lattice parameter. This minimum c value corresponds to zero stress normal to the (001) surface.

EBPs can be represented [154 – 155, 169, 191] by plotting V/V_0 versus a, V/V_0 versus c/a and c versus a where c is the value for which the total energy is a minimum for each given a lattice parameter, V is the volume per atom given by $ca^2/2$ and V_0 is the equilibrium volume per atom of the stable phase of the film material being studied. A typical Bain path is shown in Figure 5.2.

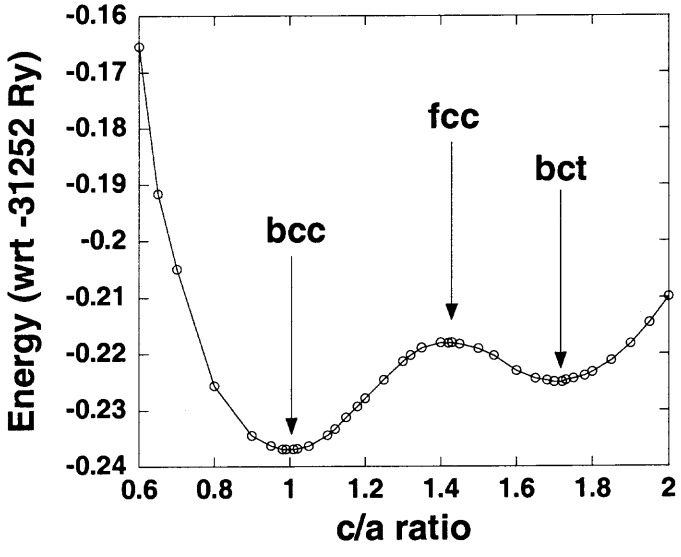


Figure 5.2 Bain path for Ta showing the bcc, fcc and bct phases.

The EBP can give important insight into the properties of different film structures. For example, a comparison of the measured structure of a deposited film with the EBP of the film material can provide information concerning the state from which the deposited film is derived by straining. The EBP also aids in selecting a suitable substrate to achieve a given film stability. To locate the point on the EBP corresponding to the deposited film,

any number of diffraction techniques could be used to measure the c value for the deposited film. Since the a parameter is usually known for the substrate, V/V_0 is also known and the appropriate point can then be located on the EBP.

Bain paths can also connect bcc and fcc phases along paths of tetragonal symmetry. Examples of Bain paths include the epitaxial Bain path produced by isotropic stress in the (001) plane and negligible stress perpendicular to the plane. The uniaxial Bain path is produced by applied stress along the [001] direction of a tetragonal phase and no stress in the (001) plane [192]. Both paths have been shown to lead to the lowest possible maximum energies of all Bain paths between tetragonal phases. Elastic properties of metastable phases, which are impossible to measure experimentally, can therefore be determined by curve fitting to *ab-initio* total energy calculations. The Bain path concept has been used to predict and prepare a metastable phase of vanadium [154 - 155].

5.4 Metastable Phases of the Group 5B – 6B Bcc Metals

Under equilibrium conditions free standing early Group 5 and Group 6 metals form bcc structures. The bcc and the fcc crystal structures can be regarded as special cases of the more general bct structure in metals in which the a and the b lattice parameters are equal to each other but not necessarily to the c lattice parameter. The bcc and fcc structures result when c/a ratios of 1 and $\sqrt{2}$ respectively are obtained while deforming a structure having the general bct symmetry ($a = b \neq c$) along the [001] direction with the constraint that $a = b$, according to the Bain transformation.

In the course of preparing thin films of bcc metals on various substrates, conditions such as impurity levels in the preparation environment, substrate strain and substrate temperature can result in structural phases other than bcc being formed. Accordingly, numerous investigators have prepared, mainly by vacuum evaporation and sputter-deposition techniques, thin film non-cubic phases of bcc metals, which were found to be metastable (*i.e.* which reverted back to the cubic phase upon heating or further deposition). These phases are often observed to have mechanical and transport properties different from those of the corresponding bcc phase and their exact structural forms are usually not immediately deducible from standard experimental analytical techniques alone. An insightful analysis usually requires applying theoretical methods in the form of *ab initio* band structure calculations in addition to typical experimental analytical tools. In order to predict and understand the bct structures formed by thin films of the bcc transition metals in columns 5 – 6 of the Periodic Table, the total energy per unit cell as functions of the unit cell volume and the c/a ratio have been calculated. It would also be worthwhile to investigate whether tetragonal distortions of the stable bcc structure of tantalum can form a metastable phase that can account for the large resistivity of β–tantalum relative to α bcc tantalum. This is done by calculating, for tantalum as well as for the rest of the bcc metals in groups 5B – 6B and periods 4 - 6 of the Periodic Table, the total energy per unit cell as functions of the unit cell volume and the c/a ratio.

Different alloy combinations of the metals in this study have been prepared as metastable phases before [170 – 172]. In addition, metastable fcc phases of Ta [96], Cr [157], Mo and W [158] have been observed experimentally, but as far as is known, except for V, no metastable bct phases of these metals which do not have the fcc or bcc

structure have been prepared in elemental form. Hence, it is hoped that this work will stimulate experimental work on the metastable bct phases of the early group 5B and group 6B metals.

5.4.1 Implementation Details

The WIEN97 code developed by Blaha *et al* [186] implementing the FLAPW method in the DFT formalism was used for the calculations. The WIEN97 code solves the single particle Kohn-Sham equations for the Kohn-Sham eigenvalues, ground state density and total energy of a many electron system using the FLAPW method without making any shape approximations to the potential. Core states are treated in a fully relativistic manner while the valence states are treated semi-relativistically. The exchange-correlation part of the total energy is represented with the generalized gradient approximation (GGA) using the parameterization of Perdew-Burke-Ernzerhof [176]. DFT in the GGA is now supplanting DFT in the local density approximation (LDA) for electronic structure study and prediction of ground state properties of different materials and systems. The GGA formalism addresses the non-uniformity of the electron gas using its density gradients, while the LDA formalism assumes electron homogeneity. Lattice constants are usually overestimated by the GGA and underestimated by the LDA, but lattice constant values tend to be closer to experiment when calculated with the GGA than when calculated with the LDA [177].

The calculations consisted of calculating the total energy per unit cell at T = 0 K as a function of the unit cell volume at a series of c/a ratios in order to determine the energy minimum E_0 in the (V,E) plane at each c/a ratio and thereby determine the

absolute energy minima in the $(c/a, E_0)$ plane. All calculations were begun by defining a unit cell, which was subsequently divided into non-overlapping atomic sphere and interstitial regions. The solutions to the Kohn-Sham equations were expanded in a combined Linear Augmented Plane Wave (LAPW) basis set consisting of plane waves in the interstitial region of the unit cell and a linear combination of products of spherical harmonics and radial functions inside the atomic spheres. The muffin-tin radii chosen for the atomic spheres ranged from 2.3 bohr for V and Cr, through 2.5 bohr for Nb and Mo, to 2.7 for Ta and W. The charge densities and the wave functions were expanded in partial waves up to l = 10 inside the atomic spheres and the potential was expanded up to l = 4 inside the spheres. The RKMAX parameter, which is the product of the muffin-tin radius and the cutoff for the plane-wave expansion, was chosen to be 10. The magnitude of the largest vector in the Fourier expansion of the charge density, represented by the GMAX parameter, was chosen to be 14 bohr $^{-1}$.

The structure optimization computations were carried out using over 1000 irreducible k-points in the first Brillouin zone for all the studied elements. The k-points were generated on a grid, which was used in a modified tetrahedron integration method [193] to calculate the Fermi energy. The total charge density per unit cell was converged to within 10^{-5} electrons/a_0^3 self consistently and then converged with respect to the number of k-points in the reciprocal lattice. Twenty-one structures, each with a different c/a ratio, were created for each cubic element studied. The following procedure was carried out for each structure: keeping the c/a ratio constant, the unit cell volume of the structure was optimized by calculating the total energy per unit cell as a function of the unit cell volume, which was varied in various sub-ranges of the range of volumes

consisting of –20% to 25% of the initial unit cell volume, depending on the magnitude of the volume corresponding to the energy minimum at that c/a ratio. The LAPW calculated energies were then fitted to the Murnaghan [189] equation of state to determine the energy minimum E_0 in the (V,E) plane.

5.4.2 Calculated Metastable Phases and Energy Differences

Two energy minima in the $(c/a,E_0)$ plane, corresponding to metastable phases, were obtained for all the studied cubic elements. The equilibrium total energy per unit cell E_0 as a function of c/a ratio is shown in Figure 5.3 for the bcc metals in Groups VB-VIB.

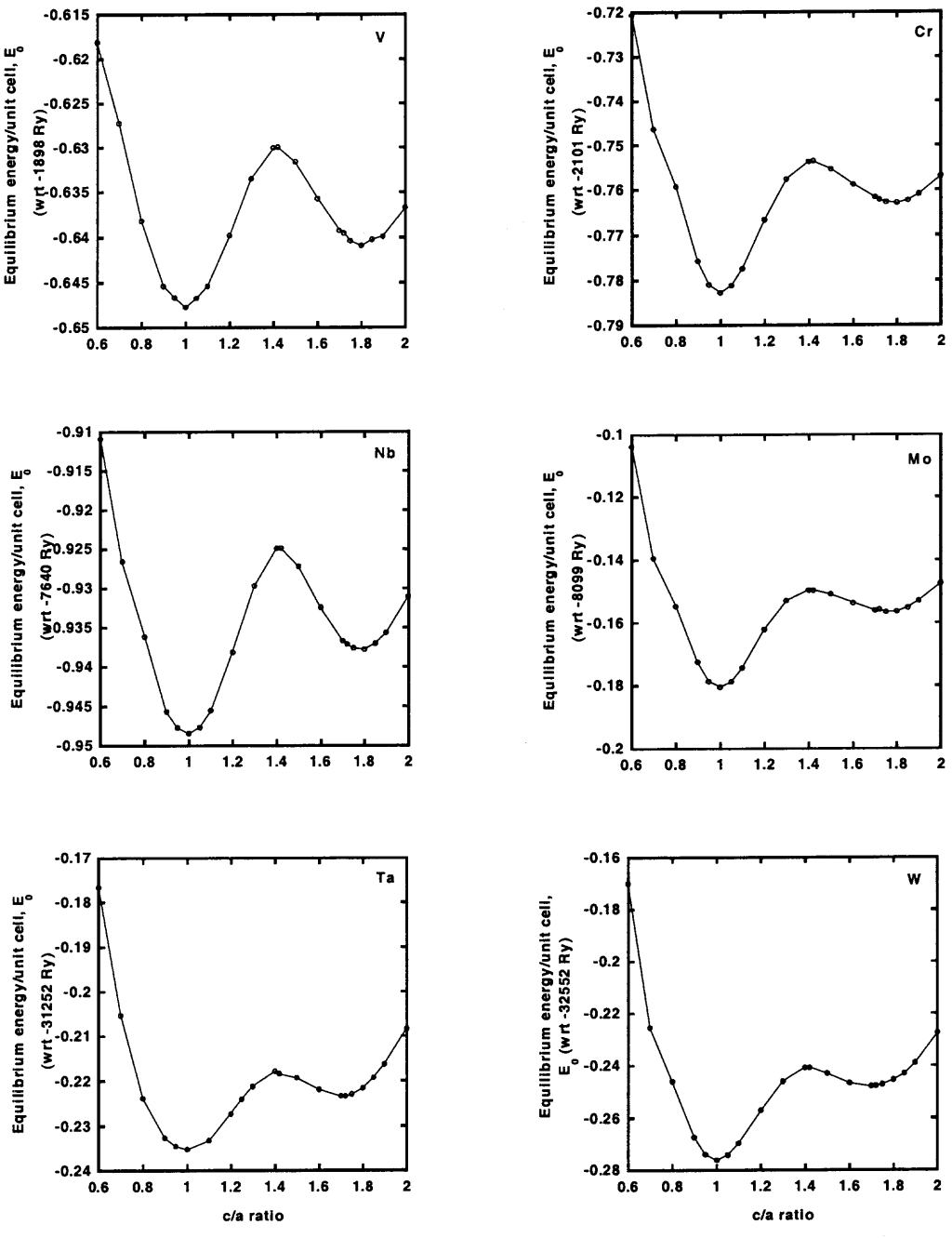


Figure 5.3 Clockwise from top-right, the EBPs of Cr, Mo, W, Ta, Nb and V represented as the total energy per unit cell plotted against the c/a ratio.

A saddlepoint at c/a ~ 1.41 can be observed in Figure 5.3 for all the cubic metals, corresponding to theoretical fcc phases of the metals, while the bcc phases occur at a c/a ratio of 1 and the bct phases at c/a ratios greater than $\sqrt{2}$. The c/a ratios at which bct phases are formed as well as the energy differences, $\Delta E_0^{bct-bcc}$, between corresponding bcc and bct phases, are listed in Table 5.1. The lattice parameters of the predicted bct phases are also shown in Table 5.1. The calculated bct equilibrium volume increases slightly (always by 2%) over the calculated bcc equilibrium volume for all the metals. The predicted bcc lattice parameters agree quite well with experiment [194], differing by 0.6% on average from the data.

Table 5.1 Calculated c/a Ratios, Equilibrium Lattice Parameters and Total Energies Per Unit Cell of Metastable and Stable Phases of Studied Cubic Metals and Phase Energy Differences $\Delta E_0^{bct-bcc}$

Metal	a_{bcc} **(Å)**	a_{bct} **(Å)**	**Bct** c/a **Ratio**	E_0^{bcc} **(Ry)**	E_0^{bct} **(Ry)**	$\Delta E_0^{bct-bcc}$ **(eV)**
Ta	3.3246 (3.3058)*	2.7894	1.72	-31252.2353	-31252.2234	0.162
Nb	3.3110 (3.3004)	2.7402	1.80	-7640.9485	-7640.9378	0.146
V	2.9978 (3.0240)	2.4747	1.80	-1898.6478	-1898.6409	0.094
W	3.1865 (3.1647)	2.6895	1.70	-32332.2762	-32332.2480	0.384
Mo	3.1625 (3.1473)	2.6429	1.75	-8099.1806	-8099.1566	0.326
Cr	2.8475 (2.8839)	2.3571	1.80	-2101.7828	-2101.7629	0.271

* Source: R. W. G. Wyckoff, *Crystal Structures*, (John Wiley and Sons, New York, 1963).

Within the studied groups, the magnitude of the c/a ratio at which a bct phase is formed is inversely proportional to the atomic number Z, while $\Delta E_0^{bct-bcc}$ is proportional

to Z. In each studied period, bct phases formed from group 5B elements had the same or higher c/a ratios than bct phases formed from group 6B elements. The $\Delta E_0^{bct-bcc}$ values calculated for the group 6B elements were all greater than those calculated for the group 5B elements. The $\Delta E_0^{bct-bcc}$ values for all the bcc metals ranged from 0.09 eV (vanadium) to 0.38 eV (tungsten).

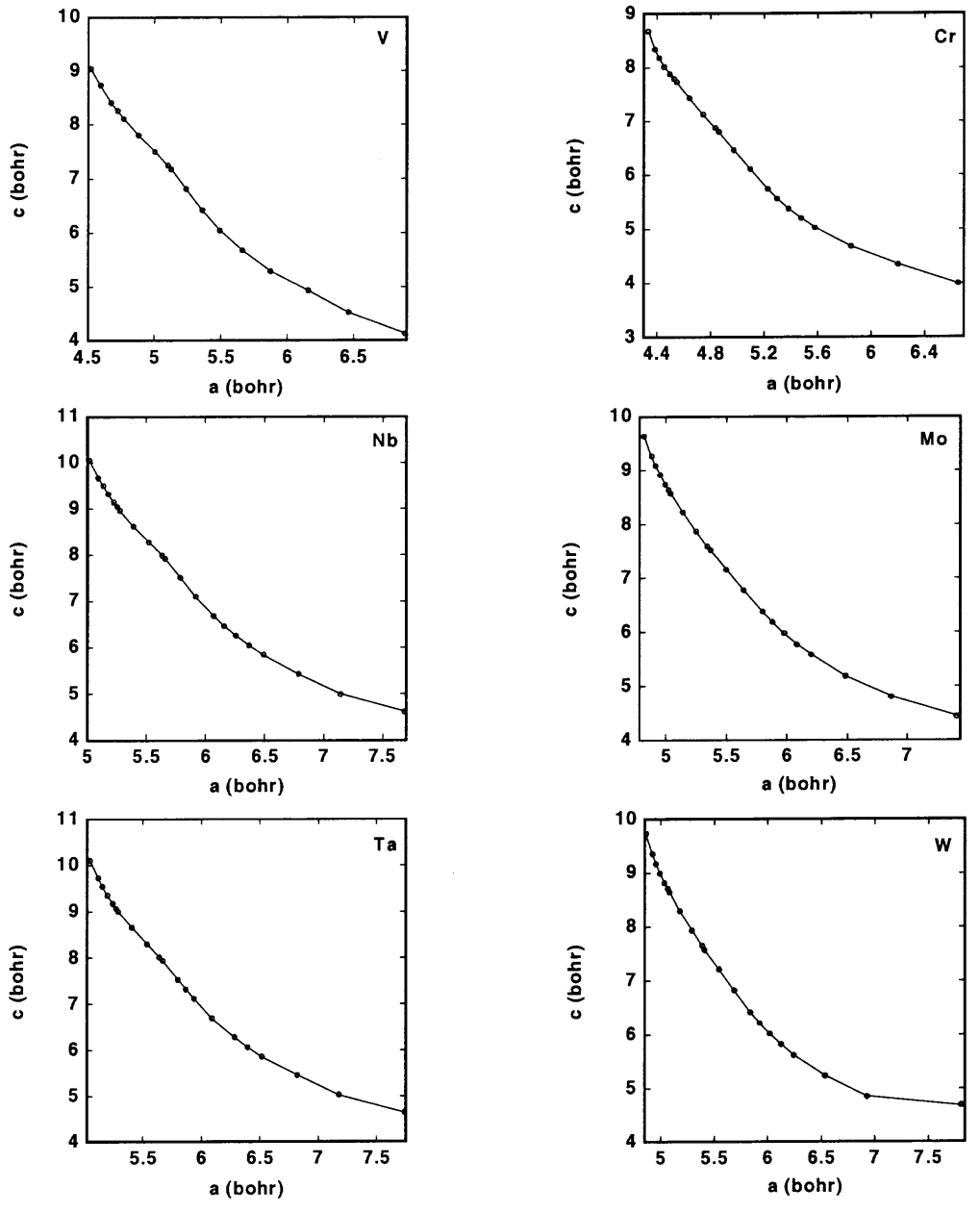


Figure 5.4 Clockwise from top-right, the EBPs of Cr, Mo, W, Ta, Nb and V represented as the c lattice parameter vs. the a lattice parameter, each set of parameters (a,c) belonging to the minimum energy structure at a given c/a ratio.

The total energy per unit cell as a function of the unit cell volume at a c/a ratio of 1, which corresponds to the bcc phase, is shown for each metal in Figure 5.5.

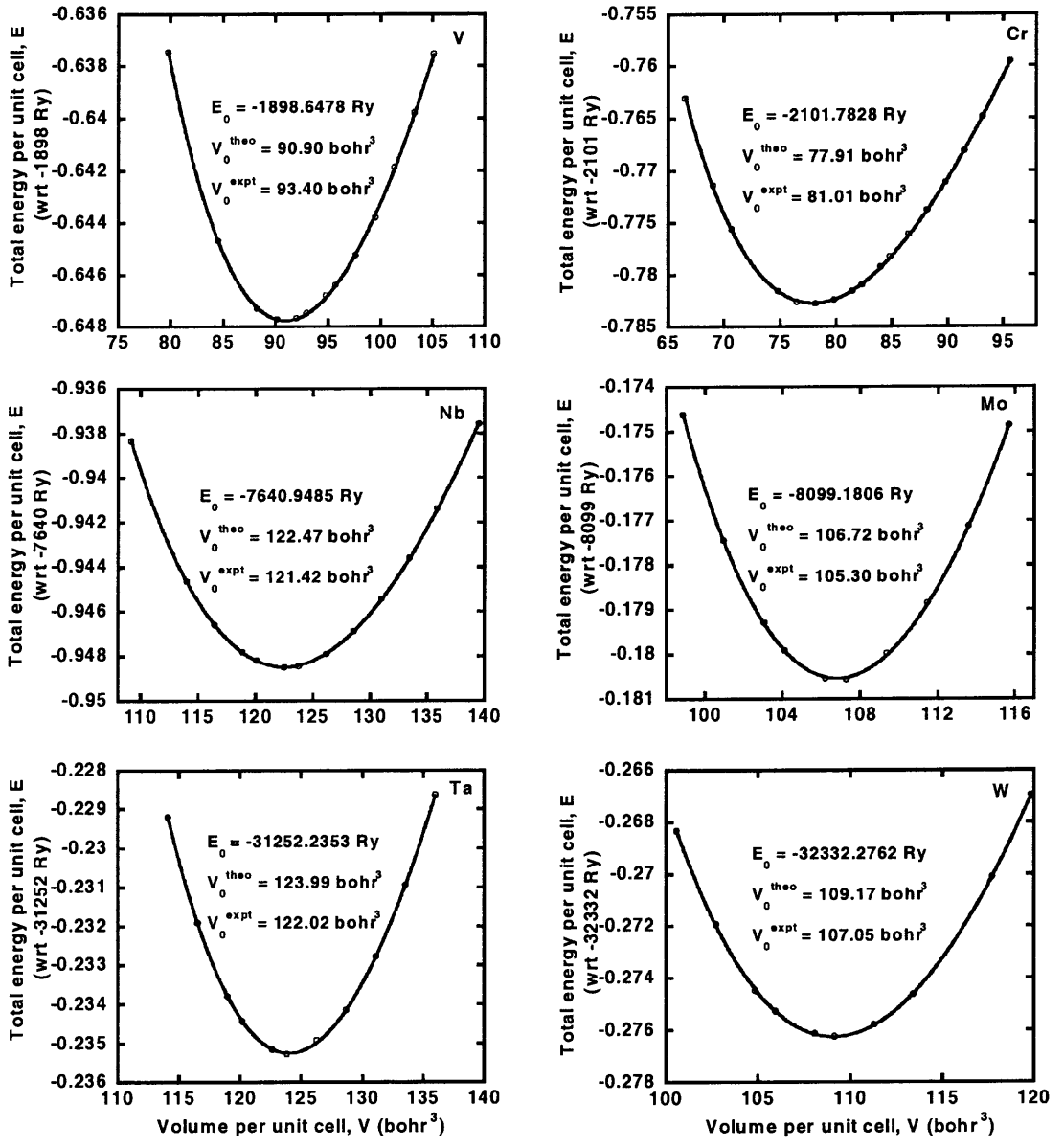


Figure 5.5 The total energy per unit cell E as a function of unit cell volume V for the bcc group 5B elements Ta, Nb and V, and the bcc group 6B elements W, Mo and Cr. The total energies shown for Ta, Nb and V are with respect to –31252 Ry, -7640 Ry and –1898 Ry respectively. The total energies shown for W, Mo and Cr are with respect to –32332 Ry, -8099 Ry and –2101 Ry respectively. The open circles in each of the figures represent the LAPW calculated energies and the solid curve is the fit to the Murnaghan equation. The experimental volumes V_0^{expt} are calculated according to $V_0^{expt} = a^3/2$ using lattice parameters taken from Reference [194].

The calculated bcc theoretical volumes differ from experiment by +2% (2% greater than the experimental volume) for Ta, +1% (Nb), -3% (V), +2% (W), +1% (Mo) and -4% (Cr). The total energy per unit cell as a function of the unit cell volume is shown in Figure 5.6 for each metal at c/a ratios of 1.72(Ta), 1.8(Nb), 1.8(V), 1.7(W), 1.75(Mo) and 1.8(Cr), for which the minimum in energy E_0 in the (V,E) plane corresponds to a local minimum in the $(c/a,E_0)$ plane and therefore to a metastable bct phase.

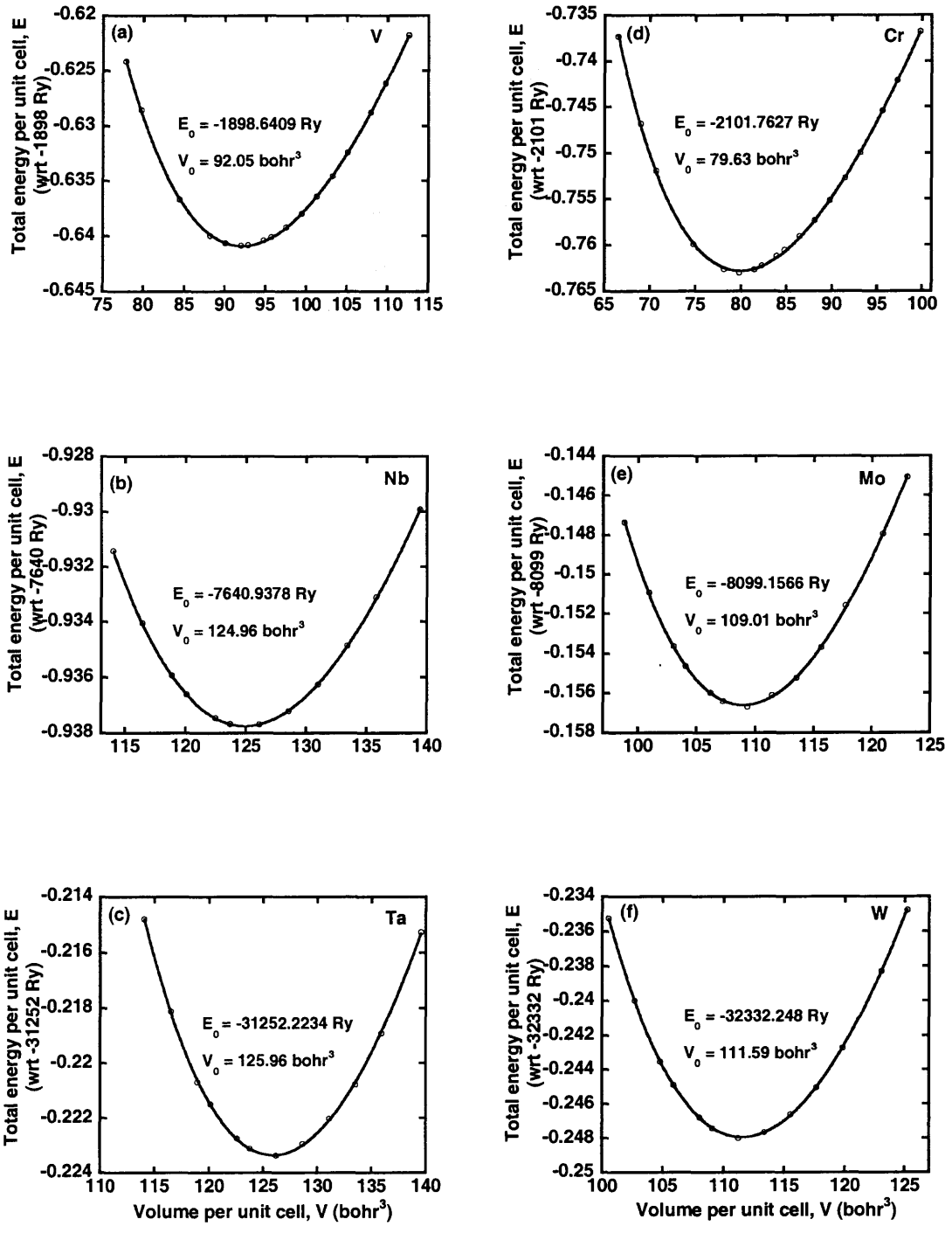


Figure 5.6 The total energy per unit cell E as a function of volume V for the predicted bct phases of the group 5B elements Ta, Nb and V, and the predicted bct phases of the group 6B elements W, Mo and Cr. The total energies shown for bct Ta, Nb and V are with respect to –31252 Ry, -7640 Ry and –1898 Ry respectively. The total energies shown for bct W, Mo and Cr are with respect to –32332 Ry, -8099 Ry and –2101 Ry respectively. The open circles in each of the figures represent the LAPW calculated energies and the solid curve is the fit to the Murnaghan equation.

5.4.3 The Bcc-Fcc Energy Difference and the Tetragonal Shear Constant

The tetragonal shear constant is given by Equation (5.2.1.17). Its magnitude for the transition metals has been determined [195] to be directly proportional to the energy differences between their bcc and fcc phases. This is illustrated in Figure 5.7.

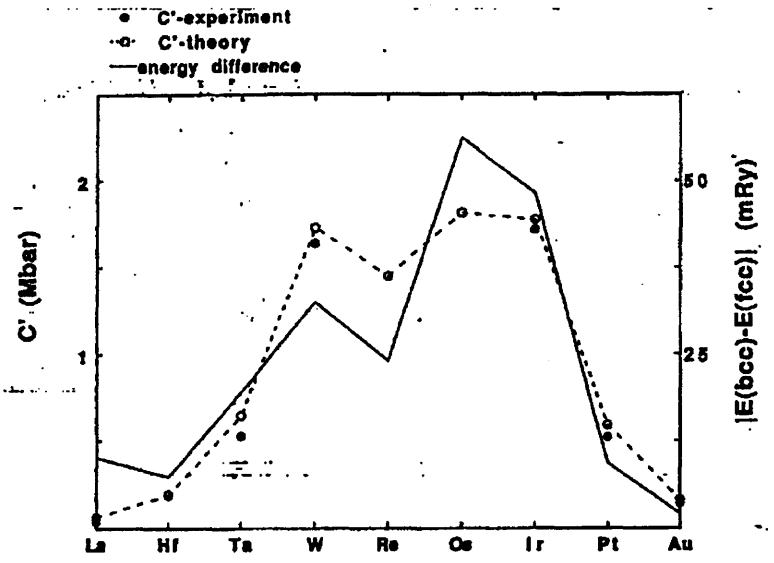


Figure 5.7 Theoretically calculated bcc-fcc energy differences are shown together with experimental and calculated values of the tetragonal shear constant (taken from Reference [195]).

The calculated lattice constants for the bcc and fcc phases of all the studied metals as well as the fcc-bcc energy differences are given in Table 5.2.

Table 5.2 Calculated Lattice Constants for Bcc and Fcc Phases of Metals in Groups VB-VIB of the Periodic Table and Phase Energy Differences

Metal	a_{bcc} (Å)	a_{fcc} (Å)	$\Delta E_0^{fcc-bcc}$ (eV)	ΔT (K)
Ta	3.3246	5.6403	0.229	2657
Nb	3.3110	5.6301	0.322	3737
V	2.9978	5.1003	0.242	2808
W	3.1865	5.3879	0.482	5593
Mo	3.1625	5.3409	0.420	4894
Cr	2.8475	4.8384	0.396	4595

From Table 5.2, it can be seen that Ta exhibits the lowest energy difference $\Delta E_0^{fcc-bcc}$ between the bcc and fcc phases, while tungsten exhibits the highest. The fcc-bcc energy difference is directly proportional to the magnitude of the tetragonal shear constant C'.

In order to calculate C', small curvatures of the total energy per unit cell as a function of the c/a ratio around a c/a ratio of 1 are used. The energies as a function of c/a ratio are fitted to a second order polynomial in δ_1, where δ_1 is given by Equation (5.2.1.14). The results are shown for tungsten in Figure 5.8.

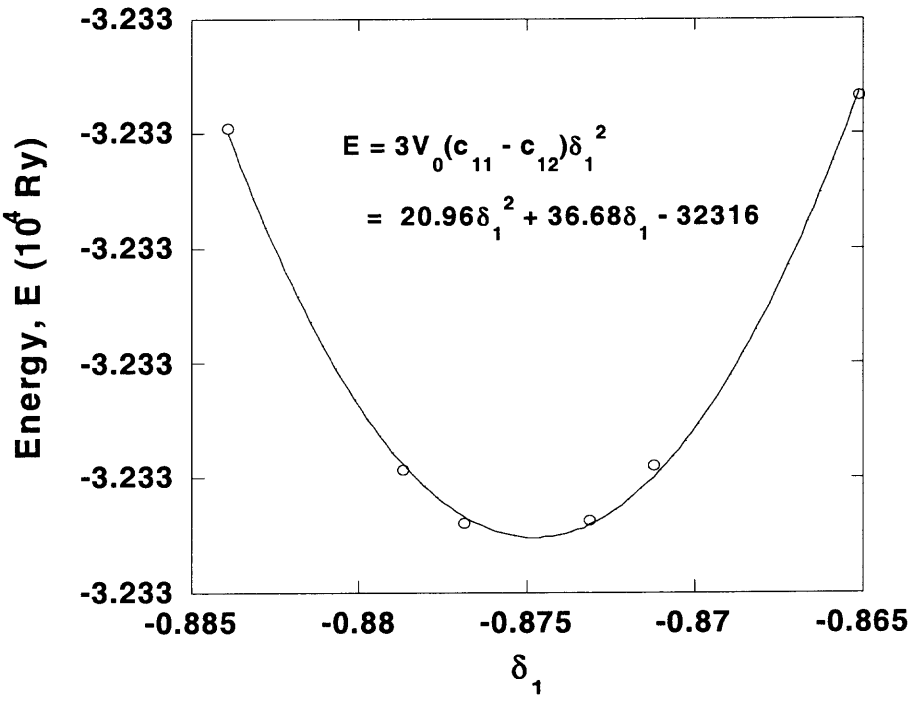


Figure 5.8 A plot of total energy per unit cell vs. δ_1 for W showing the fit to a quadratic equation used to obtain the shear constant.

A value for the shear constant of 235.9 GPa is obtained for W from this figure. The shear constant values are calculated for the rest of the studied metals (Ta, Nb, V, Cr, Mo) in the same manner at T = 0 K and are given in Table 5.3. The bulk moduli calculated at T = 0 K for all the studied metals are also shown in Table 5.3. Experimental values of the shear constant measured [196] at high temperatures of the same order of magnitude as would be encountered in thin film sputter-deposition are shown in parenthesis next to the calculated shear constants in Table 5.3. Among all the studied Group 5 and Group 6 metals, Ta fcc phases are the most often reported in the thin film literature, suggesting that they are the most easily formed.

Table 5.3 Calculated Bcc Phase Shear Constants and Bulk Moduli for the Bcc, Fcc and Bct Phases of the Studied Metals

Metal	Bulk Modulus B (GPa)	Shear Constant C' (GPa)	B (GPa) Fcc	B (GPa) Bct
Ta	195.99(200)‡	77.24 (65.3, 773K)	189.59	189.80
Nb	171.15(170)‡	90.33(39.5, 763K)	161.41	164.84
V	167.86(160)‡	90.70(47.4, 300K)	158.41	157.35
W	299.94(310)‡	235.90(152.2, 773K)	282.02	284.57
Mo	259.83(230)‡	228.26(113.1, 973K)	240.14	237.53
Cr	245.88(160)‡	240.32(113.2, 500K)	255.48	203.28

‡Source: G.V. Samsonov (Ed.) in *Handbook of the physicochemical properties of the elements*, (IFI-Plenum, New York, 1968).

This would seem to be consistent with the results in Table 5.3 showing that the lowest value of C' is calculated for Ta, because lower shear constants permit easier rearrangements of atoms, which would allow the bcc to fcc transformation, but it is not borne out by the high-temperature shear constant data. The reasons for the comparatively low value of C' calculated for Ta are rooted in electronic structure considerations, which are discussed in following sections.

5.5 Electronic Structure Dependence of Elastic Properties

It has long been known that the occurrence of a pseudogap feature separating hybrid bands of bonding and anti-bonding character in the density of states (DOS) figure of pure materials and alloys is a strong indicator that the material is structurally stable [197]. A pseudogap feature maximizes the filling of states having bonding character, while leaving the anti-bonding states unoccupied. In the most stable structures, all the valence electrons can be accommodated in the bonding bands thereby bringing the Fermi level into a

valley-like position separating bands of bonding and anti-bonding character and increasing the bonding strength. A low value of the density of states at the Fermi level therefore indicates an energetically favorable atomic arrangement in a crystal structure, and thereby a stable structure.

The Fermi level occurs near a minimum in the DOS for the studied Group VIB bcc metals, while it occurs to the left of the DOS minimum in the group VB bcc metals. Consequently, the Group VIB metals have higher bulk moduli than the group VB metals in the same period. Elastic properties are also critically affected by electronic structure. The bulk modulus exhibits a parabolic trend across the transition series of the periodic table. An increase in the bulk modulus across the transition series has been demonstrated to be due to the progressive filling of bands of d character. The bulk modulus begins to decrease when the Fermi energy increases to the extent that the bonding states are completely filled and the anti-bonding bands begin to be filled. The degree of filling of d bands of anti-bonding character is directly responsible for the decrease in the values of the bulk modulus across the transition series. The d band effect is also important in determining the values of the shear constant, which will be studied to investigate the issue of the tendency towards hexagonal phase formation in the studied bcc metals.

5.6 Densities of States

The density of states (DOS), $N(\varepsilon)$ is given by [178]

$$N(\varepsilon) = \sum_{nk} \delta(\varepsilon - \varepsilon_{nk}), \tag{5.6.1}$$

while the Fermi energy ε_F is defined [178] as the energy at which the integrated density of states $I(\varepsilon)$ defined as

$$I(\varepsilon) = \int_{-\infty}^{\varepsilon} N(\varepsilon) d\varepsilon, \tag{5.6.2}$$

equals the number of valence electrons. The DOS diagram of a material is a rich source of information concerning the electronic conductivity behavior of the material and is thus a characteristic property of any given quantum-mechanical system. The availability and degree of mobility of charge carriers in a material can be inferred by studying the DOS and its attributes such as the DOS at the Fermi energy and the bandwidth of different bands. A particularly important attribute is the DOS at the Fermi energy $N(\varepsilon_F)$, which appears in all theoretical expressions for the resistivity. The density of electron states $N(\varepsilon_F)$ at the Fermi level per atom and per spin is [178]

$$N(\varepsilon_F) = \frac{V}{8\pi^3} \int_{S_n} \frac{dS}{|\nabla_k \varepsilon_{nk}|}, \tag{5.6.3}$$

therefore $N(\varepsilon_F)$ is a Fermi surface average of $1/v$, where v is the electronic velocity. $N(\varepsilon_F)$ is inversely proportional to the average velocity on the Fermi surface.

The densities of states (DOS) as functions of energy, $N(\varepsilon)$, have been calculated for all the studied bcc metals, including the calculated metastable bct phases of these metals. From the $N(\varepsilon)$ vs. ε figures, the total and angular momentum decomposed densities of states at the Fermi energy, $N(\varepsilon_F)$ and $N_l(\varepsilon_F)$ respectively, were computed

for all phases. The total and d DOS figures for the bcc and bct phases are shown in Figures 5.9(a) – 5.9(c).

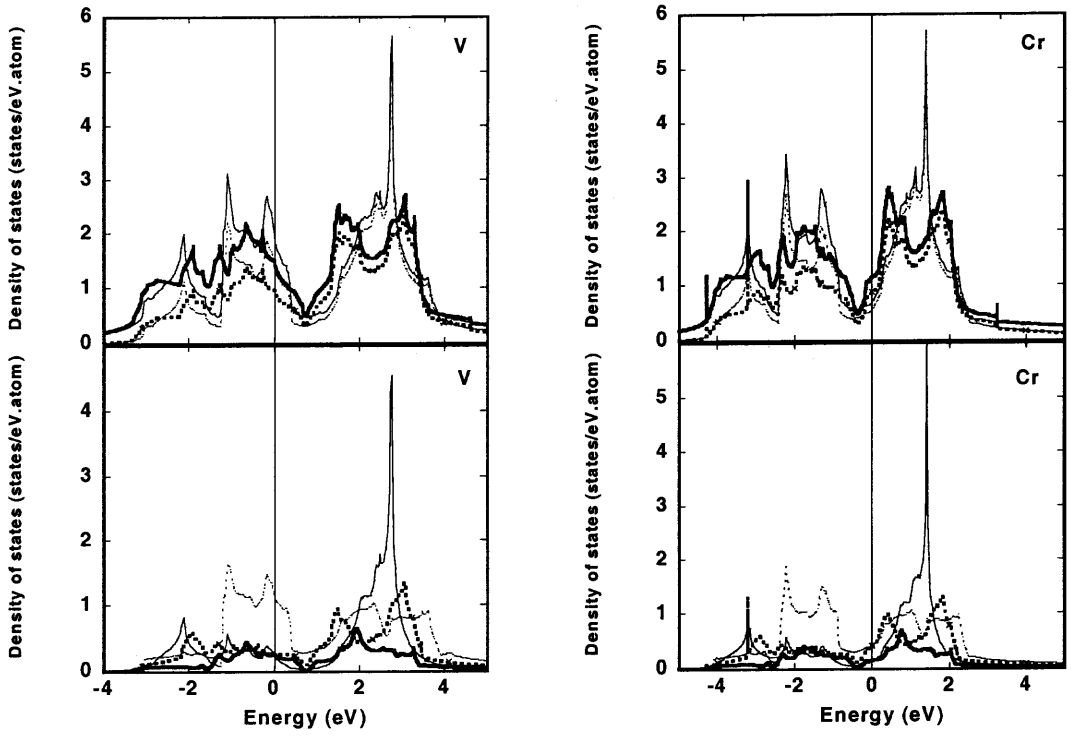


Figure 5.9(a) The densities of states as functions of energy are shown for the cubic and the tetragonal phases of the elements V and Cr. The Fermi energy is at 0 eV. In the upper panel of each figure, the thickened full lines and dashed lines represent respectively the total and d densities of states for the tetragonal phase. The thin full and dashed lines represent respectively the total and d densities of states for the cubic phase. In the lower panel of each figure, the thickened full and dashed lines represent respectively, the d_{z^2} and the d_{xz+yz} densities of states. The thin full and dashed lines represent respectively the d-e_g and the d-t_{2g} densities of states.

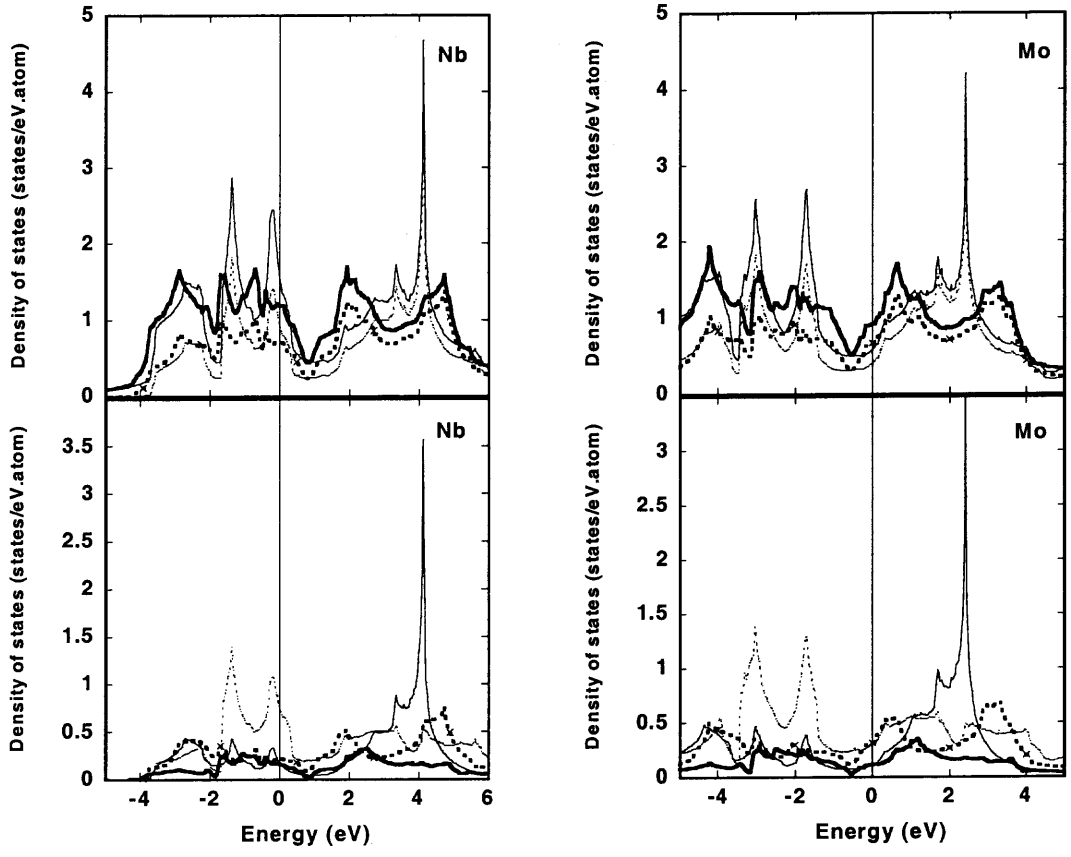


Figure 5.9(b) The densities of states as functions of energy are shown for the cubic and the tetragonal phases of the elements Nb and Mo. The Fermi energy is at 0 eV. In the upper panel of each figure, the thickened full lines and dashed lines represent respectively the total and d densities of states for the tetragonal phase. The thin full and dashed lines represent respectively the total and d densities of states for the cubic phase. In the lower panel of each figure, the thickened full and dashed lines represent respectively, the d_{z^2} and the d_{xz+yz} densities of states. The thin full and dashed lines represent respectively the d-e_g and the d-t_{2g} densities of states.

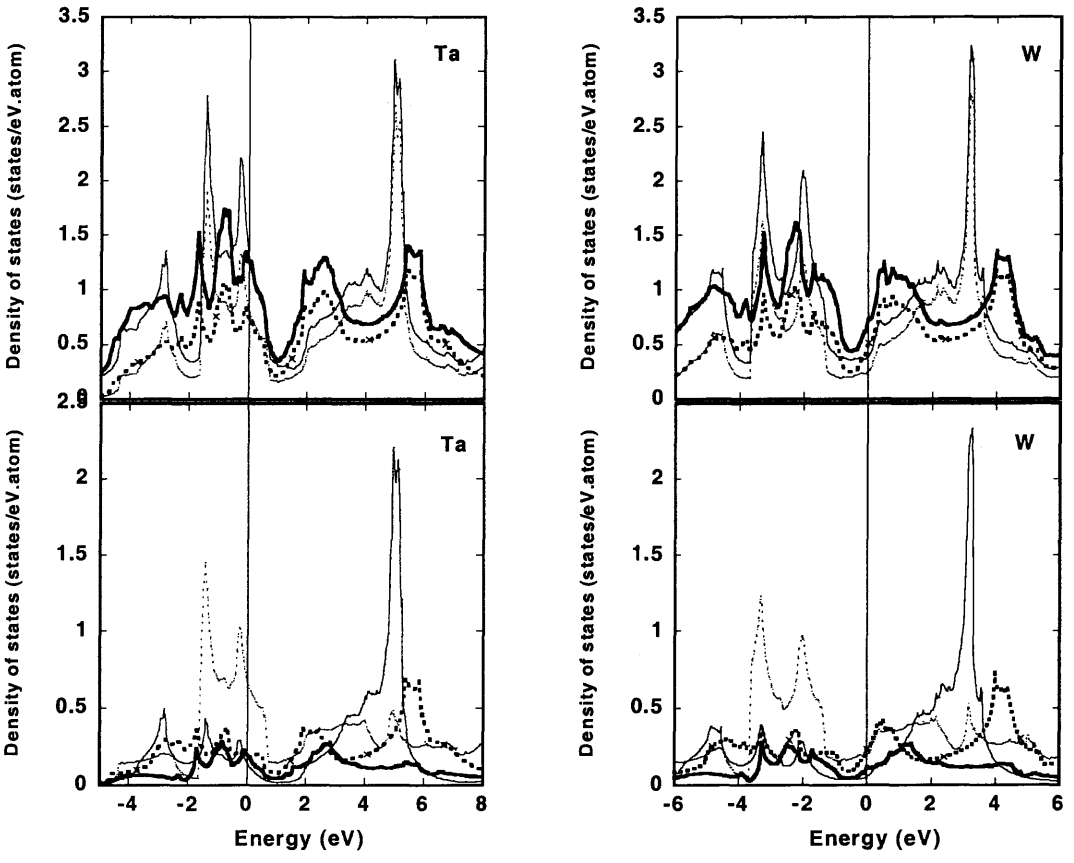


Figure 5.9(c) The densities of states as functions of energy are shown for the cubic and the tetragonal phases of the elements Ta and W. The Fermi energy is at 0 eV. In the upper panel of each figure, the thickened full lines and dashed lines represent respectively the total and d densities of states for the tetragonal phase. The thin full and dashed lines represent respectively the total and d densities of states for the cubic phase. In the lower panel of each figure, the thickened full and dashed lines represent respectively, the d_{z^2} and the d_{xz+yz} densities of states. The thin full and dashed lines represent respectively the d-e_g and the d-t_{2g} densities of states.

5.6.1 Bcc Phase DOS

The calculated band properties, such as $N(\varepsilon_F)$ and $N_d(E_F)$ for all the studied bcc metals are listed in Table 5.4.

Table 5.4 Calculated Total Density of States and Total d Density of States at the Fermi Energy, $N(\varepsilon_F)$ and $N_d(E_F)$, for Bct and Bcc Phases of the Studied Metals

Metal	c/a Ratio	$N(E_F)$ (States/eV.atom)	$N_d(E_F)$ (States/eV.atom)
Ta	1.00	1.34	0.83
	1.72	1.33	0.75
Nb	1.00	1.56	0.93
	1.80	1.26	0.67
V	1.00	1.85	1.32
	1.80	1.47	0.97
W	1.00	0.41	0.27
	1.70	0.74	0.48
Mo	1.00	0.59	0.38
	1.75	0.93	0.65
Cr	1.00	0.66	0.53
	1.80	1.07	0.86

Proceeding down the periods and also to the right across the groups of the studied block of the Periodic Table, $N(\varepsilon_F)$ was seen to decrease for the metals. The highest value of $N(\varepsilon_F)$ among the bcc elements was obtained for V, and the lowest value was obtained for W. The d densities of states, $N_d(\varepsilon_F)$, were responsible for the largest contribution to $N(\varepsilon_F)$ for all the studied bcc elements, and also followed the same trend as $N(\varepsilon_F)$. The cubic symmetry splits the d states further into the $d\text{-}e_g$ and $d\text{-}t_{2g}$ irreducible representations. There is a single narrow peak due to the $d\text{-}e_g$ states located in the upper conduction band region of all the DOS figures (Figures 5.9(a) – 5.9(c)), which broadens down the periods of the studied Periodic Table block (*i.e.* as the atomic

number Z and the lattice constant a increase). The conduction d bandwidth correspondingly increases down the periods. A double peak feature due to the d-t_{2g} states is also seen near the top of the valence band. For the group 5B elements, the double peak feature extends through ε_F. The double peak feature does not extend through ε_F for the group 6B elements, which occurs at a minimum in the DOS for these elements. The height ratio of the d-e_g single peak feature to the d-t_{2g} double peak feature is highest for V and Cr, and decreases down the periods. The d-t_{2g} states are responsible for the largest contribution to $N_d(\varepsilon_F)$ for all the studied bcc metals, slightly more so for the group 5B than the group 6B metals (Table 5.5). The s and p total densities of states were insignificant compared to the d and have therefore not been included in the figures.

Table 5.5 Total d Densities of States at the Fermi Energy, $N_d(E_F)$, and Symmetry Decomposed Components of $N_d(E_F)$ for Bcc and Bct Phases of the Studied Metals

Metal	$N_l(E_F)$							
	$d_{bcc}^{\ tot}$	d_{e_g}	$d_{t_{2g}}$	$d_{bct}^{\ tot}$	d_{z^2}	$d_{x^2-y^2}$	d_{xy}	$d_{xz}+d_{yz}$
Ta	0.83	0.14	0.69	0.75	0.20	0.31	0.06	0.18
Nb	0.93	0.17	0.75	0.67	0.15	0.24	0.08	0.19
V	1.32	0.22	1.10	0.97	0.24	0.38	0.09	0.27
W	0.27	0.05	0.22	0.48	0.11	0.07	0.09	0.21
Mo	0.38	0.11	0.27	0.65	0.11	0.10	0.13	0.31
Cr	0.53	0.12	0.41	0.86	0.16	0.15	0.13	0.43

5.6.2 Bct Phase DOS

The calculated $N(\varepsilon_F)$ for the bct phases derived from the group 5B bcc elements were consistently lower than those computed for their corresponding bcc phases. The reverse relationship to the bcc phase $N(\varepsilon_F)$ values was seen for the $N(\varepsilon_F)$ calculated for bct phases derived from the group 6B bcc elements. The trend in the computed $N(\varepsilon_F)$ observed for the bcc phases was mirrored by the trend in calculated $N(\varepsilon_F)$ for the bct phases, except that the calculated $N(\varepsilon_F)$ for bct Nb is less than that computed for bct Ta. The calculated $N_d(\varepsilon_F)$ for the bct phases followed exactly the same trend as the calculated $N(\varepsilon_F)$. ε_F was observed to increase for all the predicted bct structures relative to their corresponding bcc structures.

The total and d DOS figures for the bct phases are also shown in Figures 5.9(a) – 5.9(c). In the DOS figures for the bct phases, the $d\text{-}e_g$ and $d\text{-}t_{2g}$ states are split further into the two d_{z^2} and $d_{x^2-y^2}$ irreducible representations, and the three d_{xy}, d_{xz} and d_{yz} irreducible representations respectively by the tetragonal symmetry. The contributions to the DOS from the $d\text{-}e_g$ derived states have a double peak structure and are flattened relative to the sharp $d\text{-}e_g$ peak seen in the bcc DOS figures. The contributions to the bct DOS from the $d\text{-}t_{2g}$ derived states are also flattened relative to the double $d\text{-}t_{2g}$ peak seen in the bcc DOS figures and have a more complex peak structure. The $d\text{-}e_g$ derived states are responsible for the largest contribution to $N_d(\varepsilon_F)$ for the group 5B bct phases, while the $d\text{-}t_{2g}$ derived states are responsible for the largest contribution to $N_d(\varepsilon_F)$ for the group 6B bct phases (Table 5.5). The complex set of d bands derived from the $d\text{-}e_g$

states extend from the top of the valence band through ε_F to the bottom of the conduction band. There is a double peak feature due to the d-t_{2g} derived states extending through the conduction band. Significant contributions to the DOS from the d-t_{2g} derived states can also be seen in the lower part of the valence band region. The ratio of the height of the d-e_g derived feature to the height of the d-t_{2g} derived double peak feature increases down the periods, being lowest for bct V and bct Cr and highest for bct Ta and bct W. The ε_F for the bct phases derived from the group 5B elements, similar to the ε_F in their corresponding bcc phases, occur in a relatively high density of d bands. The ε_F for the bct phases derived from the group 6B elements occur in a higher density of d bands relative to the ε_F in the corresponding bcc phases.

The bcc metals in this study span periods 4, 5 and 6, and groups 5B – 6B of the Periodic Table. The calculated total densities of states, $N(\varepsilon_F)$, for the bcc metals followed a decreasing trend when proceeding down the periods and moving rightwards across the groups of the studied block of the Periodic Table. The magnitude of $N(\varepsilon_F)$ is inversely proportional to the d bandwidth and directly proportional to the d band energy. It also follows a trend inversely proportional to the magnitude of the atomic number Z. The order of the bcc $N(\varepsilon_F)$ values down the periods is due to the d bandwidth, which increases in this direction. The order of the bcc $N(\varepsilon_F)$ values across the groups is due to the decrease in the d band energy caused by band filling across the groups. The dominant effect of the open shell d occupancy is expected because from Table 5.4, the d densities of states were responsible for the largest contribution to

$N(\varepsilon_F)$ for all the studied bcc metals. The calculated $N(\varepsilon_F)$ for the bct phases derived from the group 5B bcc elements were consistently lower than those calculated for their corresponding bcc phases. The reverse relationship to the bcc phase $N(\varepsilon_F)$ values was seen for the $N(\varepsilon_F)$ calculated for bct phases derived from the group 6B bcc elements. The distortion of the bcc structure to form the bct structure causes ε_F to shift upward in energy for all the metals studied, while leaving the structure of the DOS in the immediate vicinity of ε_F essentially unaltered. ε_F is located to the left of the DOS minimum in the group 5B bcc metals while it is located to the right of the DOS minimum in the group 6B bcc metals. The upward shift in ε_F due to the bct transformation therefore shifts it closer to the DOS minimum for the group 5B metals, causing a reduction in $N(\varepsilon_F)$, while for the group 6B metals, ε_F is shifted away from the DOS minimum, increasing $N(\varepsilon_F)$.

Within each studied group, the magnitude of the c/a ratio at which a bct phase is formed from the bcc phase is inversely proportional to Z, while the size of $\Delta E_0^{bct-bcc}$ is directly proportional to Z. Within each period, the group 6B elements formed bct phases with the same or lower c/a ratios than the group 5B elements, and had higher values of $\Delta E_0^{bct-bcc}$. This dependence of the c/a ratio on Z and the number of outer shell electrons is exactly the same noted earlier for $N(\varepsilon_F)$, while the dependence of $\Delta E_0^{bct-bcc}$ on these quantities is the exact opposite. This can be explained by noting that the percent change in the fraction of occupied d states having bonding character for the bcc to bct phase transition is appreciably greater (47 – 50%) for the group 5B elements than for the group 6B elements (3 – 19%). The conclusion can therefore be reached that within each studied

group and across each studied period, the magnitude of the bcc d bandwidth is inversely proportional to the c/a ratio at which a bct phase is formed from the bcc phase and also inversely proportional to $\Delta E_0^{bct-bcc}$. If the magnitude of $\Delta E_0^{bct-bcc}$ is taken as a measure of the degree of difficulty in forming a bct phase from the bcc structure, this suggests that heavier elements are less likely to form bct phases and that the bct phases derived from the group 6B elements are less readily formed from the corresponding bcc phases than bct phases derived from group 5B elements, because $\Delta E_0^{bct-bcc}$ increases appreciably across groups.

5.7 d -band Filling and the Fcc – Bcc Phase Energy Difference

From Table 5.2, it can be seen that Ta exhibits the lowest energy difference between the bcc and fcc phases. The temperature differences ΔT between the bcc and fcc phases, calculated according to

$$\Delta T = \Delta E_0^{fcc-bcc} / k_B \,, \tag{5.7.1}$$

are shown in Table 5.2. The lowest value of ΔT is obtained for Ta. The bcc - fcc structural energy difference is determined by the degree of d band filling in the bcc and fcc phases [195]. In order to understand the reasons for this, DOS figures are calculated for fcc and bcc phases of Ta, Nb and Mo. The d DOS figures for these phases are shown in Figure 5.10.

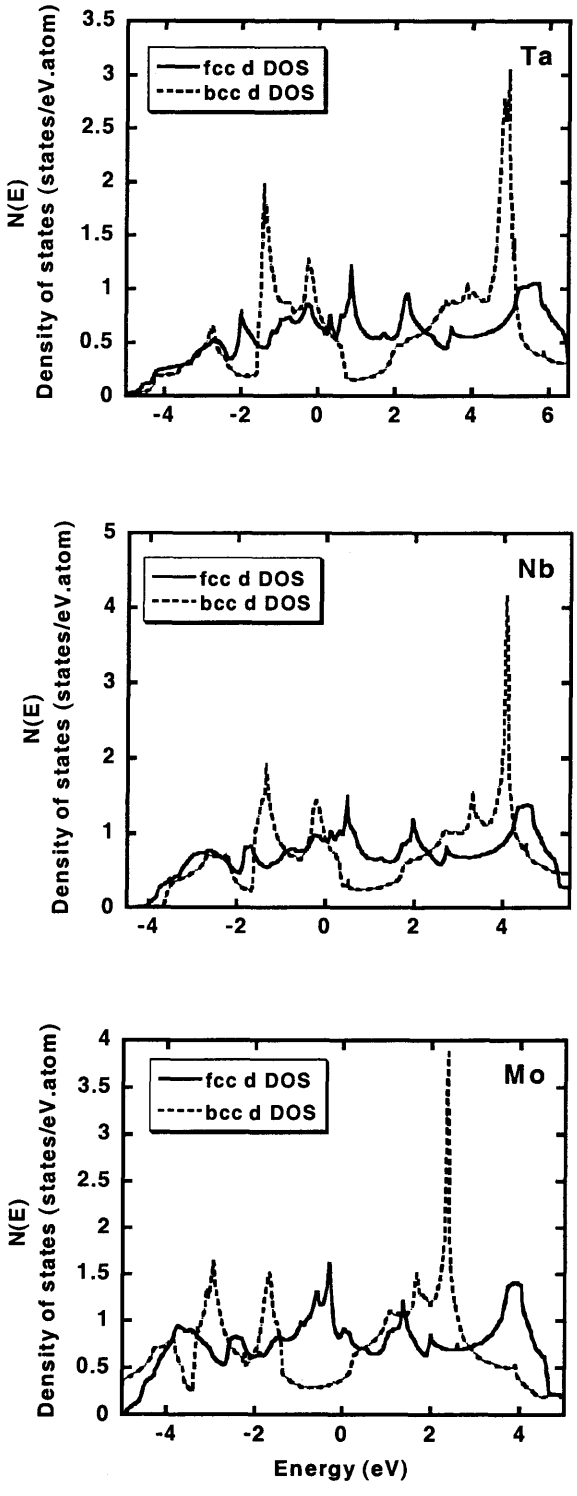


Figure 5.10 Total *d* DOS for the bcc and fcc ($c/a = \sqrt{2}$) phases of Ta, Nb and Mo. The dashed lines represent the bcc *d* DOS and the Fermi energy is at 0 eV.

The d occupancies N_d of the bcc and fcc phases of Ta, Nb and Mo are calculated from the DOS figures in Figure 5.10 according to

$$N_d = \int_{-\infty}^{\varepsilon_F} N_d(\varepsilon) d\varepsilon , \tag{5.7.2}$$

and are shown in Table 5.6. The magnitude of the fcc-bcc energy difference is proportional to the degree of d-band filling in the fcc phase relative to the bcc phase, $\Delta N_d^{fcc-bcc}$, given by

$$\Delta N_d^{fcc-bcc} = \int_{\varepsilon_{b,fcc}}^{\varepsilon_F} N_d^{fcc}(\varepsilon) d\varepsilon - \int_{\varepsilon_{b,bcc}}^{\varepsilon_F} N_d^{bcc}(\varepsilon) d\varepsilon , \tag{5.7.3}$$

where $\varepsilon_{b,fcc}$ and $\varepsilon_{b,bcc}$ are respectively, the bottom of the fcc and bcc valence bands. The fcc-bcc energy differences are also shown in Table 5.6.

Table 5.6 Total and Fractional d Occupancies for Ta, Nb and Mo Bcc and Fcc Phases

Metal	N_d **Bcc**	N_d **Fcc**	$\Delta N_d^{fcc-bcc}$
Ta	2.4722	2.2187	-0.2535
Nb	2.6282	2.5440	-0.0842
Mo	3.2645	3.6780	0.4135

The lowest value of the relative fcc to bcc d-band filling is calculated for Ta, while the highest value is calculated for Mo, with the Nb value falling between the Ta and Mo

values. The trend in values of $\Delta N_d^{fcc-bcc}$ mirrors the trend (Table 5.6) in the magnitudes of the fcc - bcc energy differences.

5.8 Correlation between the Bain and Burgers Transformations

The $P4_2/mnm$ structure, in which β-Ta is widely believed to crystallize, is a pseudo-hexagonal phase. Bearing this in mind, it is worthwhile to examine the similarities between the Bain path, which has been studied in detail in this chapter, and the most common mechanism that transforms a bcc phase to a hexagonal close packed (hcp) phase, namely the Burgers [198] transformation mechanism. The bcc metals in groups VB and VIB can undergo martensitic transitions such as the Burgers transformation, as well as the Bain transformation. As previously noted, the Bain transformation path can take a bcc structure to a bct structure, forming an fcc phase as an intermediate phase. The Bain transformation involves uniaxial strain along the [001] direction and the shear of bcc (101) planes in the $[\bar{1}01]$ direction so that the (111) fcc planes are parallel to the (101) bcc planes.

The Burgers transformation consists of shear of alternating (110) bcc planes in $[\bar{1}10]$ and $[1\bar{1}0]$ directions, followed by compression along the [001] direction and decompression along the $[\bar{1}10]$ and $[1\bar{1}0]$ directions. An intermediate fcc phase is also formed during the Burgers transformation. Both the Bain and Burgers transformation involve shear and uniaxial strain perpendicular to the shear direction. They both result in the formation of an intermediate fcc phase and can be described in a unified way as special cases of a more general transformation [199 – 200].

Bcc Ta transforms into an Ω phase [201 - 202] under shock loading at high pressures of approximately 45GPa. Ω phase formation also involves relative movement of atoms in adjacent (111) planes, leading to the collapse of the planes. Incomplete collapse of the (111) planes results in a trigonal structure belonging to space group $P\bar{3}m1$ while complete collapse leads to a hexagonal structure belonging to space group $P6/mmm$. The elements in group IVB all form hexagonal close packed (hcp) phases at room temperature. They can form the Ω phase by first forming the bcc phase at high temperatures, and then undergoing quenching to form the Ω phase or they can form the Ω phase directly at room temperature under high-pressure shock loading. The bcc and Ω phases can be described by a common hexagonal unit cell having three atoms per unit cell as shown in Figure 5.11.

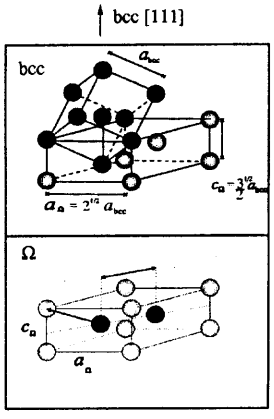


Figure 5.11 The common hexagonal cell describing the bcc and Ω structures. Open circles represent type A atoms, and closed circles represent type B atoms (taken from Reference [203]).

The atomic positions are $(0,0,0)$, $(1/3,2/3,1/3+z_\Omega)$ and $(2/3,1/3,2/3-z_\Omega)$. The z_Ω parameter describes a gradual transition from the bcc phase to the Ω phase. The bcc phase occurs when $z_\Omega = 0$, for $0 < z_\Omega < 1/6$ the structure has trigonal $(P\bar{3}m1)$

symmetry, and for $z_\Omega = 1/6$ the Ω phase results. The light atoms in the figure remain fixed, while the dark atoms move up in the bcc [111] direction.

The crystal structure sequence of the transition metals in the Periodic Table is hcp → bcc → hcp → fcc, and has been demonstrated [195] to be due to the degree of d band filling, specifically, that the increased filling of the d band from left to right of the Periodic Table results in the observed sequence of crystal structures. These crystal structures observed at ambient temperature and pressure can be modified by applying pressure. The d occupation in the hcp metals can be increased with pressure, causing them to form bcc phases [203]. This can explain the relative densities of bcc Ta and Ω Ta, which have densities of 16.6 and 13.64 g cm^{-3} respectively, since an increase in d occupation increases the number of tightly bound electrons and reduces atomic volume. It has been observed [204] however, that fcc metals such as Rh, Pd and Pt, which have a d band occupation larger than 7, form bcc structures under pressure, contradicting the argument of phase stabilization based on d band filling, since fcc structures would be expected to be more stable with respect to pressure, due to their high d occupation number. The phase transformation in this case was explained by the broadening of the semi-core p states, which reduced the $p-d$ band splitting between the p and the d bands, and resulted in $p-d$ band hybridization.

The experimental evidence shows that the most frequently observed fcc phase is that of Ta. As can be seen from Table 5.2, Ta exhibits the lowest energy difference $\Delta E_0^{fcc-bcc}$ between the bcc and fcc phases among all the studied metals. In addition to being directly proportional to the magnitude of the tetragonal shear constant C', which is

a measure of lattice resistance to shear, $\Delta E_0^{fcc-bcc}$ has been reported to be close to the energy difference between bcc and hcp phases $\Delta E_0^{hcp-bcc}$ for the transition metals [195, 205]. Accordingly, the energy differences between the fcc and hcp phases $\Delta E_0^{fcc-hcp}$ have been found to be about a fifth of $\Delta E_0^{fcc-bcc}$ [206] and the hcp and fcc phases lie very close in energy for the transition metals. The group IVB elements Hf, Zr and Ti are hcp at room temperature, but form the bcc phase at conditions of high temperature and pressure. This suggests that bcc metals might have a tendency towards forming a hcp phase at low pressures. In fact the fcc – bcc and the hcp – bcc energy differences for Ta are very close [205] at ambient conditions, as well as over a wide range of pressures above atmospheric pressure. Also lower shear constants reduce the energy barrier associated with atomic rearrangement. These suggest that under the low-pressure conditions present in vacuum sputtering or evaporation chambers, bcc Ta atomic clusters, more than atomic clusters of any of the other studied metals, would therefore be most likely to undergo phase changes directed towards hexagonal phase formation. Taken together, these results suggest a mechanism by which bcc Ta is transformed to the fcc phase at high temperature under strong reducing conditions. The β-Ta structure could then result from frustrated fcc to hcp phase transformation attempts by the already formed fcc Ta.

Under thermodynamically controlled thin film deposition conditions (low F/D ratio), the substrate should exert more of an influence on the resulting thin film crystalline structure. In this case, substrates with filled bands will generally be less reactive with film materials, but such substrates have lower surface energies, which cause surface segregation due to thermodynamic activation forces. Transition metal substrates

usually have higher surface free energies, because of their half filled *d* bands. Under thermodynamically controlled thin film deposition conditions, it is possible that hybridization of substrate bands with film material bands or impurity bands plays an important role in metastable phase formation.

When the growth of thin films on substrates is controlled by kinetics, the nature of the substrate is of less importance in determining the structure of the growing film. In addition, the energy barrier to atomic diffusion on the substrate is enhanced. This is illustrated in Figure 5.12.

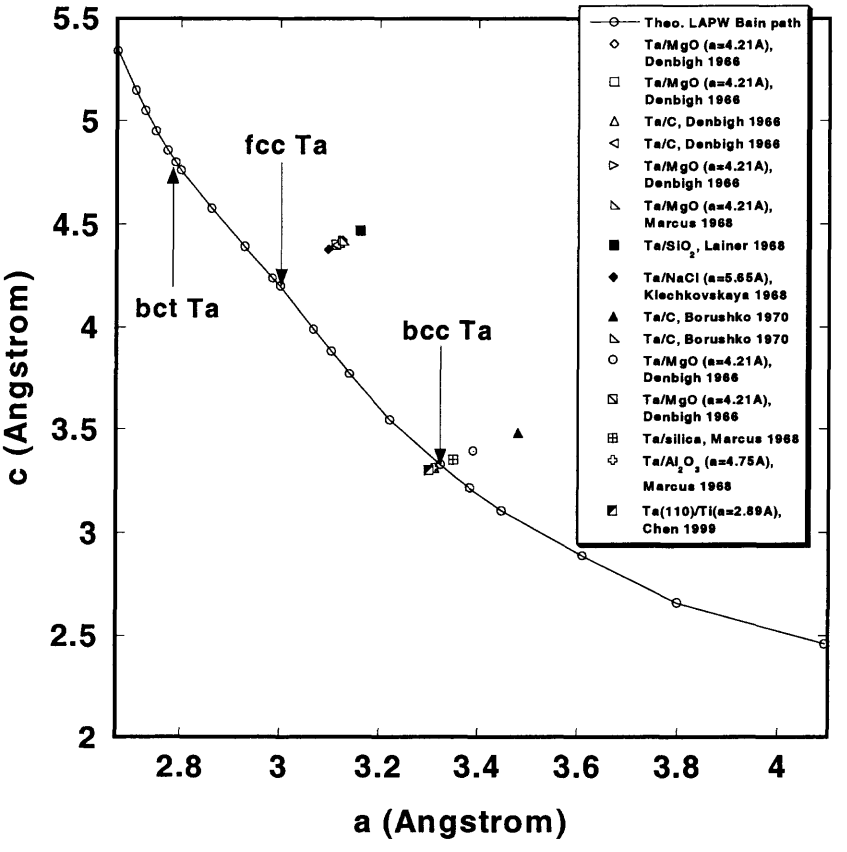


Figure 5.12 *a* and *c* lattice constants measured from Ta thin films grown on a variety of substrates, mainly by sputter deposition, superimposed on the theoretically calculated Bain path showing the equilibrium *c* lattice parameter as a function of the *a* lattice parameter.

Most of the data in Figure 5.12 comes from sputter deposition experiments, and from the figure, it can be seen that many of the prepared Ta films have lattice parameters that are expanded relative to bulk bcc Ta, and are therefore undergoing tensile stress. Since high deposition rates are frequently encountered in sputter deposition and reduction of the energy barrier encountered by diffusing atoms is believed to occur on compressively stressed substrates while the barrier increases on substrates undergoing tensile stress [207 - 209], high atomic deposition rates, which result in kinetically controlled deposition conditions, are shown by Figure 5.12 to lead to reduced substrate atomic diffusion, and therefore to increased likelihood of metastable phase formation.

In the next chapter the influence of the Bain transformation on the electrical resistivity and other transport properties is studied to determine whether the resistivity enhancement over the bcc phase, if any, in the resulting bct phases is of the same order of magnitude as seen from experimental β-Ta resistivity data.

CHAPTER 6

TRANSPORT PROPERTIES AND METASTABLE PHASES

The study of transport properties such as the electrical resistivity is important for the accurate characterization of metallic materials. The electrical resistivity of a reasonably pure material sample is an easily measured characteristic property of that material. It contains fundamental information about important phenomena such as electron-phonon and electron-electron interactions. Transport properties depend critically on electronic band structure properties such as the density of states and the mean square electronic velocity at the Fermi energy level. A great deal of effort has been devoted to theories with which the electrical resistivity can be analytically described and hence computed.

One of the first major breakthroughs in this regard were made over 70 years ago by Bloch [210], who wrote down the semi-classical Boltzmann equation describing electronic transport in metals. However, the electronic velocities and scattering probabilities occurring in the Boltzmann equation could be obtained only by a complete knowledge of the electronic structure, dispersion relation and electron-phonon scattering matrix elements, which were difficult to obtain accurately at the time. Important contributions towards the quantum theory of the resistivity were also made by Ziman [211 - 212], who developed analytic expressions grounded in quantum theory with which to describe the resistivities of pure metals and alloys. Theoretical work [212] by Ziman was able to demonstrate that the widely accepted Bloch-Grüneisen laws hitherto used to describe the electrical resistivity of materials applied only when certain special assumptions such as the Debye scattering approximation on a spherical Fermi surface had been made.

Further progress in the theoretical description of the resistivity was made [213 – 217] with the development of mathematically tractable expressions describing the electron-phonon and electron-electron scattering processes as well as scattering from non-spherical Fermi surfaces. The progress of these theories relied heavily on the availability of accurate experimental data with which computed results could be compared, while experimental work was guided to some extent by the need for experimentalists to verify predictions arising from the theories. Since the resistivity depends critically on electronic band structure properties, the successful implementations of these theories relied heavily on an accurate and reliable method for computing the electronic band structure. Such a reliable method is the LAPW method, which has been described in Chapter 4 and used to compute the density of electron states as a function of energy in Chapter 5 for the transition metals in the block of the periodic table defined by the groups 5B – 6B and the periods 4 - 6.

In this chapter, other band structure properties such as the mean square electronic velocity at the Fermi energy level will be computed for the same group of metals. These band structure properties will then be used to compute transport properties such as the electrical resistivity and the plasma frequency for the elements in this Periodic Table block and their metastable phases lying on the Bain path.

6.1 Electrical Resistivity

The electrical resistivity of a material is in most cases a readily measurable quantity and is proportional to the electrical resistance. The electrical resistance R of a thin film

material with length l, breadth b and thickness z to the flow of current is a function of the electrical resistivity ρ and the thickness of the film material and is given by

$$R = \rho l / zb . \tag{6.1.1}$$

When the thin film material is in the shape of a square, the resistance becomes

$$R = \rho / z . \tag{6.1.2}$$

This resistance is referred to as the sheet resistance, R_S, of the material. R_S can be evaluated by measuring the resistance of a rectangular film sample and then dividing the value obtained by the number of squares making up the rectangle.

The four-point probe method is very widely used for resistivity measurements. A typical four-point probe setup consists of four equally spaced (~ 1mm apart) spring-loaded tungsten metal tips. The springs allow the four metal tips to travel vertically up or down and also help to minimize sample damage. A high impedance current source is used to supply current through the outer two probes while a voltmeter is used to measure the voltage across the inner two probes in order to determine the sample resistivity.

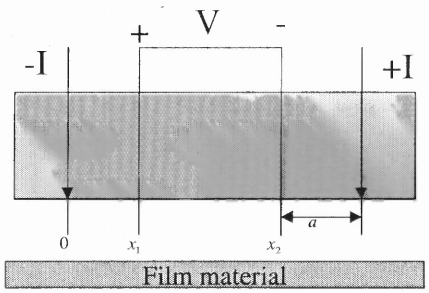


Figure 6.1 Schematic diagram of a four-point probe.

In a pure metal at room temperature (~300K), the electrical resistivity is governed mostly by conduction electron scattering due to lattice ion vibrations. Lattice vibrations result in waves whose energies are quantized in units called phonons. Electron scattering by lattice ion vibrations is proportional to the square of the amplitude of the fluctuations of the ions about their equilibrium positions, or sites, and hence on the concentration or density of states of the phonons. The wave number dependence of the angular frequency of vibration ω is termed the dispersion relation.

The simplest description of resistivity in terms of scattering processes, the Bloch-Grüneisen law, uses a model for the phonon mode density of states due to Debye [211]. The Debye model assumes a linear dispersion relation for the angular frequency of the waves arising from lattice vibrations of the form [218]

$$\omega = c_s k \,, \tag{6.1.3}$$

and defines a cutoff frequency ω_D represented as [218]

$$\omega_D = c_s \left(6\pi^2 N/V\right)^{1/3}, \tag{6.1.4}$$

which is associated with a cutoff wavevector in k space, K_D, given by [218]

$$K_D = \left(6\pi^2 N/V\right)^{1/3}, \tag{6.1.5}$$

and known as the Debye radius, where c_s is the velocity of sound, N is the total number of phonon modes with wavevector less than K_D, and V is the volume of a sphere of radius K_D. The Debye temperature Θ_D is given by [218]

$$\Theta_D = \hbar\omega_D / k_B \,. \tag{6.1.6}$$

In the Debye model, at temperatures above Θ_D, the phonon density of states is directly proportional to T and hence the resistivity is directly proportional to T, since it is directly proportional to the electron-phonon scattering rate. At temperatures below Θ_D, the lattice motion can still be decomposed into phonons, and electron-phonon scattering is still the dominant electron scattering mechanism through processes of the form [218 – 219]

$$\vec{k} + \vec{q} \Leftrightarrow \vec{k}', \tag{6.1.7}$$

where $\vec{k}$ is an electronic wave vector and $\vec{q}$ is a phonon wave vector. The average $\vec{q}$ is small at low temperatures, say 4K, therefore electrons are only scattered to small angles by phonons at these temperatures, leading to a large decrease in the electrical resistance. The resistivity due to electron-phonon scattering also decreases at low temperatures because the number of phonon modes that can take part in scattering is greatly diminished so that at these temperatures electron-impurity scattering and electron-dislocation scattering dominate the resistivity.

Matthiessen's rule states that if there are two physically differentiable scattering processes in a sample having electronic distribution relaxation times τ_1 and τ_2, and the processes do not interact, the total relaxation time τ is given by [178, 218]

$$\frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_2}. \tag{6.1.8}$$

If the relaxation times associated with each scattering process are taken to be $\vec{k}$ independent, then since the resistivity is proportional to $1/\tau$, the total resistivity ρ due to both scattering processes will be the sum of the resistivities [178, 218] due to each process,

$$\rho = \rho_1 + \rho_2 . \tag{6.1.9}$$

Matthiessen's rule breaks down if τ has a $\vec{k}$ dependence. If electron-impurity and electron-electron scattering dominate all scattering processes, Matthiessen's rule predicts that the resistivity will be of the form [178, 218 - 219]

$$\rho = AT + BT^2 , \tag{6.1.10}$$

where A and B are temperature independent coefficients. If ρ_1 is taken as the resistivity due to electron-phonon scattering and ρ_2 is taken as the resistivity due to electron-dislocation scattering, Matthiessen's rule expresses the empirical observation that ρ_2 is often temperature independent and ρ_1 is independent of the dislocation density when the dislocation density is small. The residual resistivity ρ_0 is the temperature independent part of the resistivity and its value is approached as $T \rightarrow 0$. ρ_0 is sample dependent for different samples of the same metal, but the quantity $\rho_L(T)$ defined as [219]

$$\rho_L(T) = \rho(T) - \rho_0 , \tag{6.1.11}$$

and referred to as the lattice resistivity, is sample independent for any particular metal. The residual resistivity ratio of a sample, which can be used as an estimate of its purity, is the ratio of its room temperature resistivity to its residual resistivity. Very pure material samples can have residual resistivity ratios as high as 10^6 [218].

6.1.1 Electron-Phonon Scattering

The electron-phonon contribution to the electrical resistivity is composed of two parts [219],

$$\rho_{e-p}(T) = \rho_N(T) + \rho_U(T). \tag{6.1.1.1}$$

The first part, $\rho_N(T)$, is due to normal scattering and the other part to Umklapp scattering. The Umklapp process combines the phonon scattering and Bragg reflection of an electron. For a normal electron-phonon scattering process [178, 218 – 219],

$$\vec{k}_2 - \vec{k}_1 = \vec{q}, \tag{6.1.1.2}$$

and for an Umklapp electron-phonon scattering process [178, 218 - 219],

$$\vec{k}_2 - \vec{k}_1 = \vec{q} + \vec{G}, \tag{6.1.1.3}$$

where $\vec{G}$ is a reciprocal lattice vector. These processes are depicted in Figure 6.2.

The temperature dependences of normal and Umklapp electron-phonon scattering processes are well known for ideal metals [219]. The alkali metals are the closest to ideal metals because their almost spherical Fermi surfaces do not make contact with the first

Brillouin zone. In these metals, there always exists a minimum value of the wavevector length $q_{\min}$, which is always exceeded by the length of a phonon wavevector $\vec{q}$.

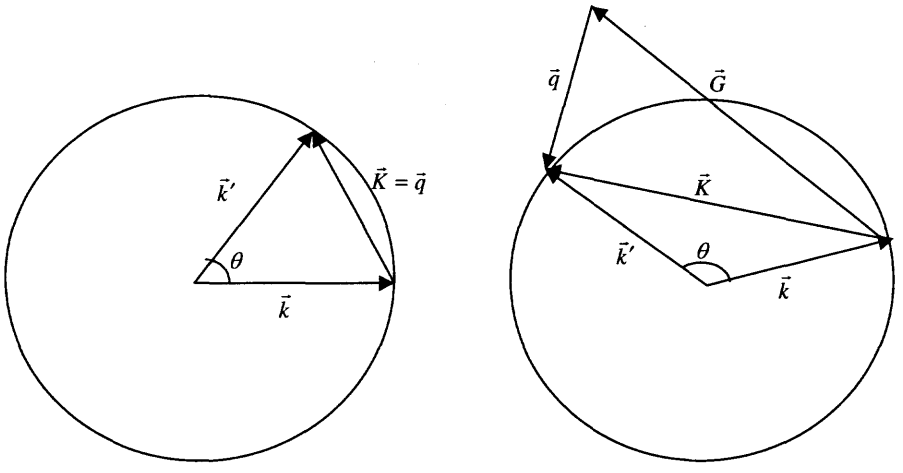


Figure 6.2 A normal (N) electron-phonon scattering process (left) and an Umklapp (U) electron-phonon scattering process (right) on a spherical Fermi surface.

At the lowest temperatures, phonons having this value of $\vec{q}$ are excluded from the Bose-Einstein distribution, and occur with a density $n_{q_{\min}}$ given by the approximate relation [219],

$$n_{q_{\min}} \approx \exp(-\hbar\omega/k_B T). \tag{6.1.1.4}$$

The contribution of the Umklapp processes to the resistivity can then be expected to be of the form [219]

$$\rho_U \propto \exp(-q_0 \Theta_D / K_D T), \tag{6.1.1.5}$$

where q_0 is the shortest distance between the Fermi surface and any of its images located in adjacent Brillouin zones. The following electron-phonon temperature dependencies are found for the alkali metals [219]

$$\rho_N(T) = K_N T^5, \tag{6.1.1.6}$$

and [219]

$$\rho_U(T) = K_U T^n \exp(-T_0/T), \tag{6.1.1.7}$$

where K_N and K_U are constants. Equation (6.1.1.6) is similar to the Bloch T^5 law. At low temperatures, $\rho_U(T)$ should become negligible, and $\rho_N(T)$ should dominate, leading to the Bloch T^5 law. However, phonon drag effects result in the phonon system drifting in the electric field direction and lead to the following expression for the electron-phonon contribution to the resistivity [219]

$$\rho_{e-p}(T) = \rho_N(T) + \rho_U(T) - \rho_{p-d}(T), \tag{6.1.1.8}$$

where the $-\rho_{p-d}(T)$ term is due to the phonon drag. As T approaches zero, $\rho_{e-p}(T)$ approaches $\gamma(T)\rho_U(T) \propto \exp(-T_0/T)$ [219], where $\gamma(T)$ is a weakly temperature dependent function that is always less than unity.

In the noble and polyvalent metals, the absence of a minimum phonon wavevector $\vec{q}_{\min}$ leads to a different functional dependence for $\rho_U(T)$. The Fermi surfaces belonging to non-alkali metals intersect the first Brillouin zone therefore there is no minimum

wavevector for the occurrence of Umklapp process, resulting in a non-exponential Bose-Einstein factor [219]. $\rho_U(T)$ is instead given by [219]

$$\rho_U(T) = K_U T^5, \tag{6.1.1.9}$$

which is of the same functional form as $\rho_N(T)$. Because the Fermi surfaces intersect the Brillouin zone boundaries, the phonon-drag effect is negligible for non-alkali metals and there is a preponderance of Umklapp processes at even the lowest temperatures. The temperature dependence of the electron-phonon contribution to the electrical resistivity for non-alkali metals is then given by [219]

$$\rho_{e-p}^{n-a}(T) = \rho_N(T) + \rho_U(T) = K_{e-p}^{n-a} T^5, \tag{6.1.1.10}$$

with K_{e-p}^{n-a} constant. The exact nature of this temperature dependence is not the same as the Bloch dependence, even though the general form is the same. The magnitude of K_{e-p}^{n-a} is much greater than that of the constant K_{Bloch} appearing in the Bloch resistivity because of the deviations of the electronic wavefunctions in non-alkali metals from free electron behavior and also because of the non-spherical Fermi surface [219]. Both expressions are proportional to T^5 as the temperature goes to zero, but the T^5 behavior of the Bloch resistivity goes up to $0.1\,\Theta_D$, while there are significant deviations from T^5 behavior for temperatures as low as $0.001\,\Theta_D$ in the case of $\rho_{e-p}^{n-a}(T)$.

6.1.2 Electron-Electron and Other Forms of Scattering

At very low temperatures, below about 2 - 3K, a contribution from electron-electron scattering to the total resistivity ρ_{tot} arises [219]. Writing the total resistivity as

$$\rho_{tot}(T) = \rho_0 + \rho_{e-p}(T) + \rho_{e-e}(T), \tag{6.1.2.1}$$

The electron-electron scattering term $\rho_{e-e}(T)$ exhibits a quadratic temperature dependence incorporating a constant c_{e-e} proportional to the number of electron-electron scattering processes and is given by [219]

$$\rho_{e-e}(T) = c_{e-e}T^2. \tag{6.1.2.2}$$

Only at these low temperatures can the term due to the electron-phonon interaction, $\rho_{e-p}(T)$, be prevented from overwhelming the $\rho_{e-e}(T)$ term to the extent that it can be measured. A typical value of the ratio $\rho_0/\rho_{e-e}(T)$ for a pure metal is about 10^4 [219], so that very high relative accuracy is needed in the measurement of the total resistivity $\rho_{tot}(T)$.

Adequate relative accuracy in the measurement of $\rho_{tot}(T)$ was not achieved until the mid-1970s, when experimental breakthroughs were achieved in the measurement of the low temperature electrical resistivity as represented by devices such as the SQUID, the SLUG and highly accurate equipment for the measurement of superconductivity [67, 219]. It was subsequently discovered, using these highly accurate instruments, which are capable of ppm relative accuracy measurements, that the values of c_{e-e} measured

experimentally for any particular metal were sample dependent. The measured magnitudes of ρ_0 and c_{e-e} have been found to increase with strain in the measured samples, and to decrease when the samples are annealed, then measured. Explanations for these observed phenomena have pointed out that the electronic relaxation time over the Fermi surface is not isotropic due to a significant density of anisotropic scattering centers in the measured samples. Anisotropic electronic relaxation times invariably lead to the enhancement of c_{e-e}. The contribution of electron-electron scattering to c_{e-e} has been proposed to be of the form [219]

$$P_{e-e}(\vec{k}_1,\vec{k}_2;\vec{k}_3,\vec{k}_4)\alpha_E\,, \tag{6.1.2.3}$$

where $P_{e-e}(\vec{k}_1,\vec{k}_2;\vec{k}_3,\vec{k}_4)$ is the transition probability for electrons in states $\vec{k}_1$ and $\vec{k}_2$ to scatter off each other into final states $\vec{k}_3$ and $\vec{k}_4$, and α_E is an effectiveness factor which measures the contribution of electron-electron scattering to the resistivity.

In the simple case of a free electron being scattered through an angle θ from a state $\vec{k}_1$ to a final state $\vec{k}_2$ by spherical impurities, α_E is equal to $1-\cos\theta$. A straightforward analysis [219] also shows that c_{e-e} is zero in this case because its two components, c_{e-e}^{U} and c_{e-e}^{N}, due to Umklapp and normal electron-electron scattering respectively, are both zero. Umklapp electron-electron scattering does not occur for free electrons, and c_{e-e}^{N} is zero because normal electron-electron scattering does not contribute to the electrical resistivity. c_{e-e}^{U} is not equal to zero when the distortions of the electronic

wave functions by the lattice potential are taken into account. c_{e-e}^{N} is also not equal to zero when electron-dislocation scattering occurs.

At very low temperatures, electron-impurity scattering and electron-dislocation scattering, with relaxation times of $\tau_{e-i}(\vec{k})$ and $\tau_{e-d}(\vec{k})$ respectively, dominate electron scattering processes [219]. The relaxation time associated with electron-dislocation scattering depends on $\vec{k}$. The total relaxation time τ influences the effectiveness factor α_E in the following manner [219],

$$\alpha_E \propto [\tau(\vec{k}_1)\vec{k}_1 + \tau(\vec{k}_2)\vec{k}_2 - \tau(\vec{k}_3)\vec{k}_3 - \tau(\vec{k}_4)\vec{k}_4]^2 . \tag{6.1.2.4}$$

From this expression, it is evident that τ depends on all electron scattering processes, not just electron-electron scattering. The residual resistivity ρ_0 can be written as [219]

$$\rho_0 = \rho_{imp} + \rho_{dis} . \tag{6.1.2.5}$$

Two limiting behaviors of the residual resistivity have been identified [219], based on whether scattering processes are dominated by isotropic scattering centers or by anisotropic scattering centers. Impurities are generally isotropic scattering centers with a constant scattering relaxation time. Normal electron-electron scattering with a $\vec{k}$ dependent relaxation time redistributes electrons into different Fermi surface states and changes the resistivity, leading to a non-zero value of c_{e-e}^{N}. The electron-electron scattering contribution depends on particular sample properties and can be represented as

$$c_{e-e}^{iso} = c_{e-e}^{U}, \tag{6.1.2.6}$$

in the isotropic limit and as

$$c_{e-e}^{aniso} = c_{e-e}^{U} + c_{e-e}^{N,aniso}, \tag{6.1.2.7}$$

in the anisotropic limit. For samples in which ρ_{imp} and ρ_{dis} are the same order of magnitude, the exact value of the electron-electron scattering contribution to the resistivity will depend on the isotropic to anisotropic scattering ratio [219], which can be determined from the value of ρ_{imp}/ρ_{dis} for the particular sample.

The phenomena consisting of the sample dependence of the measured resistivity at low temperatures for different samples are known as deviations from Matthiessen's rule (DMR) [219]. DMR effects occur because of non-isotropic scattering. Electrons in small areas of the Fermi surface, particularly the areas around the intersection points with the first Brillouin zone, are scattered much more strongly than electrons in other parts of the Fermi surface. The strong scattering regions of the Fermi surface result in very small relaxation times $\tau(\vec{k})$ for the electrons in this region, which as a consequence do not deviate significantly enough from their equilibrium distributions to contribute to the electrical current. The main contribution to the current is therefore from weakly scattered electrons and resistivity is minimal. When electron-impurity scattering is the dominant form of scattering, as seen in impure metals, an isotropic $\tau(\vec{k})$ and therefore a high resistivity will result since scattering from all electrons contribute to the resistivity and the electrons in the strongly scattering regions of the Fermi surface will make the

dominant contribution to the resistivity. For samples in which the residual resistivity is dominated by electron-impurity scattering, the degree of dominance of electron-impurity or electron-phonon scattering depends on the sample purity.

When there is a predominance of electron-impurity scattering processes, the relaxation time $\tau(\vec{k})$ approaches the electron-impurity scattering relaxation time $\tau_{e-i}(\vec{k})$. If electron-phonon scattering dominates, $\tau(\vec{k})$ approaches $\tau_{e-p}(\vec{k})$. $\tau(\vec{k})$ is therefore a function of the residual resistivity ρ_0 and deviates from Matthiessen's rule for large values of ρ_0. As ρ_0 increases, the electron-phonon scattering process is dominated by electron-impurity scattering, for which the relaxation time $\tau_{e-i}(\vec{k})$ is isotropic. In the limit of large ρ_0, the isotropic nature of $\tau_{e-i}(\vec{k})$ indicates that all electrons are involved in the scattering process. The resistivity due to electron-phonon scattering $\rho_{e-p}(\tau(\vec{k}))$ is then greater than $\rho_{e-p}(\tau_{e-p}(\vec{k}))$ for all values of $\tau(\vec{k})$ not equal to $\tau_{e-p}(\vec{k})$. In the large ρ_0 limit, when the initial state $\vec{k}_1$ and the final state $\vec{k}_2$ are both contained in the regions near the points of contact of the Fermi surface with the first Brillouin zone, the electron-phonon scattering probabilities greatly increase at low temperatures for Umklapp and normal scattering processes because of the enhancement of the effectiveness factor α. The dominant contribution to $\rho_{e-p}(\tau(\vec{k}))$ will then come from electrons in this region.

In the limit of vanishing ρ_0, electron-phonon scattering, which is an inelastic scattering process, determines the functional form of the relaxation time $\tau(\vec{k})$ because electron-impurity scattering is now negligible. In regions where the Fermi surface intersects the first Brillouin zone, $\tau_{e-p}(\vec{k})$ decreases and is negligible at the Brillouin

zone boundary. Contrary to what is observed in the limit of large ρ_0, scattering from electrons in states near these regions contribute very little to ρ_{e-p}.

These theories of the temperature and residual resistivity dependence of the experimentally measured electrical resistivity $\rho_{\exp}$ have been confirmed for Al, for which the sample dependence of $\rho_{\exp}$ is especially pronounced at low temperatures [219]. Highly contaminated samples of Al have been observed to show DMR behavior over a wide temperature range. ρ_{e-p} does not become constant at a value $\rho_{e-p}(\tau_{e-i})$ and thus independent of ρ_0 as the value of ρ_0 increases because there is coupling between the electron-phonon and electron-impurity scattering transition probabilities.

In cases when ρ_0 for a sample is mostly composed of electron-dislocation scattering, putting the sample under strain can increase its value. In such samples, $\rho_{\exp}$ decreases as ρ_0 decreases, exhibiting negative DMR. Electron-dislocation scattering has long been known [220] to contribute significantly to the sample dependence of the resistivity. The relaxation time associated with electron-dislocation scattering $\tau_{e-d}(\vec{k})$, is in magnitude between the relaxation times associated with electron-phonon scattering and electron-impurity scattering, $\tau_{e-p}(\vec{k})$ and $\tau_{e-i}(\vec{k})$ respectively. $\tau_{e-d}(\vec{k})$ is not reduced as much as $\tau_{e-p}(\vec{k})$ in the regions near the points of contact of the Fermi surface with the first Brillouin zone, because $\tau_{e-d}(\vec{k})$ does not have a scattering probability factor associated with it, and $\tau_{e-i}(\vec{k})$ is not reduced at all in these regions. The transition probability for $\tau_{e-d}(\vec{k})$ is therefore dominated by large angle scattering, while the

transition probability for $\tau_{e-i}(\vec{k})$ is dominated by small angle scattering. Negative DMR arises when a sample is strained, increasing ρ_0 and driving ρ_{e-p} towards the strained limit. ρ_{e-p} decreases when this strained limit has a lower value than the initial unstrained value. Figure 6.3 depicts positive and negative DMR behavior for impure and strained samples.

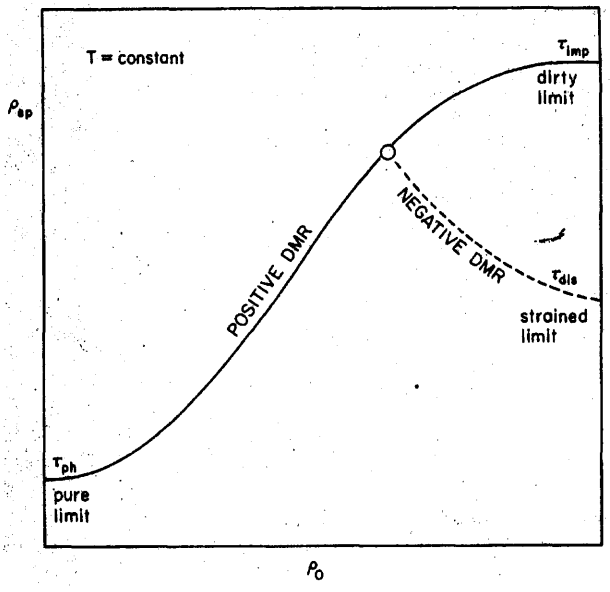


Figure 6.3 DMR graphs depicting positive and negative DMR for impure and strained samples respectively (taken from Reference [219]).

In ultra-pure samples, ρ_0 can be as low as 10^{-14}μΩcm [219], and the residual dislocation density becomes significant and influences the form of the relaxation time for electron scattering processes. These theories have explained many observed resistivity characteristics of the noble and some polyvalent metals, but more advanced theories are necessary to describe the resistivity behavior of the more complex transition metals.

6.2 Theories of the Electrical Resistivity

A simple description of the electrical resistivity is given by Bloch-Grüneisen [212, 219] theory, which predicts an expression of the form

$$\rho = \rho_0 + \rho(T)\,, \tag{6.2.1}$$

where the residual resistivity ρ_0 is linear in the defect density and the temperature dependent part $\rho(T)$ is linear in the density of phonons and therefore linear in temperature for temperatures greater than the Debye temperature, Θ_D. The first comprehensive theories to address the subject of metallic resistivity are due to Ziman [211, 212]. The Ziman theory adequately explained the Bloch-Grüneisen resistivity law as a special case that arises when the system of phonons is assumed to be in thermal equilibrium.

6.2.1 Ziman Theory of Resistivity

The central point of this theory concerns electron scattering processes that affect the electronic distribution and influence the canonical equation with scattering terms such as [212]

$$\begin{aligned} f_k^{scatt} &= \frac{1}{k_B T}\iint[\phi_{\vec{k}} + \phi_{\vec{q}} - \phi_{\vec{k}'}]P(\vec{k},\vec{q};\vec{k}')d\vec{k}'d\vec{q} \\ &+ \frac{1}{k_B T}\iint[\phi_{\vec{k}} - \phi_{\vec{q}} - \phi_{\vec{k}'}]P(\vec{k};\vec{q},\vec{k}')d\vec{k}'d\vec{q}\,, \end{aligned} \tag{6.2.1.1}$$

which describe the dependence of the electronic distribution $\phi_{\vec{k}}$ in state $\vec{k}$ upon the electronic distributions $\phi_{\vec{k}'}$ in other states and upon the phonon distribution function $\phi_{\vec{q}}$.

$P(\vec{k},\vec{q};\vec{k}')$ is the probability of scattering from state $(\vec{k},\vec{q})$ to state $\vec{k}'$. The canonical equation in this form cannot be solved exactly and various approximations must be employed to make it tractable. One such approximation [212], due to Bloch, is that the phonon system can be assumed to be in thermal equilibrium, that is

$$\phi_{\vec{q}} = 0. \tag{6.2.1.2}$$

With this approximation, a variational expression can be immediately obtained for the electrical resistivity [212]

$$\rho_L = \frac{1/k_B T \iiint [\phi_{\vec{k}} - \phi_{\vec{k}'}] P(\vec{k},\vec{q};\vec{k}') d\vec{k} d\vec{k}' d\vec{q}}{\left| \int e v_{\vec{k}} \phi_{\vec{k}} \partial f^0_{\vec{k}} / \partial \varepsilon_{\vec{k}} \right|^2}. \tag{6.2.1.3}$$

The solution of this variational expression using a trial function $\phi_{\vec{k}}$ defined as [212]

$$\phi_{\vec{k}} = \vec{k} \cdot \vec{u}, \tag{6.2.1.4}$$

yields [212]

$$\rho_{TOT} = \frac{9\pi\hbar}{2me^2 N k_B T} \frac{1}{R^2 A^2} \iint \frac{(\vec{K}\cdot\vec{u})^2 (\vec{K}\cdot\vec{e})^2 C^2(\vec{K})}{(1-e^{-\hbar\omega/k_B T})(e^{\hbar\omega/k_B T}-1)} \frac{dA}{v} \frac{dA'}{v'}, \tag{6.2.1.5}$$

where A is the Fermi surface area. The integration procedures resulting in this expression do not depend on the particular shape of the Fermi surface and whether it touches the Brillouin zone boundaries. They also are independent of the detailed form of the scattering matrix element, or whether the scattering processes are normal or Umklapp.

In the case of a simple monovalent metal having a spherical Fermi surface and undergoing normal scattering processes, $(\vec{K}\cdot\vec{u})$ and $(\vec{K}\cdot\vec{e})$ reduce to $K^2/3$ and K^2 respectively for a cubic lattice. With

$$K^2 = q^2 = 2R^2(1-\cos\theta);\ KdK = R^2\sin\theta d\theta\,, \tag{6.2.1.6}$$

the following expression [212]

$$\rho_N = \frac{3\pi\hbar}{4me^2Nk_BTR^4v_F^2}\int_0^{K_D}\frac{C^2(q)dq}{(1-e^{-\hbar\omega/k_BT})(e^{\hbar\omega/k_BT}-1)}, \tag{6.2.1.7}$$

is obtained for ρ_N, where K_D is the Debye radius, which is the largest allowed value of K and thereby q. Taking ϑ to be the angle between K and g, the magnitude of a reciprocal lattice vector, in an Umklapp process [212],

$$q^2 = K^2 + g^2 - 2gK\cos\vartheta, \tag{6.2.1.8}$$

and the contribution to the total resistivity from Umklapp processes is [212]

$$\rho_U = \frac{3\pi\hbar}{4me^2Nk_BTR^4v_F^2}\int_{g-2R}^{K_D}\frac{q}{(1-e^{-\hbar\omega/k_BT})(e^{\hbar\omega/k_BT}-1)}\left[\frac{N_z}{2g}\int_{g-q}^{2R}K^4C^2(K)dK\right]dq\,, \tag{6.2.1.9}$$

where N_z is the coordination number of the reciprocal lattice type. The total resistivity can then be written as [212]

$$\rho_{TOT} = \frac{3\pi\hbar}{4me^2 Nk_B TR^4 v_F^2} \int_0^{K_D} \frac{q^5[C^2(q) + C_U^2(q)]dq}{(1-e^{-\hbar\omega/k_B T})(e^{\hbar\omega/k_B T} - 1)}, \tag{6.2.1.10}$$

where

$$C_U^2(q) = \frac{N_z}{2gq^4} \int_{g-q}^{2R} K^4 C^2(K) dK. \tag{6.2.1.11}$$

Using the following functional forms [212] for $C(K)$,

$$C(K) = C(0) = \frac{n}{N(\varepsilon_F)}, \tag{6.2.1.12}$$

for normal scattering processes and

$$C(K) = 0, \tag{6.2.1.13}$$

for Umklapp scattering processes results in the following expression [212] for ρ_{TOT},

$$\rho_{TOT} = \frac{3\pi\hbar C^2(0)}{4me^2 Nk_B TR^4 v_F^2} \int_0^{K_D} \frac{q^5 dq}{(1-e^{-q\Theta_D/R_D T})(e^{q\Theta_D/R_D T} - 1)}, \tag{6.2.1.14}$$

or

$$\rho_{TOT} = \frac{3\pi\hbar}{4me^2 Nk_B TR^4 v_F^2} \frac{C^2(0) K_D^6 T^6}{\Theta_D} \int_0^{\Theta_D/T} \frac{z^5 dq}{(1-e^{-z})(e^z - 1)}, \tag{6.2.1.15}$$

which is equivalent to

$$\rho_{TOT} = \frac{3\pi\hbar C^2(0)K_D^6}{4me^2 N k_B T R^4 v_F^2}\left(\frac{T}{\Theta_D}\right)^5 I\left(\frac{\Theta_D}{T}\right), \tag{6.2.1.16}$$

where $I(\Theta_D/T)$ can be represented as $I(z)$, a Debye integral. The expression for ρ_{TOT} can finally be represented as

$$\rho_{TOT} = 4\left(\frac{T}{\Theta_D}\right)^5 I\left(\frac{\Theta_D}{T}\right)\rho_{\Theta_D}, \tag{6.2.1.17}$$

where ρ_{Θ_D} is a constant. This is the Bloch-Grüneisen equation for the resistivity.

At high temperatures, $I(z)$ takes the form $1/4(\Theta_D/T)^4$ [212] so that

$$\rho_{TOT} \approx \left(\frac{T}{\Theta_D}\right)\rho_{\Theta_D} \qquad T > \Theta_D. \tag{6.2.1.18}$$

The electrical resistance is proportional to the mean square thermal fluctuation of the lattice, which is proportional to the absolute temperature. At low temperatures, $I(z)$ tends towards a numerical value of 124.4 [212]. The resistivity is then given by

$$\rho_{TOT} \approx 497.6\left(\frac{T^5}{\Theta_D}\right)\rho_{\Theta_D}, \tag{6.2.1.19}$$

and is proportional to the fifth power of the temperature. According to the Debye law, the total density of phonons at low temperature is proportional to T^3, the number of phonons with wavevectors on the scattering surface is proportional to T^2, and the scattering matrix element is proportional to T. The effectiveness of scattering processes depends on

the square of the scattering angle, which is proportional to T. The combined effects of these factors cause ρ_{TOT} to be proportional to T^5. The Bloch-Grüneisen curve is well fitted by experimental data from a wide range of metals, as can be seen from Figure 6.4. The Bloch theory of resistivity is based on constant volume considerations and uses a very simple trial function to solve the Boltzmann equation.

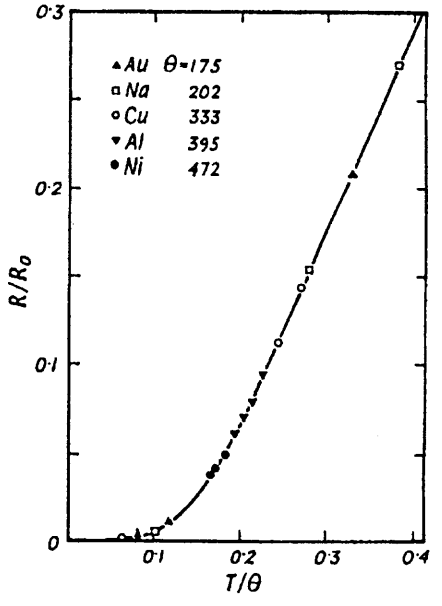


Figure 6.4 The Bloch Grüneisen formula fitted to metallic resistivity data (taken from Reference [212]).

Trial functions more accurate than Equation (6.2.1.4) result in reductions of approximately 10% in the calculated resistivity at a T/Θ_D value of 0.1, and negligible differences at extremely low and high temperatures [212]. Most of the discrepancies between the simple Bloch-Grüneisen model and experimental data stem from the manner in which Umklapp scattering processes are treated by the model. The form assumed for the electron-phonon scattering matrix element neglects Umklapp processes. In reality, the

matrix element should vary with $\vec{K}$. To treat more complicated metals in the high-temperature limit, where the Bose-Einstein distribution reduces to the classical distribution $(k_B T/\hbar\omega)^2$, an electron-phonon interaction of the form [212]

$$C(0) = \frac{n}{N(\varepsilon_F)} = \frac{4\pi^3 \hbar v_F n}{A}, \tag{6.2.1.20}$$

is used to produce the expression [212]

$$\rho_{TOT} = \frac{3\pi^3 \hbar^3 K_D T N_a}{4\sqrt{2} m e^2 k_B \Theta_D^2} \left(\frac{A^{free}}{A}\right)^2 \iint \left(\frac{\vec{K}\cdot\vec{u}}{R}\right)^2 \left(\frac{\vec{K}\cdot\vec{e}}{q}\right)^2 \left(\frac{C(K)}{C(0)}\right)^2 \frac{v_F^2}{vv'} \frac{dAdA'}{A^2}, \tag{6.2.1.21}$$

where N_a is the valency, and A^{free} is the free electron Fermi surface area. A parameter R, which depends mainly on the area of the Fermi surface and represents the reduced resistance, can be calculated from this formula [212]

$$R = \frac{4\sqrt{2} e^2 m k_B \Theta_D^2}{3\pi^3 \hbar^3 K_D T} \rho_{TOT}. \tag{6.2.1.22}$$

This parameter correlates very well with the position of the elements in the periodic table when plotted against atomic number, sinking to deep valleys at the positions corresponding to the monovalent metals, and rising to high plateaux at the positions corresponding to the transition metals, as shown in Figure 6.5.

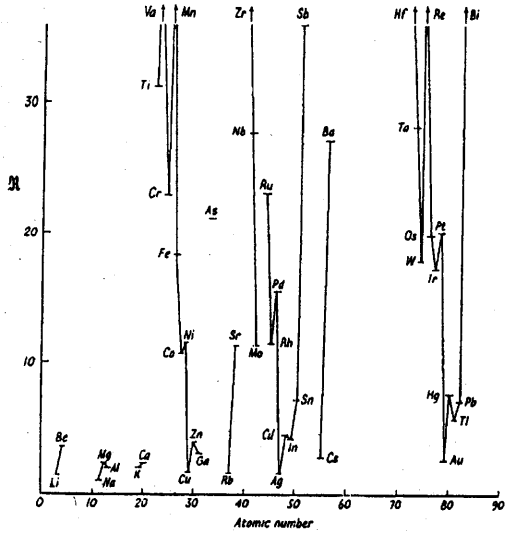


Figure 6.5 Reduced resistance as a function of atomic number (taken from Reference [212]).

R varies with the Fermi surface area according to [212]

$$R \approx N_a^{3/2}(A^{free}/A)^2 . \tag{6.2.1.23}$$

Low values of the electron-phonon scattering probability at large scattering angles cause R values for the transition metals to decrease relative to R values for the monovalent metals.

In transition metals, groups of electrons occupying different bands have very different velocities. Most of the current is carried by s electrons, there is considerable overlap of s and d bands, and there are many d electrons contained in narrow bands. The contribution of s-s scattering to the resistivity will be unaffected by the presence of d electrons, making its magnitude the same as in non-transition metals [212]. A new contribution to the resistivity, namely s-d scattering, must however be considered in the case of the transition metals. This contribution is given by [212]

$$\rho_{s-d} = \frac{9\pi\hbar}{2me^2 N k_B T v_s^2 A_s^2} \iint \frac{(v_s - v_d)^2 (\vec{K}\cdot\vec{e})^2 C_{s-d}^2(\vec{K})}{(1-e^{-\hbar\omega/k_B T})(e^{\hbar\omega/k_B T}-1)} \frac{dA_s}{v_s} \frac{dA_d}{v_d}, \tag{6.2.1.24}$$

where A_s is the Fermi surface area due to s electrons, A_d is the Fermi surface area due to d electrons, v_s and v_d are respectively the velocities of s and d electrons, with $v_s >> v_d$. The contribution to the electrical resistivity by s-d scattering of electrons by phonons is greater than the s-s scattering contribution by roughly the ratio of the density of states in the d bands to the density of states in the s bands.

Ziman [212] describes another theory of transition metal resistivity, which hypothesizes that d electrons are localized around transition metal atoms, making s-d scattering impossible. An exchange interaction can result with an energy that depends on the spin orientation of moving conduction electrons relative to the spin of the fixed moment of the other electrons. The exchange interaction affects the resistivity only when the d electrons are not all of the same spin. Raising the temperature leads to disordered d electron spin orientations, and thereby incoherent s electron scattering. The incoherent scattering would result in an extra resistivity, which rises sharply toward the Curie point, and then flattens off.

The Mott theory of transition metal resistivity states that the d band is a high-state density electron trap, into which s electrons can be scattered and lost. This theory has been confirmed somewhat with Pd-Ag alloys. Upon initial addition of Pd to Ag, a slight additional increase in the resistance is first observed, and after a certain weight percent of Pd is achieved, the conduction electrons of the Ag are no longer enough to quench the Pd d band holes, causing the resistance to rise rapidly. The resistivities of

transition and noble metal alloys as functions of composition are shown in Figure 6.6, where this effect can be seen.

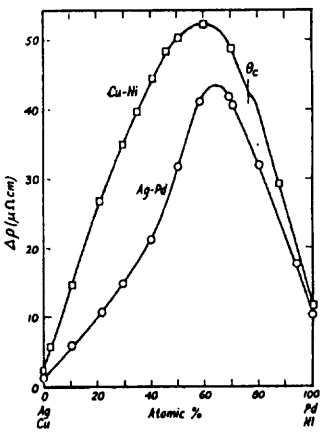


Figure 6.6 Resistivities of transition and noble metal alloys as functions of composition (taken from Reference [212]).

Ferromagnetic effects also affect the temperature dependence of the electrical resistivity of some transition metals. The ferromagnetic effects have been ascribed to the electron interactions, which undergo magnetic polarization when the number of spin-up electrons increases at the expense of spin-down electrons, causing the Fermi level of the spin-up electrons to increase while the Fermi level of the spin-down electrons decreases. An exchange interaction brings the two Fermi levels into coincidence, thereby stabilizing the system. If a current of spin-up s electrons is present and only the spin-up d states are occupied, then spin-up s-d scattering cannot occur, and the spin-polarized metal will be less resistive than the unpolarized metal by a magnitude [212]

$$\Delta\rho \approx -\rho_0/9(M/M_s)^2, \tag{6.2.1.25}$$

where M is the net magnetization and M_s is the saturation magnetization. The temperature dependent resistivity of ferromagnetic Ni is compared to that of non-magnetic Pd in Figure 6.7.

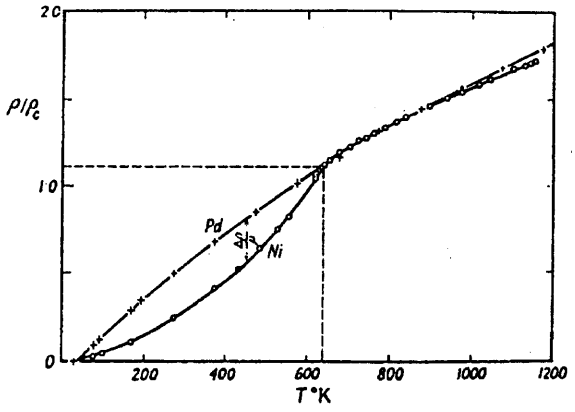


Figure 6.7 Resistivity of the ferromagnetic transition metal Ni compared with the non-magnetic metal Pd (taken from Reference [212]).

Normal electron-electron scattering does not contribute to the electrical resistivity, however Umklapp electron-electron scattering can contribute to the resistivity a term ρ_{e-e} given by [212]

$$\rho_{e-e} \approx \frac{\pi^3 z'}{32} \frac{e^2}{v_F \varepsilon_F} \gamma^2 \left(\frac{gR}{\xi}\right)\left(\frac{k_B T}{\varepsilon_F}\right)^2, \tag{6.2.1.26}$$

where ξ has the units of inverse length and γ is an interference factor. Electron-electron scattering decreases as T^2 while electron-phonon scattering decreases as T^5, therefore electron-electron scattering contributes significantly to the low temperature resistivity and even more so in the case of transition metals if the two band s-d scattering model is adopted. The collision of the light and fast moving s electrons, which carry the current,

with the relatively heavy d electrons affects the resistivity appreciably, the normal electron-electron scattering processes now contribute as well and assuming a spherical Fermi surface and a variational function of the sort [212]

$$\phi_k \approx (v_s - v_d)^2 (\vec{K} \cdot \vec{u}) / R^2 , \tag{6.2.1.27}$$

the electron-electron contribution to the resistivity is [212]

$$\rho_{e-e}^{sd} \approx \frac{\pi^4 z'}{16} \frac{e^2}{v_s \varepsilon_s} \left(\frac{R}{\xi} \right) \left(\frac{v_s - v_d}{v_d} \right)^2 \left(\frac{k_B T}{\varepsilon_s} \right)^2 , \tag{6.2.1.28}$$

where ε_s is the Fermi energy of the s electrons, assumed to lie in a parabolic energy band. The factor $[(v_s - v_d)/v_d]^2$, which is large compared to the interference factor γ occurring in the electron-electron contribution to the resistivity for simple metals, accounts for the higher resistive effect of the electron-electron collisions in transition metals. The low temperature resistivities of some transition metals such as Pt and Pd follow a T^2 dependence [221], confirming the importance of electron-electron scattering at lower temperatures.

Metals that exhibit large negative deviations from linearity at high temperatures include Nb [222 - 223], Pd, Pt, as well as the actinide metals uranium (U), neptunium (Np) and plutonium (Pu) [221]. The low temperature resistivities of some of these metals have been discovered [221] to obey a T^2 law, resulting in their being represented in terms of spin-fluctuations. Many metallic compounds exhibiting finite superconducting transition temperatures also display similar high-temperature resistivity behavior. The low temperature resistivities of these compounds have been discovered [221] to obey a

T^3 law, leading to descriptions in terms of s to d phonon scattering. Above the T^3 temperature regime, the temperature dependence of the contribution to ρ due to s to d phonon-induced scattering is insensitive to the phonon spectrum. For Pd, the resistivity due to s-d inter-band scattering accounts adequately for the temperature dependence of the resistivity at high temperatures [221]. Structure in the density of states at the Fermi level is also important in determining the high-temperature resistivity behavior.

The electron-phonon interaction also leads to a decrease in the density of states near the Fermi energy. The resistance then varies approximately as [212, 219]

$$\rho \propto K_{Bloch} T^5 \qquad T < \Theta_D . \tag{6.2.1.29}$$

The almost inevitable presence of impurities and imperfections in metallic samples results in deviations from T^5 behavior. The electrical resistivity due to impurities is temperature independent because their electron scattering cross-section is virtually constant. At low enough temperatures, when the phonon scattering becomes insignificant, this type of resistivity still remains and is therefore termed the residual resistivity. In some metals such as Au, the resistivity is observed to decrease to a small value and then increase by a few percent as the temperature is increased [212]. This phenomenon is illustrated by Figure 6.8, which shows the resistivity minimum in a Cu-Fe system.

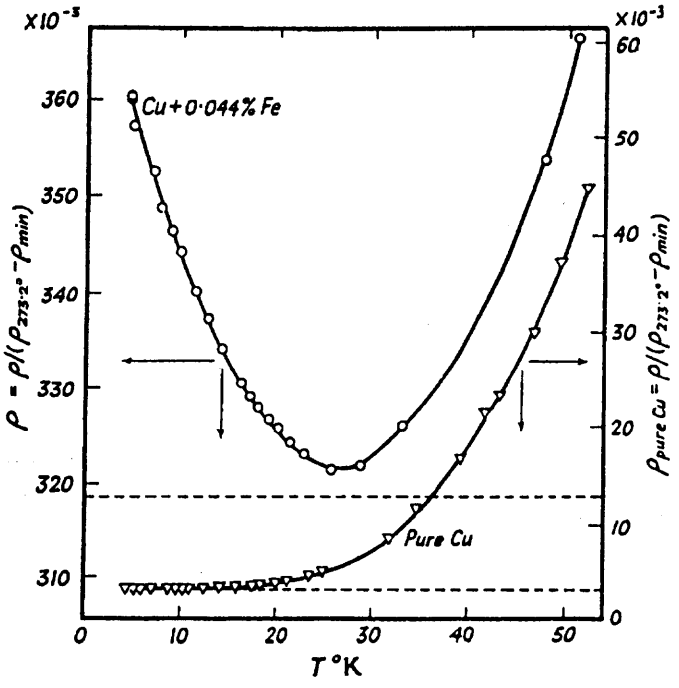


Figure 6.8 The resistivity minimum in a Cu-Fe system (taken from Reference [212]).

Small amounts of impurities in metals can cause this resistance minimum effect, but not all metals show a resistance minimum, whether in the pure or impure state. The temperature at which the resistance minimum occurs has been observed to be a function of the residual resistivity at the resistance minimum ρ_m which exhibits the following proportionality to the temperature at the resistance minimum T_m [212],

$$\rho_m \propto T_m^{\,5}. \tag{6.2.1.30}$$

In most metals, the electrical resistance increases when a magnetic field is applied, but in some metallic alloys with resistance minima, a negative magnetoresistance is observed. Crystalline imperfections such as dislocations, stacking faults and grain boundaries also contribute to the residual resistivity. Grain boundaries are not believed

[219] to contribute significantly to the resistivity, or at least to the extent that stacking faults and dislocations can. Approximately 10^{14} dislocations per cm^2 are required to cause as much residual resistance as 1% of a chemical impurity. Scattering from dislocations is an example of scattering from anisotropic scattering centers. The scattering probability from an array of dislocations is [219]

$$P_{dis}(\vec{k}_1,\vec{k}_2)=\frac{N_D}{2\pi\hbar^2 v_{\vec{k}}}\left[F(\vec{k}_2-\vec{k}_1)\right]^2\vec{k}_2 d\phi_2, \tag{6.2.1.31}$$

for scattering of an electron from state $\vec{k}_1$ into the azimuthal angle $d\phi_2$ of state $\vec{k}_2$, and N_D is the number of parallel dislocation lines crossing a unit area of a plane normal to the dislocation lines. The electronic mean free path, λ_{dis}, under such conditions is given by [212]

$$\frac{1}{\lambda_{dis}}=2N_D\left[u_x^2\int_0^{2\pi}\beta(\alpha)\cos^2\alpha d\alpha+u_y^2\int_0^{2\pi}\beta(\alpha)\sin^2\alpha d\alpha\right]\int_0^{2\pi}\frac{3\vec{K}}{8R^3}A(\vec{K})d\vec{K}, \tag{6.2.1.32}$$

where u_x and u_y are components of a unit vector in the direction of the applied electric field, $\vec{K}$ is equal to $\vec{k}_2-\vec{k}_1$, α is the angle between the electric field vector and the x-axis, and $A(\vec{K})$ and $\beta(\alpha)$ are model dependent functions. If the electric field vector were parallel to the z-axis, then λ_{dis} would be infinite and the resistivity due to electron-dislocation scattering would disappear. If the scattering cross section is constant, i.e.

independent of $\vec{K}$ and α, and if the field were confined to the plane, the mean free path λ_{dis} will reduce to [219]

$$\lambda_{dis} = \frac{1}{N_D \Gamma_D}, \tag{6.2.1.33}$$

where Γ_D is the scattering width of an edge dislocation.

6.2.2 Lowest Order Variational Approximation (LOVA) to the Resistivity

Tractable expressions for the electrical resistivity are derived in the framework of the Boltzmann equation, and the derivation in this section is taken from Reference [217]. For the Boltzmann equation to be valid, the overall mean free path $l = \langle v^2 \rangle^{1/2} \tau$, where $\langle v^2 \rangle$ is the mean square Fermi velocity, of an electron should be of sufficient magnitude that electrons can be assigned a propagation vector. Neglecting thermoelectric effects and assuming that the lattice phonons are in equilibrium, the Boltzmann equation is

$$\left(\left(\frac{\varepsilon_k}{T} \right) \frac{dT}{dx} + eE \right) v_x(k) \frac{\partial f}{\partial \varepsilon_k} = \sum_{k'} Q_{kk'} \phi_{k'}, \tag{6.2.2.1}$$

where the perturbing fields have been assumed to lie in the x direction, the subscript k represents the quantum numbers $\vec{k}$ and n, ε_k and $v_x(k)$ are respectively the energy and x component of the velocity for an electron in state k, e and E are respectively, the magnitudes of the electronic charge and the electric field, f is the Fermi distribution function, and $Q_{kk'}$ and $\phi_{k'}$ are respectively, the scattering operator and the electronic

distribution function. The Fermi function $f(\varepsilon_k)$ is a special case of the distribution function $F(\varepsilon_k)$ at equilibrium.

The electrical current $\vec{j}_e$ in terms of the deviation from equilibrium of the distribution function $F(\varepsilon_k) - f(\varepsilon_k)$ is

$$\vec{j}_e = -2e\sum_k v_k \phi_k \left(-\frac{\partial f}{\partial \varepsilon_k} \right) = \beta_{00}\vec{E} + \beta_{01}\vec{\nabla}T \,, \tag{6.2.2.2}$$

where the coefficients β_{00} and β_{01} are defined by

$$\beta_{00} = \sigma, -\beta_{00}^{-1}\beta_{01} = S \tag{6.2.2.3}$$

and σ and S are respectively, the electrical conductivity and the thermopower. Basis functions composed of polynomials in the energy and products of Fermi-surface harmonics are used to transform the Boltzmann equation into a matrix equation, which can be truncated to a 1×1 matrix to yield the usual variational solution for the electrical conductivity where ϕ_k is taken proportional to v_k.

The Fermi-surface harmonics are used to account for the changes in ϕ_k due to the Fermi-surface variation of $\vec{k}$. They are generalizations of the spherical harmonics and are orthonormal upon integration over the Fermi surface. The orthonormality conditions for the Fermi-surface harmonics are

$$\frac{\sum_k F_J(k)F_{J'}(k)\delta(\varepsilon-\varepsilon_k)}{\sum_k \delta(\varepsilon-\varepsilon_k)} = \delta_{JJ'}. \tag{6.2.2.4}$$

The $F_J(k)$ are element and material specific. They are periodic in space as well as continuous and differentiable at the Brillouin zone boundaries. For spherical surfaces the $F_J(k)$ become the spherical harmonics $Y_{l,m}(k)$. To simplify the orthonormality properties, the $F_J(k)$ are chosen to transform under the irreducible operations of the point group of the crystal for which they are being constructed. Rewriting the current as

$$\vec{j}_{ex} = -2e\int -(\partial f/\partial\varepsilon)\sum_k v_{kx}\phi_k\delta(\varepsilon-\varepsilon_k)d\varepsilon\,, \tag{6.2.2.5}$$

and if there is only one Fermi-surface sheet, one of the functions $F_J(k)$ can be chosen proportional to v_{kx} for convenience, resulting in the rest of the $F_J(k)$ being orthogonal to v_{kx}. If there are N Fermi-surface sheets involved in the problem, then any set of N linearly independent combinations $\sum_j^N c_j v_{kx}^j$ are possible, corresponding to N independent first order polynomials in v_{kx}.

The energy polynomials $\sigma_n(\varepsilon)$ describe the energy variation and are polynomials in ε/k_BT of order n which are orthonormal with a weighting function given by $-\partial f/\partial\varepsilon$. They satisfy the following orthonormality conditions,

$$\int \sigma_n(\varepsilon)\sigma_{n'}(\varepsilon)(-\partial f/\partial\varepsilon)d\varepsilon = \delta_{nn'}. \tag{6.2.2.6}$$

They are specified by choosing σ_0 to be equal to unity and then requiring the rest of the polynomials to be orthonormal to σ_0 as well as to all lower polynomials. The coefficient of the n-th power of ε is always required to be positive. This combined basis set of functions is therefore suitable for expanding any $\vec{k}$ dependent function. When expanded in the combined basis, Equations (6.2.2.2) and (6.2.2.3) are transformed into the following

$$-\left(k_B \frac{\pi}{\sqrt{3}} \frac{dT}{dx} \delta_{n1} + eE\delta_{n0}\right)\delta_{JX} = \sum_{J',n'} Q_{J,n;J',n'} \phi_{J',n'}, \tag{6.2.2.7}$$

and

$$\vec{j}_e = -2e\phi_{X,0}, \tag{6.2.2.8}$$

where the function $\phi_{X,0}$ represents the combination of the Fermi surface harmonic F_X, proportional to the x component of the velocity, and the energy polynomial σ_0, which is proportional to unity. This leads to the following expression for the conductivity

$$\sigma = 2e^2 (Q^{-1})_{X,0;X,0}. \tag{6.2.2.9}$$

The expansion in the combined basis set converges reasonably quickly even for the most complicated band structures. In resistivity calculations, the distribution function is approximated by rigidly translating the Fermi distribution function to obtain $F[\varepsilon(k+\delta k)]$, equivalent to assuming that $\phi_k = \phi_{X,0} F_X(k)$, meaning that only the $X0$ element of ϕ_{Jn} is significant. When the off diagonal elements of the scattering matrix

$Q_{X,0;J,n}$ are small enough, perturbation theory can be used to obtain an expression for the resistivity given by

$$\rho = 1/2e^2 (Q^{-1})_{X,0;X,0}\,, \tag{6.2.2.10}$$

or

$$\rho = \frac{1}{2e^2}\left(Q_{X,0;X,0} - \sum_{\alpha'} \frac{Q_{X,0;\alpha} Q_{\alpha;X,0}}{Q_{\alpha,\alpha}} + \sum_{\alpha',\beta'} \frac{Q_{X,0;\alpha} Q_{\alpha,\beta} Q_{\beta;X,0}}{Q_{\alpha,\alpha} Q_{\beta,\beta}} - \cdots \right), \tag{6.2.2.11}$$

where α and β represent the indices (J,n). Keeping only the first term in this expression leads to the lowest order variational solution (LOVA) to the Boltzmann equation. The electrical resistivity is bounded from above by the functional $\rho_{\max}[\psi(k)]$ of an arbitrary trial function $\psi(k)$

$$\rho_{\max}[\psi_k(k)] = \frac{\sum_{k,k'} \psi_k Q_{kk'} \psi_{k'}}{2e^2 \left(\sum_k v_{kx} \psi_k (-\partial f / \partial \varepsilon_k) \right)^2}. \tag{6.2.2.12}$$

The minimum value of this functional is equal to the actual resistivity. The functional is minimized by expanding $\psi(k)$ in the combined Fermi harmonic and energy polynomial basis functions, inserting the expanded $\psi(k)$ in the functional and then varying the coefficients to find the minimum value of the functional. The LOVA solution corresponds to using v_{kx} for $\psi(k)$ in the functional. The scattering operator is given by

$$Q_{kk'} = (1/k_B T)\left(\delta_{kk'}\sum_{k''} P_{kk''} - P_{kk'}\right), \tag{6.2.2.13}$$

where $P_{kk'}$ is the equilibrium transition probability which is equal to

$$P_{kk'}^{e-i} = (2\pi/\hbar)\left|H_{kk'}\right|^2 f_k(1-f_{k'})n_i\delta(\varepsilon_k - \varepsilon_{k'}), \tag{6.2.2.14}$$

for electron-impurity scattering and

$$\begin{aligned} P_{kk'}^{e-p} &= (2\pi/\hbar)\left|M_{kk'}\right|^2 f_k(1-f_{k'})N_i\delta(\varepsilon_k - \varepsilon_{k'}) \\ &\times[(N_K+1)\delta(\varepsilon_k - \varepsilon_{k'} - \Omega_K) + N_K\delta(\varepsilon_k - \varepsilon_{k'} - \Omega_K)], \end{aligned} \tag{6.2.2.15}$$

for electron-phonon scattering. Here $\left|H_{kk'}\right|$ and $\left|M_{kk'}\right|$ are respectively the electron-impurity and electron-phonon scattering matrix elements, and n_i and N_K are the numbers of impurities and phonons respectively, while Ω_K is the phonon energy. Representing the conductivity as

$$\sigma = 2N(\varepsilon_F)\left\langle v_x^2\right\rangle e^2\tau, \tag{6.2.2.16}$$

and the resistivity as

$$\rho^{(0)} = [(n/m)_{eff}\, e^2]^{-1}(1/\tau_{e-i}^{(0)} + 1/\tau_{e-p}^{(0)}), \tag{6.2.2.17}$$

the quantity $N(\varepsilon_F)\left\langle v_x^2\right\rangle$ is given by

$$N(\varepsilon_F)\langle v_x^2\rangle = \sum_k v_x^2(k)(-\partial f/\partial\varepsilon_k)$$
$$\cong \sum_k v_x^2(k)\delta(\varepsilon-\varepsilon_k). \qquad (6.2.2.18)$$

The scattering rates being

$$\frac{1}{\tau_{e-i}^{(0)}} = \frac{2\pi N_i N(\varepsilon_F)}{\hbar}\frac{\sum_{kk'}(v_{kx}-v_{k''x})^2\left|H_{kk'}\right|^2\delta(\varepsilon_k)\delta(\varepsilon_{k'})}{2\sum_{jj'}v_{jx}^2\delta(\varepsilon_j)\delta(\varepsilon_{j'})}, \qquad (6.2.2.19)$$

for electron-impurity scattering and

$$\frac{1}{\tau_{e-p}^{(0)}} = \frac{2\pi}{\hbar}\frac{N(\varepsilon_F)}{k_BT}\int d\varepsilon\int d\varepsilon'\int d\Omega\frac{\sum_{kk'}(v_{kx}-v_{k'x})^2\left|M_{kk'}\right|^2\delta(\varepsilon-\varepsilon_k)\delta(\varepsilon-\varepsilon_{k'})}{2\sum_{jj'}v_{jx}^2\delta(\varepsilon_j)\delta(\varepsilon_{j'})}$$

$$\times\delta(\Omega-\Omega_K)f(\varepsilon)[1-f(\varepsilon')]\{[N(\Omega)+1]\delta(\varepsilon-\varepsilon'-\Omega)+N(\Omega)\delta(\varepsilon-\varepsilon'-\Omega)\}, \qquad (6.2.2.20)$$

for electron-phonon scattering. Here $N(\Omega)$ is the phonon density.

An electron-phonon spectral function is defined in general as

$$\alpha_{tr}^2F(\Omega,\varepsilon,\varepsilon') = N(\varepsilon_F)\frac{\sum_{kk'}(v_{kx}-v_{k'x})^2\left|M_{kk'}\right|^2\delta(\varepsilon-\varepsilon_k)\delta(\Omega-\Omega_K)}{2\sum_{jj'}v_{jx}^2\delta(\varepsilon_j)\delta(\varepsilon_{j'})}, \qquad (6.2.2.21)$$

and specifically, the spectral function representing scattering-in and scattering-out processes is

$$\alpha^2_{out(in)}(\Omega)F(\Omega) = \frac{\left(V^2/(2\pi)^6\right)\int dS_k/\hbar v_k \int dS_{k'}/\hbar v_{k'} \sum_{kk'} \left|M^j_{kk'}\right|^2 v_x(k)v_x(k')\delta(\Omega-\Omega^j_K)}{\hbar N(\varepsilon_F)\left\langle v_x^2\right\rangle}, \tag{6.2.2.22}$$

where V is the unit cell volume. The transport spectral function $\alpha^2_{tr}(\Omega)F(\Omega)$ is the difference between the scattering-in and the scattering-out spectral functions

$$\alpha^2_{tr}(\Omega)F(\Omega) = \alpha^2_{out}(\Omega)F(\Omega) - \alpha^2_{in}(\Omega)F(\Omega). \tag{6.2.2.23}$$

The spectral function usually depends only weakly on ε and ε' compared to the variables k_BT and $\hbar\Omega_K$. The ε and ε' dependences can therefore be neglected, resulting in $\alpha^2_{tr}F(\Omega)$. Defining an electron-phonon coupling constant λ_{tr} by

$$\lambda_{tr} = 2\int_0^\infty \frac{\alpha^2_{tr}F(\Omega)}{\Omega}d\Omega = N(\varepsilon_F)\frac{\sum_{kk'}(v_{kx}-v_{k'x})^2\left(\left|M_{kk'}\right|^2/\hbar\Omega_K\right)\delta(\varepsilon_k)\delta(\varepsilon_{k'})}{2\sum_{jj'}v^2_{jx}\delta(\varepsilon_j)\delta(\varepsilon_{j'})}, \tag{6.2.2.24}$$

at high temperatures, the electron-phonon scattering rate can be written as

$$1/\tau^{(0)}_{e-p} = (2\pi/\hbar)k_BT\lambda_{tr}. \tag{6.2.2.25}$$

The coupling constant λ, which determines the superconducting transition temperature of superconductors, differs from λ_{tr} by velocity factors.

The superconducting function $\alpha^2F(\Omega)$ behaves as Ω^2 in the limit of small Ω, while the velocity factor $(v_{kx}-v_{k'x})^2$ modifies the behavior of $\alpha_{tr}^2F(\Omega)$ to a Ω^4 dependence, resulting in a T^5-like contribution to the resistivity. These equations represent the generalization of the Bloch-Grüneisen formula to metals such as the transition metals, which have highly non-spherical Fermi surfaces and for which phonon spectra more realistic than the simple Debye phonon spectrum are needed in order to obtain accurate calculations of the resistivity. The Bloch-Grüneisen equations for the resistivity are obtained by taking $\alpha_{tr}^2F(\Omega)$ to equal $2\lambda_{tr}(\Omega/\Omega_D)^4\Theta(\Omega_D-\Omega)$ in the Debye approximation.

The LOVA equations are most accurate for temperatures $T \geq \Theta_D$ and represent transition metal resistivities very well when accurate energy bands are used in the calculations, even with the approximation $\lambda_{tr} \approx \lambda$ [222]. LOVA neglects the energy dependence and anisotropy of the electronic distribution function, and therefore assumes that the Fermi surface displaces rigidly in response to an external field [224]. At low temperatures, significant corrections due to $\vec{k}$ dependence and ε dependence of ϕ_k are usually necessary. The phonon density of states curves $F(\Omega)$ for many superconducting metals have been found to be similar in shape to $\alpha^2F(\Omega)$, and therefore by extension, to $\alpha_{tr}^2F(\Omega)$ except in the low Ω region where there is a change in the power law behavior from Ω^2 to Ω^4. The LOVA expression for the electrical resistivity, which is an upper bound to the true resistivity, is given by [222]

$$\rho(T) = \frac{6\pi V k_B T}{e^2 \hbar N(\varepsilon_F) < v^2 >} \int_0^\infty \frac{\alpha^2{}_{tr} F(\Omega) x^2}{\sinh^2 x} \frac{d\Omega}{\Omega}, \tag{6.2.2.26}$$

where $x = \hbar\Omega / 2k_B T$.

In the high-temperature limit, the expression for the resistivity represented by Equation (6.2.2.27) becomes

$$\rho = \frac{6\pi V k_B T}{e^2 \hbar N(\varepsilon_F) < v^2 >} \lambda_{tr} \left(1 - \frac{\hbar^2 \langle \Omega^2 \rangle_{tr}}{12 k_B{}^2 T^2} + \cdots \right). \tag{6.2.2.27}$$

The resistivity can then be rewritten as

$$\rho = (4\pi / \omega_p{}^2)(1/\tau), \tag{6.2.2.28}$$

where ω_p is the Drude plasma frequency and $1/\tau$ is the electron-phonon scattering rate given by [222]

$$1/\tau = 2\pi k_B T \lambda_{tr} \left(1 - \frac{\hbar^2 \langle \Omega^2 \rangle_{tr}}{12 k_B{}^2 T^2} + \cdots \right), \tag{6.2.2.29}$$

with the electron-phonon interaction parameter λ_{tr} given by [224]

$$\lambda_{tr} = 2 \int_0^\infty \alpha^2{}_{tr}(\Omega) F(\Omega) \frac{d\Omega}{\Omega}. \tag{6.2.2.30}$$

The mass enhancement factor λ determines the superconducting transition temperature T_c according to McMillan's formula [225],

$$T_c = \frac{\Theta_D}{1.45}\exp\left(-\frac{1.04(1+\lambda)}{\lambda-\mu^*(1+0.62\lambda)}\right). \tag{6.2.2.31}$$

λ_{tr} is closely related to the mass enhancement factor λ by [226]

$$\lambda_w = N(\varepsilon_F)\frac{\sum_{k,k'} w^{k,k'}\left|M^{k,k'}\right|^2(\hbar\omega_{k-k'})^{-1}\delta(\varepsilon-\varepsilon_k)\delta(\varepsilon-\varepsilon_{k'})}{\sum_{k,k'} w^{k,k'}\delta(\varepsilon-\varepsilon_k)\delta(\varepsilon-\varepsilon_{k'})}, \tag{6.2.2.32}$$

where $\left|M^{k,k'}\right|$ is the electron-phonon matrix element of scattering from state k of energy ε_k and group velocity v_k to a state k' by emission of a phonon of energy $\hbar\omega_{k-k'}$.

The weighting factor $w^{k,k'}$ is 1 for λ and $(v_{kx}-v_{k'x})^2$ for λ_{tr}. In an isotropic system, λ_{tr} and λ differ by the factor $\langle 1-\cos\theta\rangle$ which is of the order of unity, θ being the transport process scattering angle. λ_{tr} and λ usually agree to within 10% for d band metals [226 - 227], so in this work, λ_{tr} is approximated by λ, which is obtained from electronic tunneling data and McMillan's formula for T_c [227].

6.3 Plasma Frequency

Charge density waves in a metal are quantized in units of plasmons and oscillate with a frequency known as the plasma frequency, which is given by [228 – 229]

$$\omega_p = 4\pi n_c e^2 / m_c , \tag{6.3.1}$$

in the Drude limit ($\omega \to 0$) of free-electron behavior. ω_p is the Drude plasma frequency, n_c is the density of conduction electrons and m_c is the effective mass of the conduction electrons. When the frequency ω is sufficiently lower then the minimum frequency needed for interband absorption to occur, the following relation holds between the real part $\bar{\varepsilon}_1(\omega)$ of the dielectric constant $\bar{\varepsilon}(\omega)$ of a cubic metal and the Drude plasma frequency [228 – 229],

$$\bar{\varepsilon}_1(\omega) = \bar{\varepsilon}_1^0 - \omega_p^2 \tau^2 / (1 + \omega^2 \tau^2) , \tag{6.3.2}$$

where ω is the angular frequency of the incident light and $\bar{\varepsilon}_1^0$ is the static ion-core contribution. Interband transitions strongly affect the plasma frequency of metals, making the concept of an effective electron mass to describe it ineffective [228]. The real part of the dielectric constant is also given by [229]

$$\bar{\varepsilon}_1(\omega) = \eta^2 - \kappa^2 , \tag{6.3.3}$$

where η is the index of refraction and κ is the extinction coefficient. The Drude plasma frequency ω_p of a metal can therefore be extracted by measuring its optical spectrum. ω_p can also be derived from calculated band structure properties and therefore provides a useful measure of the band structure near the Fermi energy level. It is useful for studying transport properties such as the resistivity, as can be seen from Equation (6.2.2.28). It is

directly proportional to zero temperature metallic electrical conductivity and can be represented as [230]

$$\hbar^2\omega_p{}^2 = \frac{8\pi e^2}{(3V)4\pi\varepsilon_0} N(\varepsilon_F) < v^2 > . \tag{6.3.4}$$

The quantity $N(\varepsilon_F) < v^2 >$ can be represented as [230]

$$N(\varepsilon_F) < v^2 > = \sum_{nk} \left|\nabla_k \varepsilon_{nk}\right|^2 \delta(\varepsilon_{nk} - \varepsilon_F), \tag{6.3.5}$$

where the sums on the indices n and k are over occupied states. It is illustrative to transform the sum in Equation (6.3.5) into an integral over the Fermi surface [178, 230]

$$\sum_{nk} \left|\nabla_k \varepsilon_{nk}\right|^2 \delta(\varepsilon_{nk} - \varepsilon_F) = \frac{V}{8\pi^3} \int_S \nabla_k \varepsilon_{nk} dS, \tag{6.3.6}$$

showing that the plasma frequency is directly proportional to the Fermi surface average of the velocity [230]

$$\hbar^2\omega_p{}^2 = \frac{e^2}{12\varepsilon_0\pi^2} \int_S \nabla_k \varepsilon_{nk} dS . \tag{6.3.7}$$

The quantity $N(\varepsilon_F) < v^2 >$ has been calculated according to Equation (6.3.5) for all the bcc metals in this study and their corresponding bct phases using the FP-LAPW method. $N(\varepsilon) < v^2 >$ is shown plotted as a function of energy for the bcc and bct phases of all the studied metals in Figures 6.9(a) – 6.9(b).

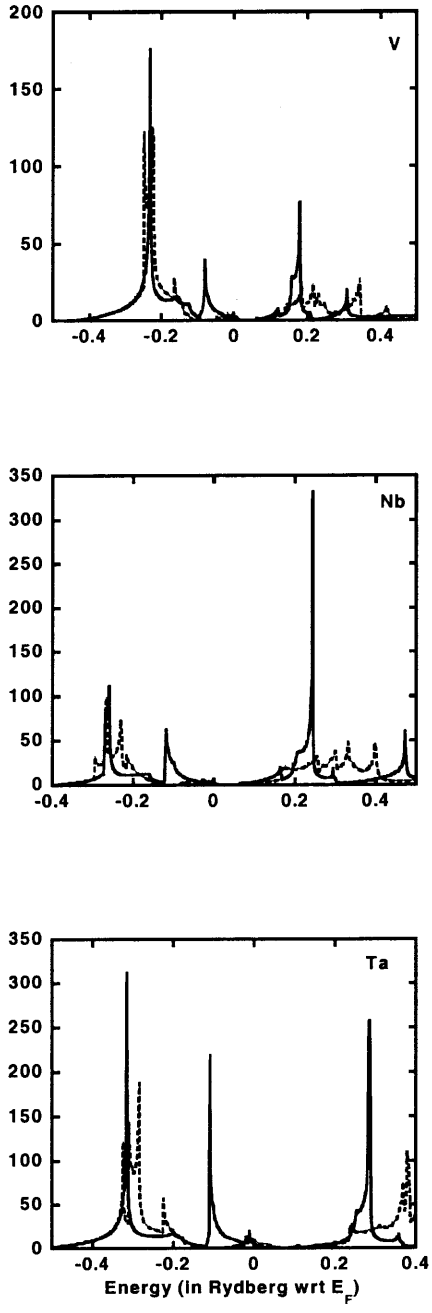


Figure 6.9(a) $N(\varepsilon) < v^2 >$ as a function of energy is shown for both the bcc (solid lines) and bct (dashed lines) phases of the studied group 5B metals V, Nb and Ta. In each panel of Figure 6.9(a), the Fermi energy is at 0 Rydberg.

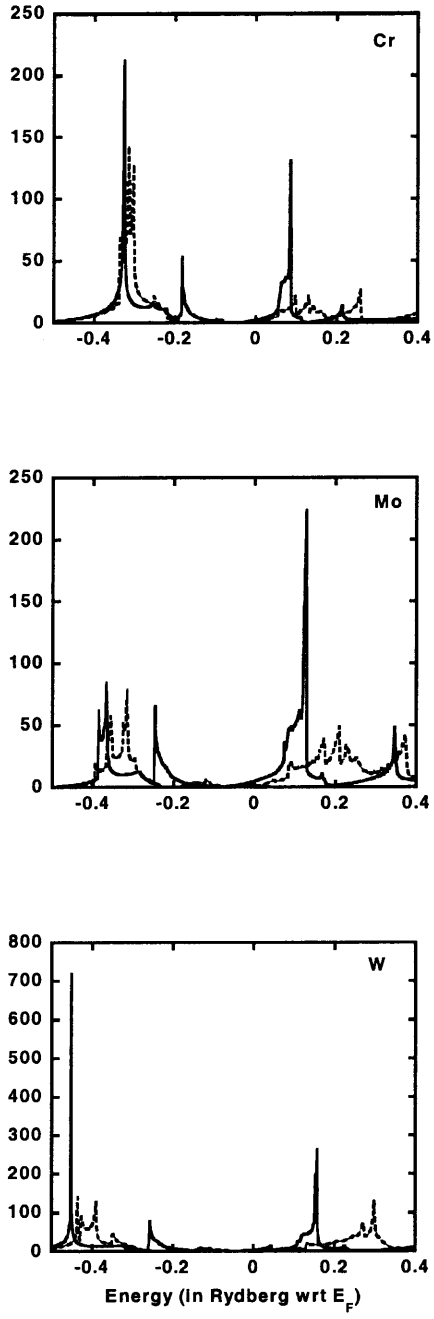


Figure 6.9(b) $N(\varepsilon) < v^2 >$ as a function of energy is shown for both the bcc (solid lines) and bct (dashed lines) phases of the studied group 6B metals Cr, Mo and W. In each panel of Figure 6.9(b), the Fermi energy is at 0 Rydberg.

Four peaks are noticeable from the figures for all the studied elements, although the peak at highest energy is barely discernible for the period 6 elements Ta and W. The

$N(\varepsilon) < v^2 >$ vs. ε figures for the period 4 bcc elements V and Cr look remarkably similar to each other, as is the case for the period 5 elements Nb and Mo. The highest peaks in the figures occur below ε_F for the period 4 and period 6 elements, while they occur above ε_F for the period 5 elements. These peaks occur at 0.24 and 0.34eV below ε_F for V and Cr respectively, while they occur at 0.24 and 0.12eV above ε_F for Nb and Mo respectively. The period 6 elements Ta and W both exhibit their most prominent peaks in the $N(\varepsilon) < v^2 >$ figures at 0.32 and 0.46eV respectively below ε_F, farther below what is seen in the figures for the period 4 elements. The $N(\varepsilon) < v^2 >$ vs. ε figures for the bct phases of all the elements show more structure than the corresponding bcc figures and exhibit peaks of generally lower intensity, although prominent peaks below ε_F at the same positions seen in the bcc $N(\varepsilon) < v^2 >$ figures still exist in the bct V, Cr and Ta figures.

$N(\varepsilon_F) < v^2 >$ is obtained from the value of $N(\varepsilon) < v^2 >$ at ε_F, which has been adjusted to 0 Rydberg in Figures 6.9(a) and 6.9(b). The quantity $N(\varepsilon_F) < v^2 >$ began to converge within 1% only when over 33,000 k-points in the irreducible wedge of the Brillouin zone were used in evaluating Equation (6.3.5). The calculated values of $N(\varepsilon_F) < v^2 >$ are shown in Table 6.1.

Table 6.1 Calculated Bcc and Bct Plasma Frequency Values $\hbar\omega_p$ for Studied Group 5 and Group 6 Metals

Metal	Bcc $N(E_F)<v^2>$ (Ry.bohr^2)	Bcc $\hbar\omega_p$ (eV)	Bcc $<v^2>^{1/2}$ (10^7 cm/s)	Bct $N(E_F)<v^2>$ (Ry.bohr^2)	Bct $\hbar\omega_p$ (eV)	Bct $<v^2>^{1/2}$ (10^7 cm/s)
V	2.370	6.36	3.36(3.73)[a,t]	2.772	6.83	4.07
Nb	1.692	4.63	3.09(5.10)[b,e]	0.714	2.98	8.24
Ta	5.565	8.34	6.04(6.70)[c,t]	9.434	10.78	7.90
Cr	2.490	7.04	5.76	1.592	5.57	3.62
Mo	4.974	8.50	8.60(8.61)[c,t]	1.994	5.33	4.34
W	4.063	7.60	9.34(9.60)[c,t]	0.212	1.72	1.59

[a]Soukoulis *et al*, 1982, [b]Mattheiss, 1970, [c]Chakraborty *et al*, 1976, e – experimental, t – theoretical.

The highest value of $N(\varepsilon_F)<v^2>$ among the bcc elements is obtained for Ta (5.565Ry.bohr^2), and the lowest value for Nb (1.692Ry.bohr^2). Among the group 5B elements, $N(\varepsilon_F)<v^2>$ values for the bct phases of V and Ta increased by 17% and 70% respectively over the bcc values, while the $N(\varepsilon_F)<v^2>$ value for bct Nb decreased by 58% from the bcc value. Among the group 6B elements, $N(\varepsilon_F)<v^2>$ values for the bct phases of W, Mo and Cr decreased by 95%, 60% and 36% respectively from the bcc values. The square root of the mean squared Fermi velocity, $<v^2>^{1/2}$, for all the elements in their bcc and bct phases has also been calculated from Equation (6.3.5) and Equation (6.3.7), assuming a spherical Fermi surface. The highest value of $<v^2>^{1/2}$ (Table 6.1) among the bcc elements is obtained for W (9.34 x 10^7 cm/s), and the lowest value for Nb (3.09 x 10^7 cm/s). The calculated values of $<v^2>^{1/2}$ are compared in Table 6.1 to other values of $<v^2>^{1/2}$ obtained by the electron-lifetime model and the APW

band structure method [231], Slater-Koster interpolation of APW data [232], and experimental measurement [233]. These values of $< v^2 >^{1/2}$ obtained by other authors are enclosed in parentheses and are listed in Table 6.1 next to the bcc $< v^2 >^{1/2}$ values calculated in this work. The $< v^2 >^{1/2}$ calculated for the group 5B elements V and Ta and the group 6B elements Mo and W in this work differ by an average of 10% and 2% respectively from these previously reported values. However, the $< v^2 >^{1/2}$ value calculated for Nb differed by 39% from the literature reported [233] value.

In each studied period, the group 6B bcc metal always had a higher value of $< v^2 >^{1/2}$ than the group 5B bcc metal in the same period. Bct values of $< v^2 >^{1/2}$ increased relative to the corresponding bcc phase values for the group 5B elements, while the bct values of $< v^2 >^{1/2}$ decreased relative to the corresponding bcc phase values for the group 6B elements.

The plasma frequencies calculated for all the studied metals according to Equation (6.3.4) and Equation (6.3.5) are also shown in Table 6.1. The highest magnitude of ω_p is obtained for cubic Mo (8.50 eV), while the lowest is obtained for cubic Nb (4.63 eV). Plasma frequencies increase in Group 5B for the bcc to bct phase transition except for Nb, while they decrease for the bcc to bct phase transition in Group 6B. The values of ω_p calculated in this work are compared in Figure 6.10 to values obtained both experimentally, using measurement of optical spectra in the infra-red region, and theoretically by other authors and are mostly well within the ranges reported by these authors.

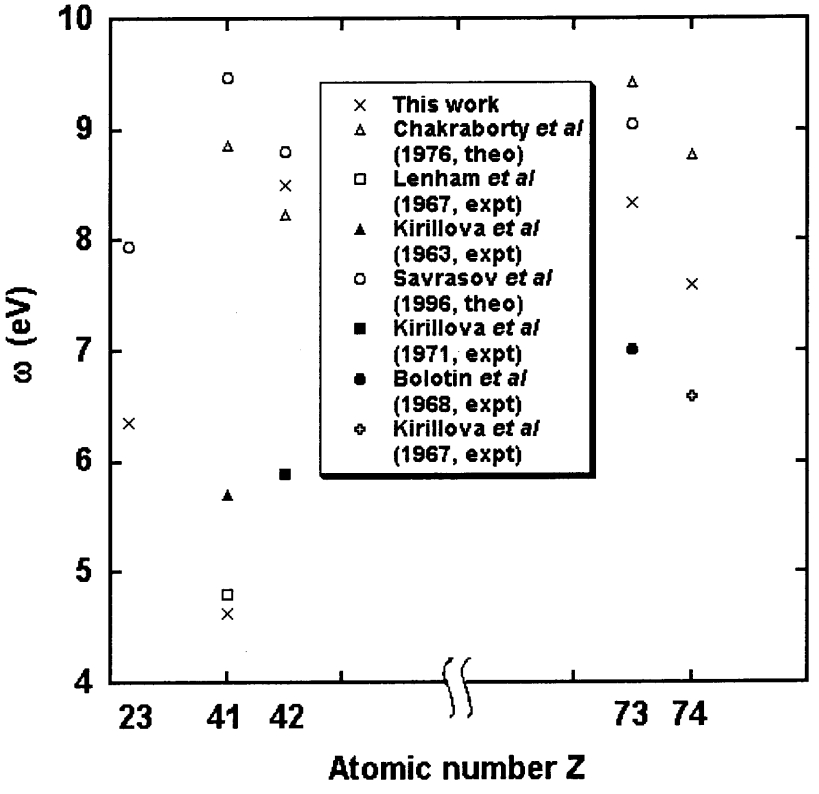


Figure 6.10 Calculated and experimental values of the Drude plasma frequency for V, Nb, Mo, Ta and W.

6.4 Pressure Dependence of Transport Properties

In order to predict the structure dependence of transport properties from band-structure calculations, it would be useful to test the accuracy of the computational method employed in this work by using it to try to reproduce results from work by Neve *et al* [230] in which the pressure dependence of transport properties such as the density of electron states at the Fermi level, $N(\varepsilon_F)$, and the plasma frequency ω_p for the transition metal Nb were calculated and experimentally measured. In this work the pressure dependence of these transport properties are calculated for Nb and Ta. The experimental method used by Neve *et al* to determine the pressure dependence of the plasma frequency

for Nb involved making use of the measured pressure dependence of the temperature gradient of the electrical resistivity $\partial\rho/\partial T$, published results for the superconducting transition temperature T_c, and the compressibility. The theoretical method consisted of a self-consistent Linear Muffin Tin Orbital (LMTO) band-structure calculation. The plasma frequency ω_p, which is essentially the product of $N(\varepsilon_F)$ and $<v^2>$, was found to be somewhat insensitive to small changes in the Fermi energy level. The density of states at the Fermi energy, $N(\varepsilon_F)$, was determined to decrease with increasing pressure, satisfying the relation

$$\frac{\partial \ln N(\varepsilon_F)}{d \ln V} = 1.2, \tag{6.4.1}$$

while the plasma frequency ω_p and $<v^2>^{1/2}$ were determined to increase with increasing pressure. The calculated energy bands were observed to broaden with increased pressure, resulting in reduced electron-electron interaction as measured by the Coulomb pseudopotential. An $N(\varepsilon_F)$ value of 10.68 states/Ry.atom.spin was obtained for Nb from our calculations, which is well within the range obtained by Neve *et al.* The calculated pressure dependence of transport properties such as $N(\varepsilon_F)<v^2>$, ω_p, $<v^2>^{1/2}$ and band properties such as ε_F and $N(\varepsilon_F)$ are shown for Nb and Ta in Figures 6.11(a) – 6.11(e).

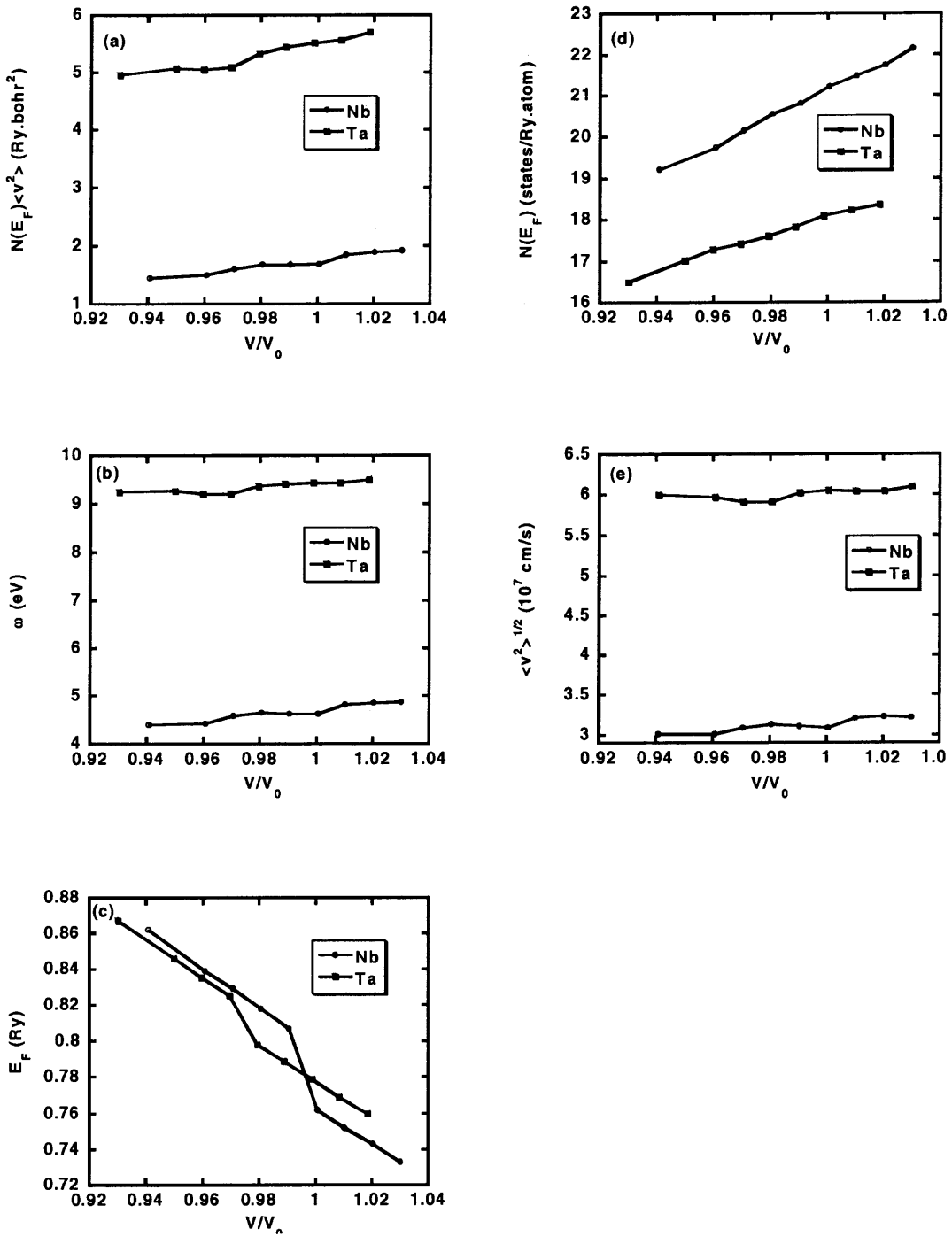


Figure 6.11 The pressure dependence of the quantities $N(\varepsilon_F) < v^2 >$, ω_p, ε_F, $N(\varepsilon_F)$ and $< v^2 >^{1/2}$ are shown for Nb and Ta. V_0 is the calculated bcc equilibrium volume.

$N(\varepsilon_F)<v^2>$, ω_p and $<v^2>^{1/2}$ show a general slight decrease with pressure while ε_F increases and $N(\varepsilon_F)$ decreases with pressure for both metals. This pressure dependence behavior of ε_F and $N(\varepsilon_F)$ is the same as previously obtained [230] for Nb, while the pressure dependence behavior of $N(\varepsilon_F)<v^2>$ and ω_p in this paper is opposite to what was observed in that work. $<v^2>^{1/2}$ showed a slight increase with pressure increase in Reference [230] which is contrary to what was observed in this study. The results obtained in this work for the pressure dependence of the plasma frequency, similar to those of Reference [230], show a slight dependence of the plasma frequency on the pressure, but the monotonic increase of the plasma frequency with pressure seen in that reference is not observed, however. The fact that the plasma frequencies calculated in this work do not increase monotonically with pressure as seen in Reference [230] might be due to the evaluation of Equation (6.3.5) without taking into account spin-polarization effects. By not including spin-polarization an average over the spin-up and spin-down states has been performed. On the other hand, the results from this work might be more accurate than those of Reference [230], where only 1053 irreducible k-points in the Brillouin zone were used in the LMTO-ASA scheme, whereas over 33,000 irreducible k-points have been used to determine the transport properties in this paper. The large discrepancy between the Nb $<v^2>^{1/2}$ value calculated in this work and the experimental Nb $<v^2>^{1/2}$ value reported in the literature [233] is also attributed to the manner in which Equation (6.3.5) was evaluated.

The observed decrease in $N(\varepsilon_F)$ with pressure increase mirrors the trend in $N(\varepsilon_F)$ down and across the studied block of the periodic table, where $N(\varepsilon_F)$ is seen to

decrease down the block and across it due mostly to respectively, an increase in Z and a slight contraction in the lattice parameter, both of which result in an increase in the electron concentration per unit cell. $N(\varepsilon_F) < v^2 >$ is observed to decrease slightly for bcc Nb and Ta with pressure increase or equivalently, with a decrease in unit cell volume. This behavior of $N(\varepsilon_F) < v^2 >$ is consistent with the previously noted increase in its bct phase values over its bcc phase values for the group 5B metals, since the transition from bcc to bct leads to a volume increase for all the studied metals. Since the bct values of $N(\varepsilon_F) < v^2 >$ decrease relative to the bcc values for the group 6B metals, it is predicted that $N(\varepsilon_F) < v^2 >$ and $< v^2 >^{1/2}$ will increase with an increase in pressure or reduction in unit cell volume for the group 6B bcc metals.

6.5 Calculated Resistivities and Implications for β-Ta Formation

The method used to calculate the resistivities is summarized in Table 6.2.

Table 6.2 Summary of Resistivity Calculation Method

Quantity	Calculated With	Comments
$\rho(T)$	$\rho(T) = \frac{6\pi V k_B T}{e^2 \hbar N(\varepsilon_F) < v^2 >} \times \int_0^\infty \frac{\alpha^2{}_{tr} F(\Omega) x^2}{\sinh^2 x} \frac{d\Omega}{\Omega}$	$\alpha^2{}_{tr} F(\Omega)$ is transport spectral function [216]
$N(\varepsilon_F) < v^2 >$	$\sum_{nk} \lvert \nabla_k \varepsilon_{nk} \rvert^2 \delta(\varepsilon_{nk} - \varepsilon_F)$	ε_{nk} is band structure
$\lambda_{tr} = \int_0^\infty \frac{\alpha^2{}_{tr} F(\Omega) x^2}{\sinh^2 x} \frac{d\Omega}{\Omega}$	$\lambda_{tr} \approx \lambda$	λ_{tr} is electron-phonon coupling constant, λ is mass enhancement factor from superconductivity theory [219]

It is attempted to use the calculations in this study to reproduce the experimental resistivities of all the studied bcc metals (except Cr) over a defined temperature range. It is also attempted to reproduce the trend in the magnitudes of the resistivities at room temperature observed from experimental data in order to have some degree of confidence in the ability of the computational method employed in this work to differentiate between possible structures of β-Ta based on the comparison of calculated resistivity values to experimental data. The method being used to calculate the resistivities does not permit calculations for Cr, because bcc Cr is not a superconductor. Accurate resistivity data for the rest of the metals [234], from which the residual resistivities have been subtracted, have been used for the comparisons to the calculations in this work. The room temperature resistivities of β-Ta thin films with thicknesses ranging from 100 – 20,000 Å have been measured by the four-point probe method [81, 91, 93, 95, 107 – 108, 111, 152] and the values obtained fall in the range 170 – 220 μΩ cm. The value of the bcc Ta room temperature resistivity is 13.6 μΩ cm [93, 115], 1200 – 1600 % less than the value for β-Ta. It is therefore crucial that the errors associated with the calculated resistivities remain well below these boundaries.

The experimental resistivity data are compared to the calculated resistivities over the temperature range 20 – 700 K in Figure 6.12 for Ta and in Figure 6.13 for the rest of the studied metals, except Cr. At room temperature, the calculated resistivities differ from the experimental data by less than 15% for all the studied bcc metals except for Nb, for which an anomalously large value of the resistivity (350% greater than experiment) is obtained from the computations. Over the temperature range 300 – 700 K, the average discrepancies between experimental and calculated resistivities are 5% (Ta), 360% (Nb),

13% (V) and 7% (Mo and W). The order of resistivity magnitudes seen in the experimental data at room temperature is exactly reproduced by the calculations if the calculated value for Nb is not considered.

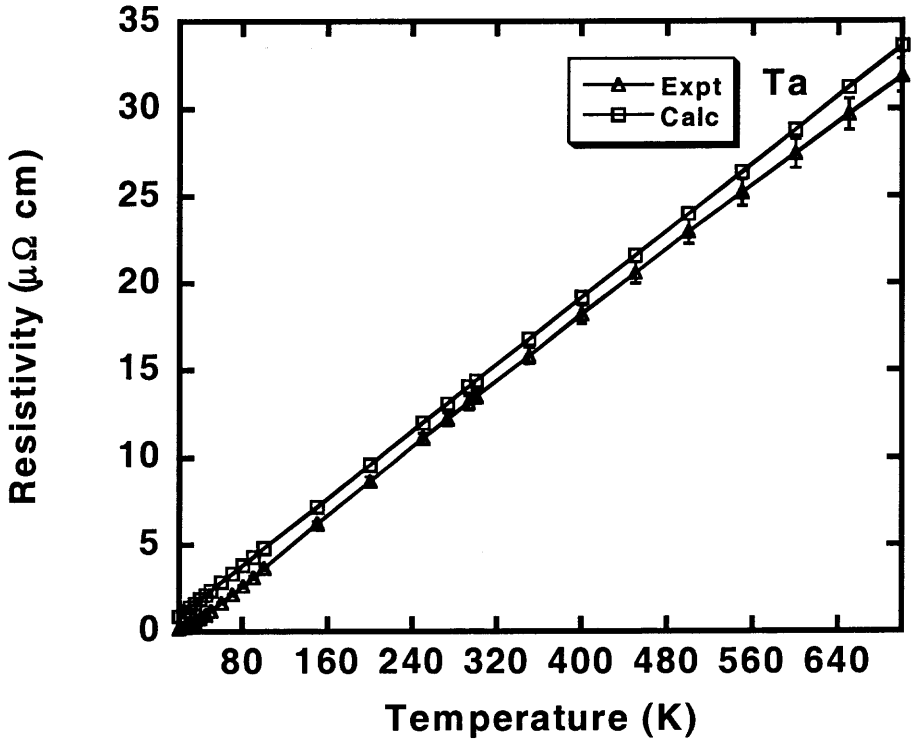


Figure 6.12 Measured and calculated values of the electrical resistivity for bcc Ta over the temperature range 20 – 700 K. The residual resistivity has been subtracted from the data.

The calculated resistivity curves appear linear over the entire temperature range and lie above the experimental curves at the lower end of the temperature scale for all the metals. The reasons for this are discussed below. As previously noted, over the temperature range 300 – 700 K, the average discrepancies between experimental and calculated resistivities are 5% (Ta), 360% (Nb), 13% (V) and 7% (Mo and W).

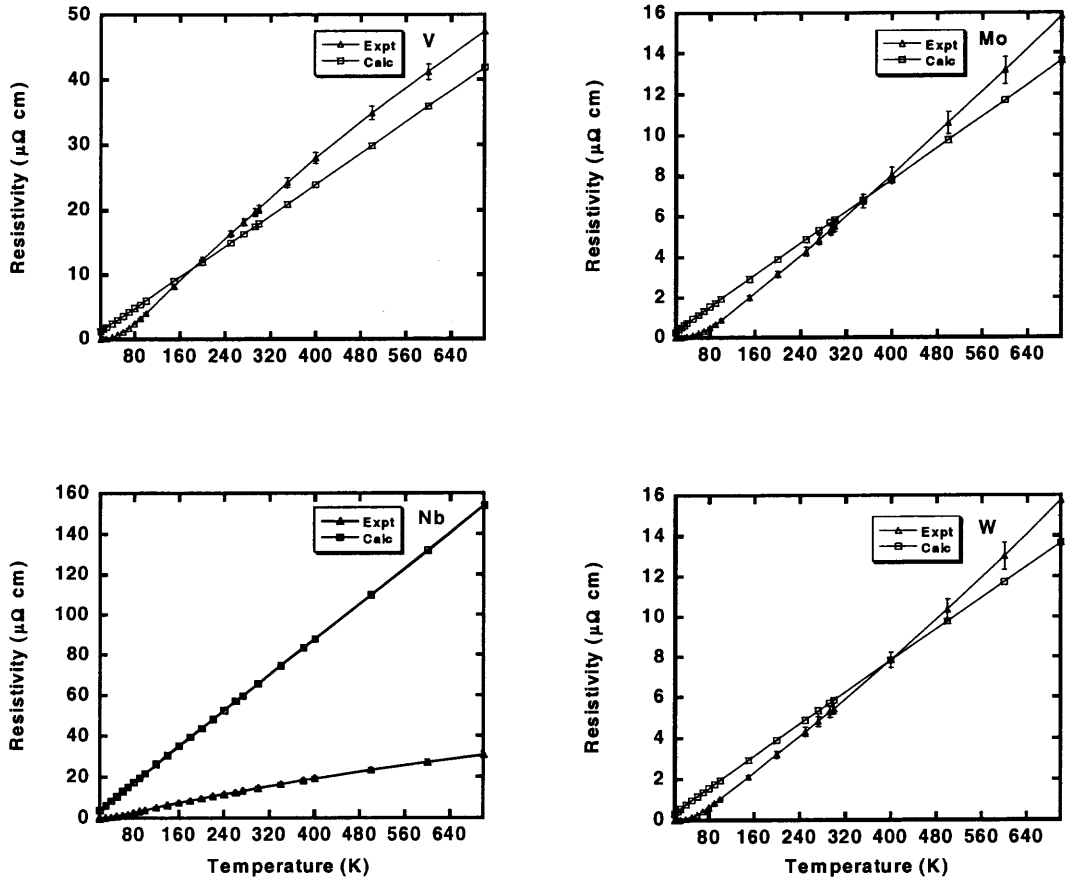


Figure 6.13 Measured and calculated values of the electrical resistivity over the temperature range 20 – 700 K are shown for the metals V, Nb, Mo and W. The residual resistivities have been subtracted from the data.

The discrepancies between the calculated resistivities and the resistivity data are attributed to the shortcomings of the LOVA model, the approximation which has been used to calculate the resistivities represented by Equation (6.2.2.27), and also to the approximation of λ_{tr} by λ. LOVA assumes that the Fermi surface displaces rigidly, and also that the electronic distribution function, and hence the electronic relaxation time, is energy independent and isotropic, the angular variation being proportional to v_{kx} [235]. Higher-order corrections to LOVA aim to incorporate the angular and energy dependence of the distribution function, as well as the effects of allowing different sheets of the Fermi

surface to displace differently, into resistivity calculations. Resistivity curves calculated using LOVA usually lie above the experimental curves at temperatures below 100K, and refinements to LOVA such as the n-sheet approximation, which allow different sheets of the Fermi surface to have different velocities, lower the calculated resistivity relative to experiment in the low temperature region [222]. The calculated resistivities appear linear because the linear term in Equation (6.2.2.27) dominates the other terms over the entire temperature range. The approximation of λ_{tr} by λ neglects the factor $\left[1-\frac{v_{kx}\cdot v_{k'x}}{|v_{kx}|^2}\right]$, which preferentially weights backscattering processes. The transition metals possess very complex, highly nested Fermi surfaces; therefore backscattering contributions to λ_{tr} could be significant and lead to λ_{tr} being significantly different from λ [236].

The calculated temperature dependent resistivity for Nb had the worst agreement with experiment of all the metals over the entire range of temperatures. This is attributed to the following factors in addition to those mentioned above. The calculated value of $N(\varepsilon_F)<v^2>$ for Nb, being smaller than usual [230], causes the calculated resistivity curve to shift significantly upward from the experimental curve. The large error in the value of $N(\varepsilon_F)<v^2>$ calculated for Nb might be due to the fact that Equation (6.3.5) was evaluated without taking into account spin-polarization effects. Superconductive tunneling experiments on Nb single crystals have revealed that the Nb Eliashberg spectral function $\alpha^2(\omega)F(\omega)$, which is related to the transport spectral function $\alpha^2{}_{tr}F(\omega)$ by Equations (6.2.2.30) and (6.2.2.32), is strongly anisotropic [237]. Due to this anisotropy, the errors introduced into the resistivity calculation by using LOVA could be more severe for Nb than for the rest of the studied metals. When the anisotropy and energy

dependence of the distribution function are accounted for in the Nb resistivity calculation, the calculated resistivities below 100 K are appreciably reduced relative to experiment and the LOVA results, while the calculated resistivities at higher temperatures are negligibly affected [222]. The measured temperature dependent resistivity of Nb is known to exhibit a significant negative deviation from linearity at high temperatures [223]. LOVA and higher order corrections to it neglect Fermi smearing effects, which are significant at higher temperatures, and which along with anharmonicity, are responsible for considerable negative deviations from linearity in the resistivity [222].

The metals with the largest values of the calculated room temperature resistivity are Nb and V, while the lowest values are obtained for W and Mo. The experimental resistivity data mirror this trend (Figures 6.12 and 6.13). Within each studied Periodic Table group, the order of the resistivity magnitudes is as follows: $\rho_{3d} > \rho_{4d} > \rho_{5d}$. The in-group order of resistivities is due to the magnitude of the d conduction bandwidth, which increases down the periods. $N(\varepsilon_F)$ is largely of d character for all the studied metals, so an increase in d bandwidth should correspond to a greater average velocity of electrons on the Fermi surface. This is confirmed by the calculated values of $< v^2 >^{1/2}$, which except for the Nb value, also increase in this direction. The Fermi energies of the group 5B metals V, Nb and Ta occur inside a high density of d bands, unlike the Fermi energies of the group 6B metals Cr, Mo and W, which occur near a minimum in the DOS (Figure 5.9). This means that the states in the range of the Fermi energy are more diffuse in the group 6B metals and have little contribution inside the muffin-tin spheres. The greater extent of the electrons with energies near ε_F in the group 6B bcc elements is

again confirmed by our calculated values of $< v^2 >^{1/2}$, which as was previously noted, are higher in each period for the group 6B elements than for the group 5B elements.

The room temperature resistivities of the group 6B metals are all less than those of the group 5B metals, so the combined effects of the d bandwidth and the position of ε_F in the band structure determine the order of resistivities. Applying this logic, the bct phases derived from the group 5B metals, including bct Ta, should in general be more conductive than their corresponding bcc phases, since the ε_F positions in the bct phases are shifted to lower DOS, while the d bandwidths are approximately the same as the bcc d bandwidths. Conversely the bct phases derived from the group 6B metals should generally be less conductive than their corresponding bcc phases, since the ε_F positions in these phases are shifted to higher DOS, without any appreciable changes in the bct d bandwidths relative to those of the bcc phases. These general conclusions can be confirmed for some of the metals by comparing the values of $N(\varepsilon_F) < v^2 >$ and $< v^2 >^{1/2}$ calculated for the bcc and bct phases. As previously noted, among the group 6B elements, $N(\varepsilon_F) < v^2 >$ values for the bct phases of W, Mo and Cr decreased by 95%, 60% and 36% respectively from the bcc values. The resistivities of the bct phases predicted for the group 6B metals Cr, Mo and W should therefore increase appreciably over the respective bcc phase values not only according to Equation (6.2.2.27), but also because the calculated bct value of $< v^2 >^{1/2}$ decreases from the bcc value by 37%, 50% and 83% for Cr, Mo and W respectively. Among the group 5B elements, $N(\varepsilon_F) < v^2 >$ values for the bct phases of V and Ta increased by 17% and 70% respectively over the bcc values, while the $N(\varepsilon_F) < v^2 >$ value for bct Nb decreased by 58% from the bcc

value. The calculated bct value of $< v^2 >^{1/2}$ increases over the bcc value by 21%, 160% and 31% for V, Nb and Ta respectively. The bct phases predicted for the studied group 5B metals V, Nb and Ta should therefore be more conductive than the respective bcc phases.

The calculated resistivities at room temperature, taken to be 293K, differ from experimental data by at most 15% (except in the case of Nb) and follow the trend in the order of magnitudes seen in the experimental data. The measured β–Ta resistivity is 1000 - 2000 % greater than that of α–Ta, therefore the calculations in this work can determine if the frequently observed β phase of Ta has the bct structure that has been predicted from the Bain transformation path in Chapter 5.

There is some evidence that impurities such as oxygen may stabilize β–Ta [91, 108, 110, 113 - 115]. Metastable phases of Ta have been observed to readily transform to the stable cubic phase upon heating, with a significant decrease in resistivity and oxygen impurity content [111, 152]. It has therefore been suggested [118] that the relatively high resistivity of β–Ta is due to the effects of incorporated impurities and its small grain size, which was determined to be about 45% less than that of bcc Ta. However studies of ion-facilitated metal film growth [238] have concluded that the decrease in grain size from bcc to beta Ta should only account for a twofold increase in resistivity, not the nearly tenfold increase that is observed. Furthermore, metastable Ta phases which have been prepared with low levels of contamination still exhibit high values of resistivity [91, 93] therefore the experimentally measured high resistivity of β–Ta has been attributed not to impurities and grain size effects, but to the intrinsic transport properties of its crystalline structure [118]. Uncertainty regarding impurity contributions to the resistivity of

metastable phases is not confined to Ta. Thermally induced structural transformation to the stable bcc phase has also been observed with metastable phases of tungsten [159]. There have however been conflicting reports regarding the oxygen content of bcc tungsten formed by heating relative to its metastable precursor [159 – 160], which is known to be about three times more resistive than bcc tungsten [162, 167].

In order to determine whether the bct phase of Ta predicted in this work is produced at any stage of the β–Ta formation process, the calculated resistivity of the bct phase is compared with resistivity values obtained from measurements on pure β–Ta films of varying thicknesses. The comparison can be carried out in two ways. The calculated transport properties of bct Ta and the measured resistivities of β–Ta can be used to obtain a range of empirical values for λ_{tr}, which should be on the order of unity. Alternatively, reasonable values of λ_{tr} for β–Ta can be used in Equation (6.2.2.26), together with the computed transport properties of bct Ta to directly compute the resistivity, which can be compared to the resistivity values measured for β–Ta. Using the value of $N(\varepsilon_F) < v^2 >$ computed for bct Ta and the measured resistivity of β–Ta at 400K [95] in Equation (6.2.2.26), an empirical λ_{tr} value of 20 is obtained for β–Ta, which is clearly unreasonable. Conversely, using a reasonable range of λ_{tr} values (0.5 – 1) in Equation (6.2.2.26) together with the value of $N(\varepsilon_F) < v^2 >$ computed for bct Ta results in a calculated bct Ta resistivity range of 6.68 – 13.38μΩ cm, which is less than the experimentally measured [234] value (18.22μΩ cm) and the value calculated (Figure 6.12) in this work (19.78μΩ cm) for bcc Ta, and therefore certainly less than the measured resistivity of β-Ta. This does not agree with numerous experimental findings

discussed previously. It is therefore highly probable that the bct Ta phase predicted in this work is not the same metastable (β) Ta phase that has been observed in so many experiments.

In the next Chapter, the experimental evidence concerning the effects of impurities and the grain size on the high measured resistivity of β-Ta is reviewed. Empirical calculations of the electronic mean free path employing the measured low temperature resistivity and band structure derived transport properties are used together with the experimental findings to elucidate the origin of the abnormally high measured resistivity of β-Ta.

CHAPTER 7

ORIGIN OF BETA TANTALUM RESISTIVITY

7.1 Structure of β-Ta

The efforts made over many years to determine the crystal structure of β-Ta have been discussed at length in Chapter 2. Out of the many structures that have been proposed to explain the observed properties of β-Ta such as the resistivity, two stand out as the most probable candidates. These structures belong to the $P4_2/mnm$ [106] and the $P\text{-}42_1m$ [126] space groups, and are both tetragonal with 30 atoms per unit cell.

7.1.1 P-42_1m β-Ta Structure

Single crystals of β-Ta have been produced [126] by electro-deposition from a molten fluoride bath containing oxofluoride melts with metallic Ta and TaO soluble anodes, in which K_2TaF_7 was used in the electrolyte as the Ta supplying component. The β-Ta crystals obtained were small (~5-20μm), and the $P\text{-}42_1m$ structure with tetragonal unit cell parameters of a=10.211Å and c= 5.3064Å was obtained with 1024 reflections from a diffractometer using Mo Kα radiation. The atomic parameters of this structure are given in Table 7.1.

Table 7.1 Atomic Positions in the P-42_1m Structure of β-Ta

Wyckoff Notation	Number of Atoms	Atomic Position Parameters		
		x	*y*	*z*
c	2	0.5	0	0.2280
f(1)	8	0.7598	0.0677	0.2350
f(2)	8	0.0343	0.1267	0.2550
e(1)	4	0.6033	0.1033	0.7640
e(2)	4	0.8142	0.3142	0.0030
e(3)	4	0.3196	0.1804	0.4910

Source: A. Arakcheeva, G. Chapuis, and V. Grinevitch, Acta. Cryst. B **58,** 1 (2002).

7.1.2 $P4_2/mnm$ β-Ta Structure

Moseley and Seabrook [106] prepared β-Ta by electrodeposition from a molten fluoride bath at 1073K. They recorded the X-ray powder pattern of the prepared samples and derived a tetragonal unit cell having the $P4_2/mnm$ structure with cell parameters of a= 10.194Å and c= 5.313Å. This structure is a CrFe-type structure [126] (the so-called Frank-Kaspar phase structures of binary compounds). The structure contains 30 atoms per unit cell in five inequivalent positions. The atomic positions involve three undetermined parameters. β-U is also believed to crystallize in the $P4_2/mnm$ structure. The structural parameters determined for β-U are adapted to the structure and then the resulting β-Ta structure is subjected to a force minimization process to refine the atomic positional parameters. The atomic position parameters adapted from β-U are shown in Table 7.2.

Table 7.2 Atomic Parameters Adapted from β-U $P4_2/mnm$ Structure and Refined in the β-Ta $P4_2/mnm$ Structure

Wyckoff Notation	Number of Atoms	β-U Atomic Position Parameters (Reference [119])			β-Ta Atomic Position Parameters After Force Refinement		
		x	y	z	x	y	z
b	2	-	-	-	-	-	-
f	4	0.1030	-	-	0.1050	-	-
j	8	0.3188	-	0.7444	0.3186	-	0.7493
i(1)	8	0.5622	0.2343	-	0.5673	0.2390	-
i(2)	8	0.3655	0.0391	-	0.3713	0.0347	-

7.2 Force Minimization of $P4_2/mnm$ Structure of β-Ta

The $P4_2/mnm$ structure of β-Ta has five inequivalent atomic positions, which are occupied by a total of 30 atoms. The first inequivalent atomic position is fixed by symmetry, as indicated in Table 7.2. The forces on the atoms with adjustable structural parameters are shown as functions of the number of force iterations in Figure 7.1. The parameters δ_m and η_m defined in Chapter 4 have been chosen to ensure that the final structure resulting from the force minimization calculations does not result from a local energy minimum.

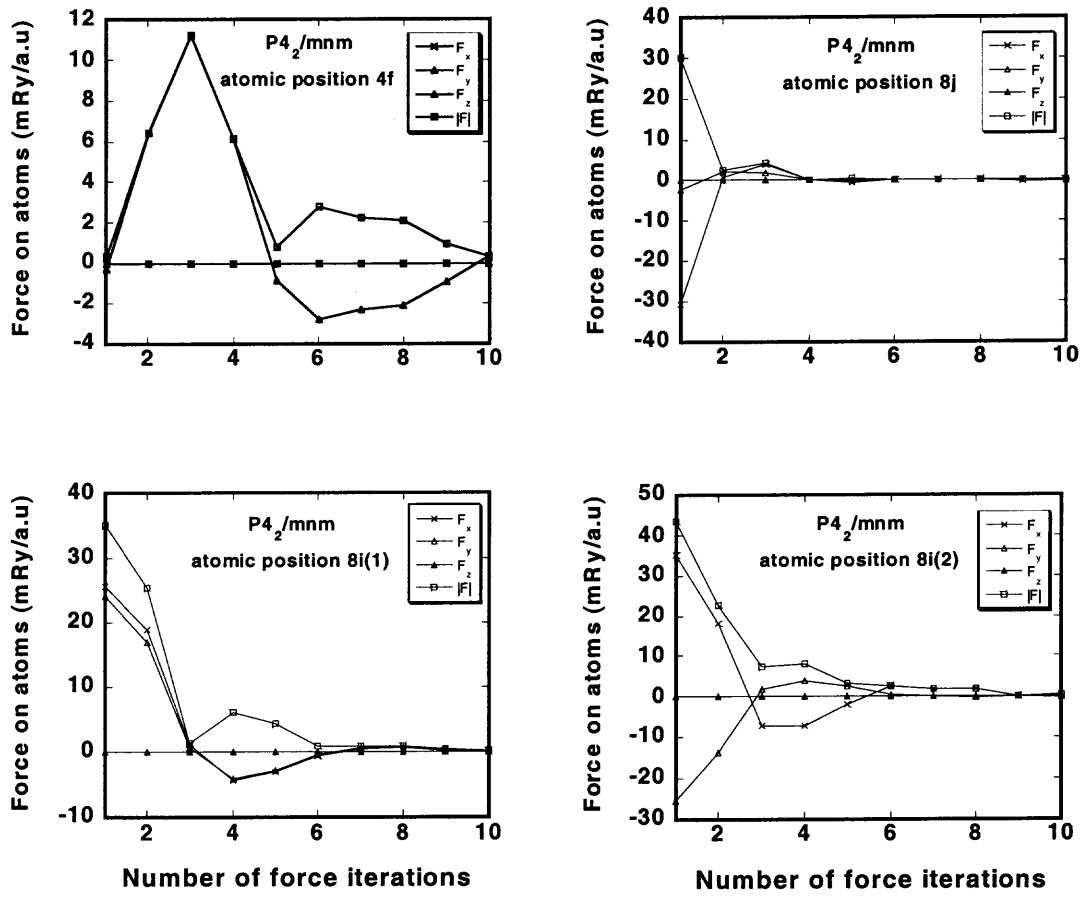


Figure 7.1 Forces on atoms with adjustable structural parameters as functions of number of force iterations.

The fastest force convergence is obtained for the atoms in the 8j setting (3rd inequivalent position), while the atoms in the 4f setting (2nd inequivalent position) need more force iterations to converge their forces to within the specified force tolerance criterion. The behavior of the total energy per unit cell as a function of the number of atomic force iterations is shown in Figure 7.2. There is a convergence to a total energy minimum within 10^{-4} Ry.

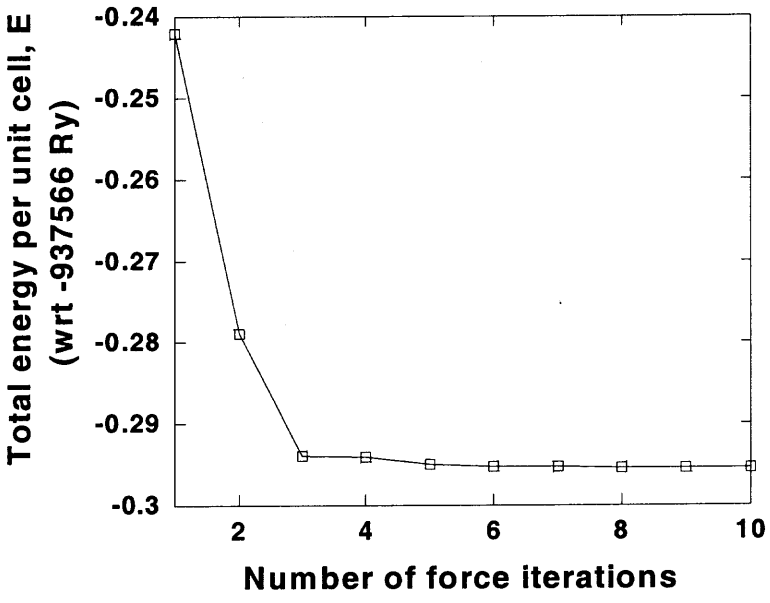


Figure 7.2 Total energy per unit cell of β-Ta $P4_2/mnm$ structure as a function of the number of force iterations. The forces are calculated using damped Newton dynamics in the LAPW formalism.

The next and next-nearest neighbor environments in the bcc, $P4_2/mnm$ and $P\text{-}42_1m$ structures are illustrated in Table 7.3, which lists all nearest and next-nearest neighbor distances. It can be seen from Table 7.3 that the $P4_2/mnm$ and $P\text{-}42_1m$ structures both have less local and long-range symmetry than the bcc structure. This will have implications for the band structure and transport properties, which can be calculated now that the structural parameters for β-Ta in the $P4_2/mnm$ structure have been refined.

Table 7.3 Atom Symmetry Types and their Nearest and Next-Nearest Neighbor Distribution in the Bcc, $P4_2/mnm$ and $P\text{-}42_1m$ Structures

Structure	Multiplicity of Atom Type	# of Nearest Neighbor Atoms	Distance to Nearest Neighbor Atoms (Å)	# of Next-Nearest Neighbor Atoms	Distance to Next-Nearest Neighbor Atoms (Å)
bcc Ta	1	8	2.878	6	3.324
β-Ta ($P4_2/mnm$)					
Position 1(2b)	2	4	2.747	4	2.933
Position 2(4f)	4	2	2.806	1	3.026
Position 3(8j)	8	1	2.649	1	2.662
Position 4(8i)	8	1	2.747	1	2.792
Position 5(8i)	8	1	2.717	1	2.806
β-Ta ($P\text{-}42_1m$)					
Position 1(2c)	2	4	2.741	2	2.878
Position 2(8f)	8	1	2.741	1	2.773
Position 3(8f)	8	1	2.680	1	2.847
Position 4(4e)	4	2	2.847	1	2.878
Position 5(4e)	4	1	2.622	1	2.685
Position 6(4e)	4	1	2.622	1	2.685

The $P4_2/mnm$ and $P\text{-}42_1m$ structures are both derived from a common parent space group and the $P4_2/mnm$ structure is centrosymmetric while the $P\text{-}42_1m$ structure is non-centrosymmetric. (001) views of the two structures are shown in Figure 7.3.

A study [239] of the bcc Ta and β-Ta structures on three distance scales, namely the local configuration of nearest neighbors around any given atom, the long-range crystalline order, and the average grain boundary size, concluded that the local

arrangement of nearest neighbors was more important than the long-range structural order in determining the electronic properties in both structures.

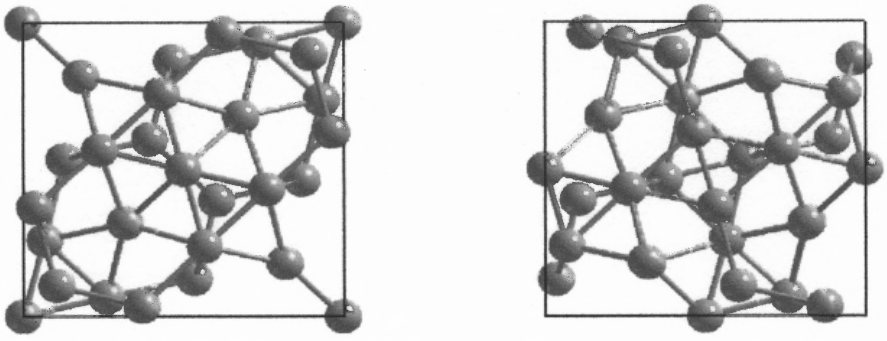


Figure 7.3 (001) views of the $P4_2/mnm$ (left panel) and $P\text{-}42_1m$ (right panel) structures.

The $P4_2/mnm$ and $P\text{-}42_1m$ structures are therefore very likely to have very similar electronic properties because from Table 7.3, the nearest neighbor atomic arrangement in the unit cell, which contains 30 atoms, is similar for both structures.

7.3 Electronic Structure of $P4_2/mnm$ Structure of β-Ta

The Brillouin zones of the bcc structure of Ta and the $P4_2/mnm$ structure of β-Ta are shown in Figure 7.4, with the high-symmetry points in the zones indicated. The band structure diagrams of β-Ta and bcc Ta along these high-symmetry directions in the Brillouin zone are shown in Figure 7.5. Apart from the much larger number of bands in the $P4_2/mnm$ band structure, there are no major differences between the two band structures. The $P4_2/mnm$ band structure is quite metallic, with no appreciable pseudogaps in the vicinity of the Fermi energy level.

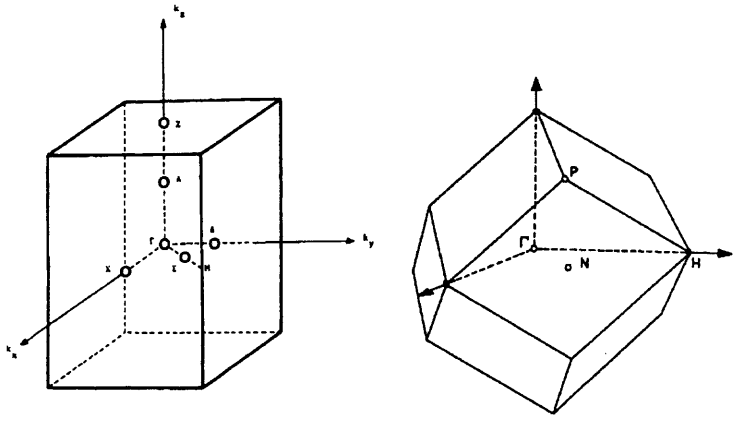


Figure 7.4 Brillouin zones for the $P4_2/mnm$ β-Ta structure (left panel) and the bcc Ta structure (right panel).

Both diagrams are calculated with the LAPW method. The resulting band structure of the 30 atom unit cell β-Ta $P4_2/mnm$ structure will necessarily be highly complex, as shown in Figure 7.5, with bands being split into several parts due to the close proximity of atoms to each other in the unit cell.

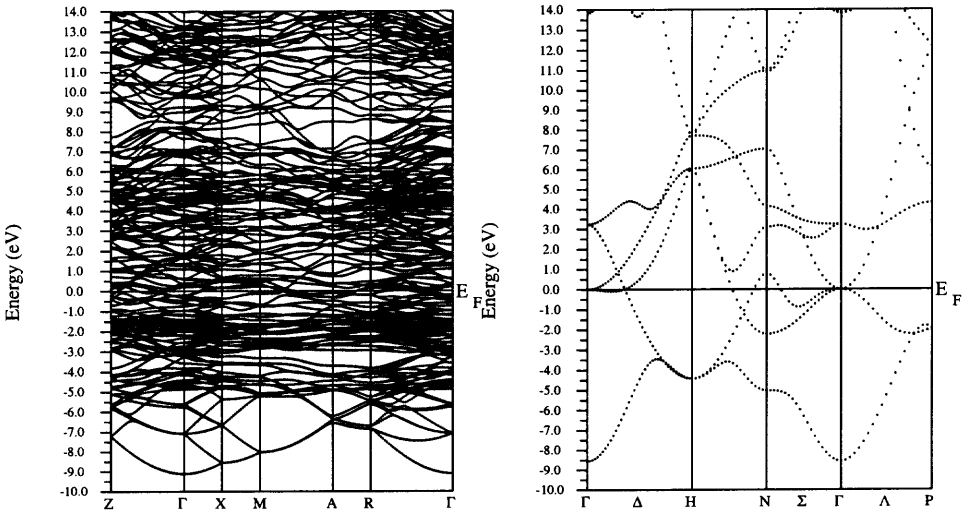


Figure 7.5 Energy band structure diagram along high symmetry directions in the Brillouin zone for the $P4_2/mnm$ β-Ta structure (left panel) and the bcc Ta structure (right panel).

The band structure of the $P4_2/mnm$ β-Ta structure closer to the Fermi energy level is shown in Figure 7.6.

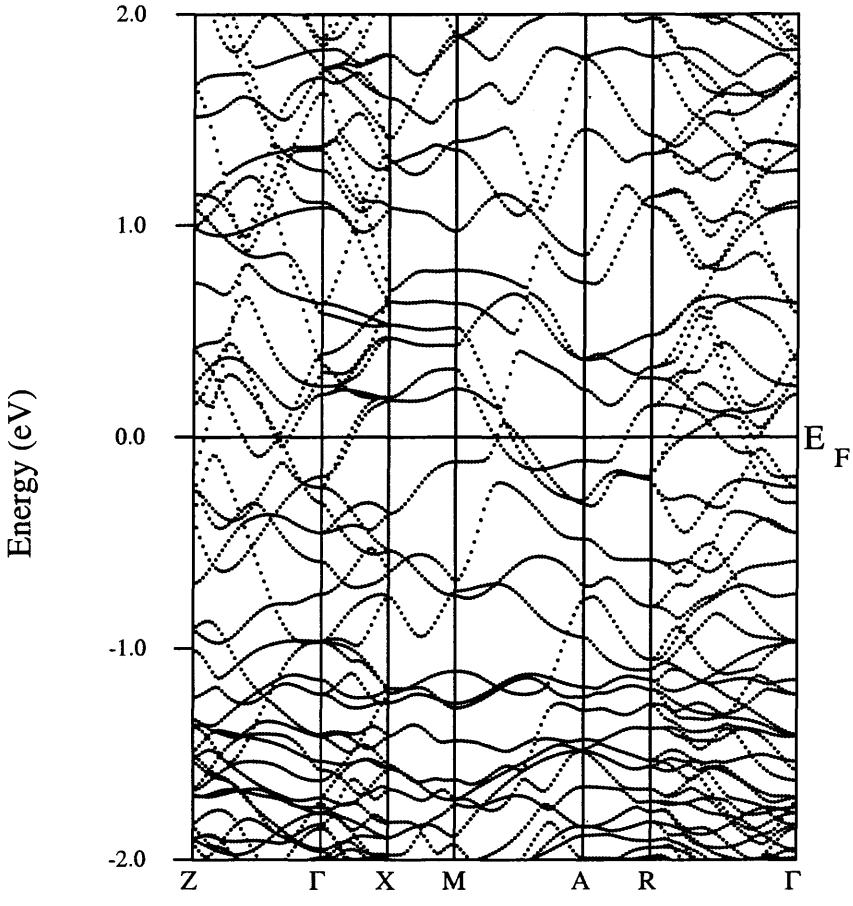


Figure 7.6 Energy band structure diagram along high symmetry directions in the Brillouin zone for the $P4_2/mnm$ β-Ta structure (expanded view near ε_F).

A higher number of bands cross the Fermi energy level in the $P4_2/mnm$ Ta structure, compared to the bcc Ta structure, therefore β-Ta in the $P4_2/mnm$ structure is very likely to have a Fermi surface that is highly fragmented compared to the bcc Ta Fermi surface. The increased fragmentation of the Fermi surface might enhance the possibilities for umklapp scattering processes and thereby increase the probability of large-angle electron scattering near the Fermi surface while decreasing the mean free path

in β-Ta. The main branches of the Fermi surface in the first Brillouin zone of the $P4_2/mnm$ structure are shown in Figure 7.7(a).

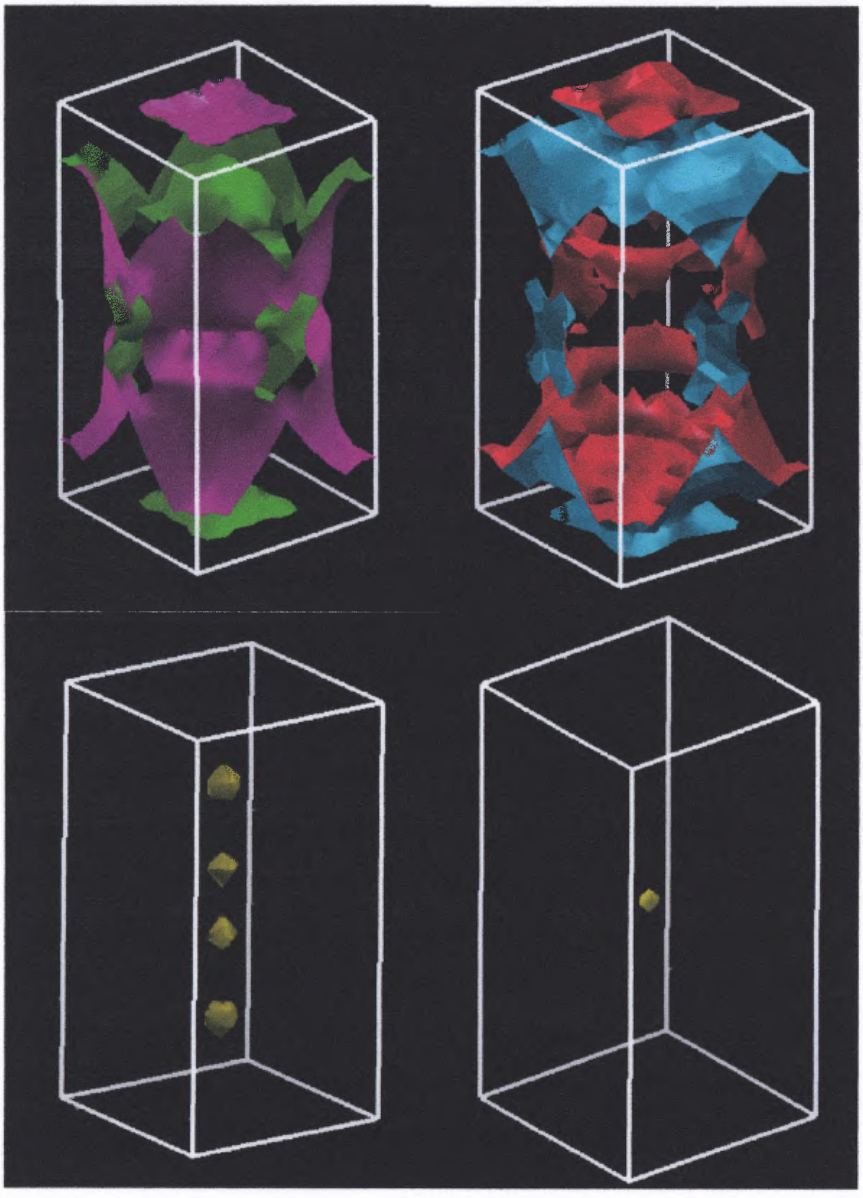


Figure 7.7(a) Branches of the Fermi surface in the first Brillouin zone of β-Ta $P4_2/mnm$ structure. Except for the branch of the surface in the lower right part of the Figure, all branches occur in bands between 0.88 and 0.91 Ry in energy.

The $P4_2/mnm$ Fermi surface exhibits more fragmentation than the bcc Ta Fermi surface, which is shown in Figure 7.7(b). The hole surface enclosing unoccupied levels in the second Brillouin zone of the bcc Ta crystal structure is shown on the left panel of Figure 7.7(b), while the hole-containing ellipsoids enclosing unoccupied levels in the third Brillouin zone [233] can be seen in the right panel of Figure 7.7(b).

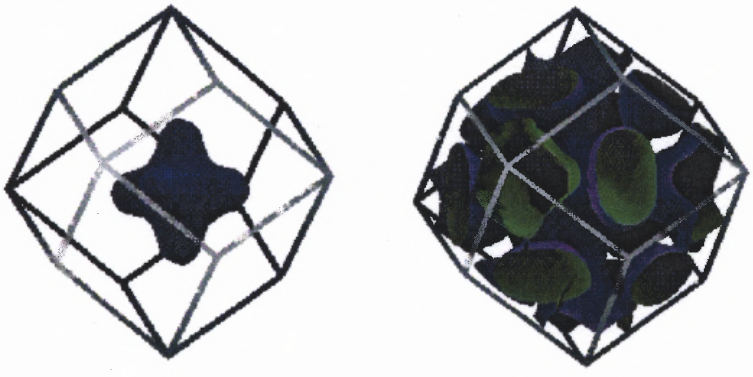


Figure 7.7(b) Branches of the Fermi surface in the first Brillouin zone of the bcc Ta structure. The branch of the surface on the left occurs in bands lying between 0.57 – 0.9 Ry, while the branch of the surface on the right occurs in bands lying between 0.73 – 1.33 Ry.

The hole surface enclosing unoccupied levels in the second Brillouin zone of the bcc Ta structure, which is contained entirely in a single band, is not present in the $P4_2/mnm$ Fermi surface. The hole-containing ellipsoids shown in the right panel of Figure 7.7(b) are also not present in the $P4_2/mnm$ Fermi surface. The hole surfaces are presumably split by the lower symmetry of the $P4_2/mnm$ structure, where they occur over several bands. The split hole surface part of the $P4_2/mnm$ Fermi surface has the form shown in the lower half of Figure 7.7(a).

The density of states (DOS) figure of the $P4_2/mnm$ β-Ta structure as a function of energy is shown in Figure 7.8. Two main peaks occur at –0.4 and –3.8eV below the Fermi level.

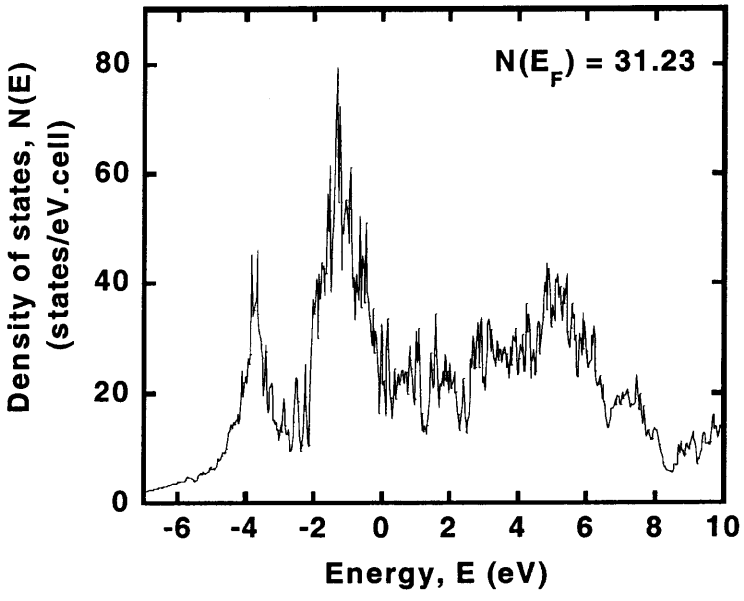


Figure 7.8 Total DOS as a function of energy for the $P4_2/mnm$ β-Ta structure. The Fermi energy is at 0 eV.

From the DOS plot in Figure 7.8, β-Ta in the $P4_2/mnm$ structure has a value of 31.23 states/eV.cell for the total density of states at the Fermi level, a value much higher than the bcc Ta value of 1.34 states/eV.cell. β-Ta and bcc Ta contain 30 atoms and 1 atom respectively per primitive unit cell, but the density of states at the Fermi level per atom is quite similar, being 1.04 and 1.31 states/eV.atom for β-Ta and bcc Ta respectively. The DOS as a function of energy is shown for both structures in Figure 7.9. As previously noted from Table 7.3, β-Ta in the $P4_2/mnm$ structure exhibits much less local symmetry than the bcc structure. The effects of the lower symmetry can be seen from Figure 7.9 as some of the features in the $P4_2/mnm$ DOS roughly corresponding in energy to sharp

features in the bcc DOS are relatively smeared out and reduced in intensity in the $P4_2/mnm$ DOS.

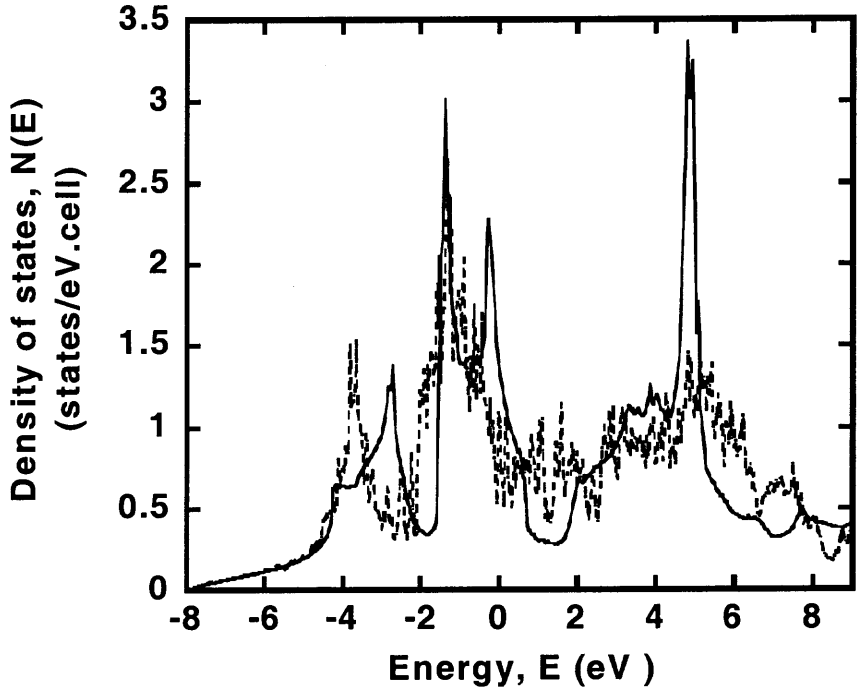


Figure 7.9 Total DOS as a function of energy for the $P4_2/mnm$ β-Ta and the bcc Ta structures. The full and dashed lines represent the bcc and the $P4_2/mnm$ DOS respectively.

The *d* DOS, which contributes the most to the total DOS, is shown in Figure 7.10 for each inequivalent atomic site in the β-Ta $P4_2/mnm$ structure. Some atomic positions in the $P4_2/mnm$ β-Ta structure have some DOS features sharpened relative to the bcc DOS, while other atomic positions have their DOS features smeared out relative to the bcc DOS, reflecting the dissimilar environments experienced by the different inequivalent atomic positions in the $P4_2/mnm$ structure of β-Ta. The overall shape of the total β-Ta DOS is however somewhat similar to the total bcc DOS.

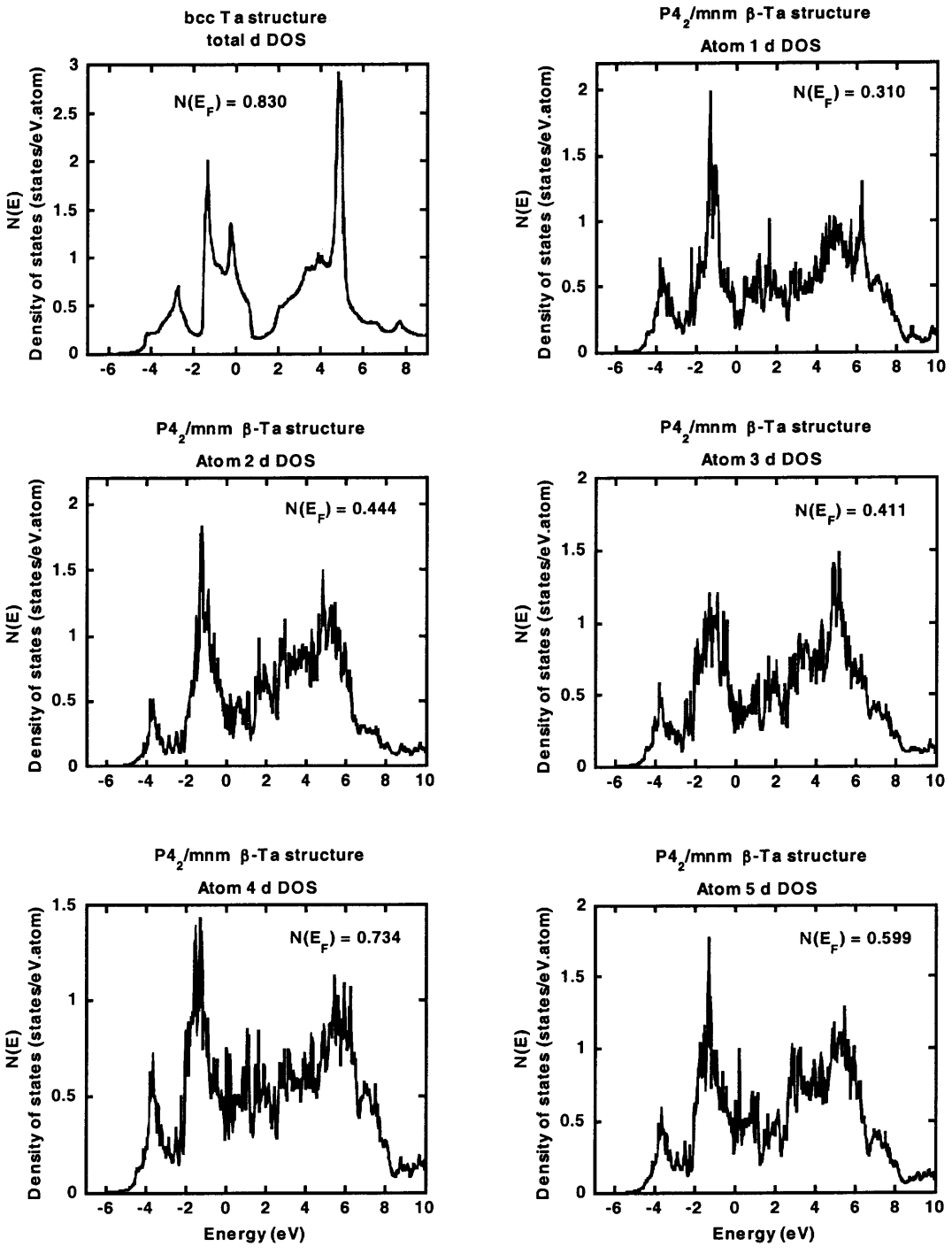


Figure 7.10 The *d* DOS as a function of energy for bcc Ta and the five inequivalent atomic positions in the $P4_2/mnm$ β-Ta structure.

To investigate whether β-Ta is a pressure stabilized phase, the f character DOS is examined in Figure 7.11. It can be seen that the f bandwidth in β-Ta is reduced relative to that in bcc Ta, so that $f-d$ hybridization, which would result under pressure-stabilized conditions, does not occur. Further confirmation of the absence of $f-d$ hybridization in β-Ta is given by the slight increase in atomic volume in the $P4_2/mnm$ β-Ta structure over the bcc Ta atomic volume. $f-d$ hybridization in β-Ta would be expected to result in a reduced atomic volume relative to bcc Ta. As a consequence of the absence of $f-d$ hybridization as demonstrated by the band structure, β-Ta in the $P4_2/mnm$ structure is probably not a pressure-stabilized phase

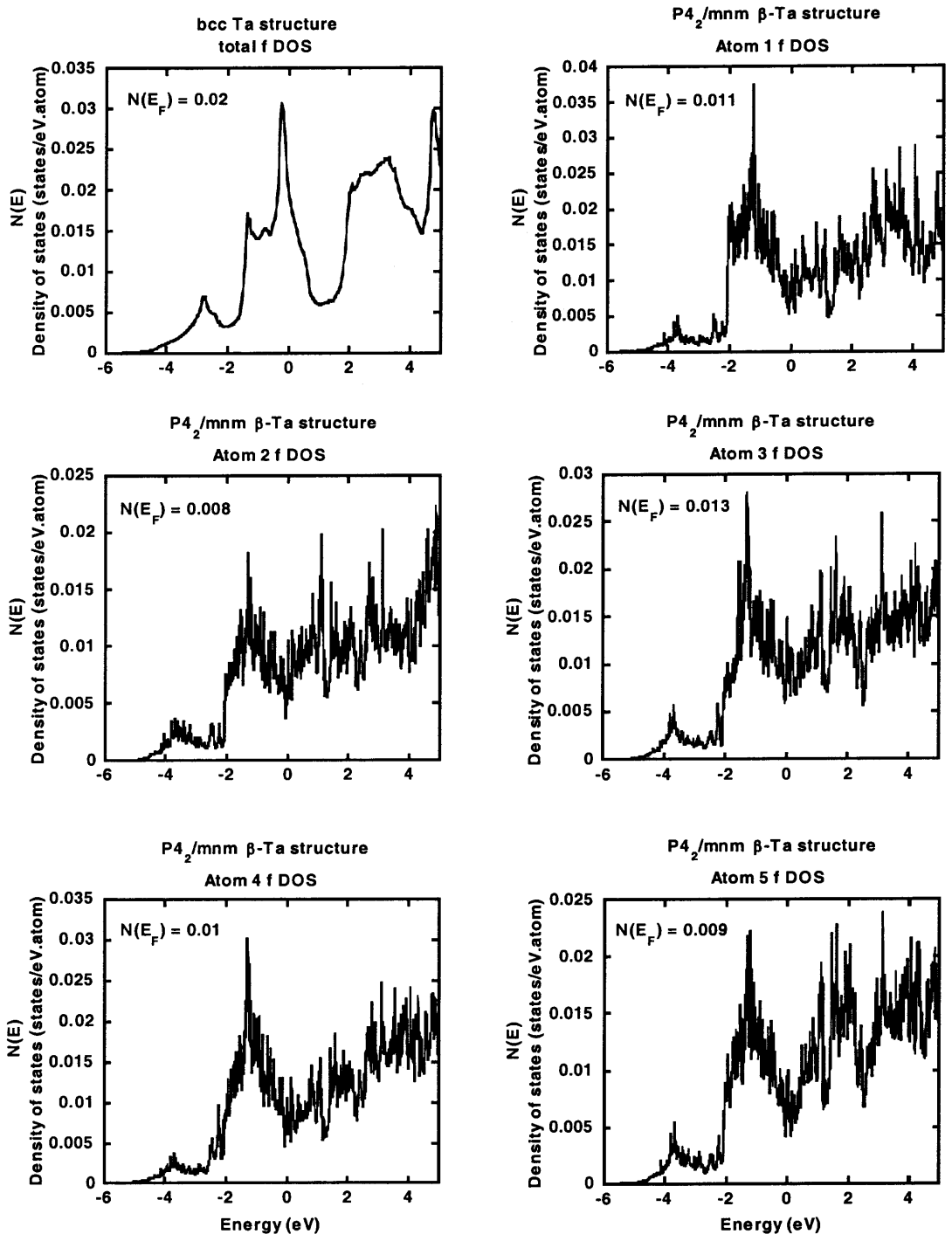


Figure 7.11 The *f* DOS as a function of energy for bcc Ta and the five inequivalent atomic positions in the $P4_2/mnm$ β-Ta structure.

7.4 Electrical Resistivity of β-Ta

As discussed in Chapter 2, numerous workers have measured the electrical resistivity of β-Ta, obtaining resistivities in the range of 170 – 270μΩcm. Compared to the bcc Ta resistivity value of 13.6μΩcm, the measured β-Ta resistivity values are quite high. The temperature-dependent resistivity measured for several samples of β-Ta from 4.2K to 300K [95] is shown in Figure 7.12. The maximum change in resistivity over the entire temperature range is only 6%.

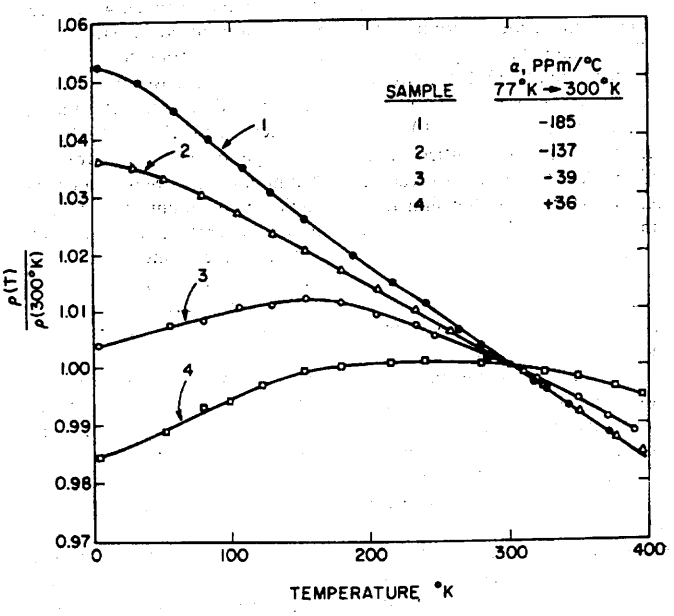


Figure 7.12 Ratio of the resistivity at a given temperature to the resistivity at 300K as a function of temperature for four different β-Ta film samples (taken from Reference [95]).

Samples 1 and 2 in the figure were determined [95] to be pure β-Ta using X-ray diffraction while samples 3 and 4 were determined to contain some bcc Ta, which accounts for the maxima in the corresponding resistivity curves in Figure 7.12. Figure 7.13 [95] shows the resistivity of β-Ta compared to the resistivity of the metals with the

two highest known resistivities at room temperature, α-plutonium (α-Pu) and α-manganese (α-Mn), which have resistivities of 143μΩcm and 134μΩcm respectively at room temperature. The electronic structure of β-Ta as represented by Figures 7.5, 7.6, 7.7 and 7.8 is typical of a normal metal; therefore it is not apparent from the data why β-Ta exhibits this behavior.

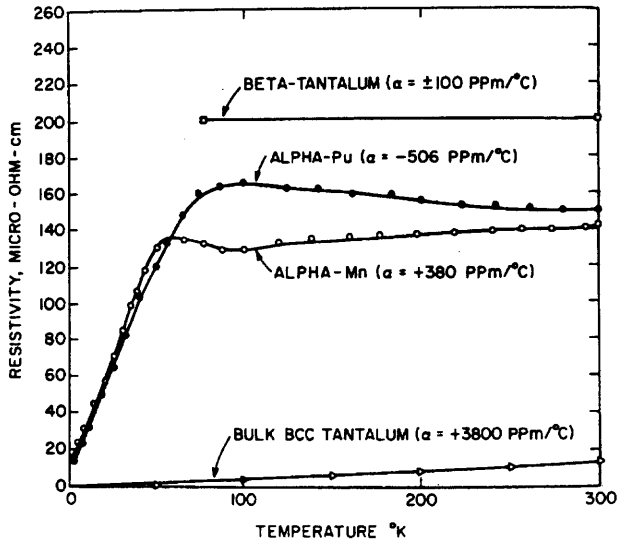


Figure 7.13 The resistivities of β-Ta, bulk α-Pu, bulk α-Mn, and bulk bcc Ta as a function of temperature (taken from Reference [95]).

To attempt to resolve this issue, transport properties of β-Ta in the $P4_2/mnm$ structure will be calculated and the electrical resistivity will be estimated according to Equation (6.3.2.27). Due to the tetragonal symmetry of the structure, the relaxation time τ will be anisotropic and can be represented as τ_{r_i}, given by [226]

$$\tau_{r_i} = \frac{4\pi}{\rho\left\langle \omega_{p_{r_i}}^2 \right\rangle}, \tag{7.4.1}$$

in direction r_i. $N(\varepsilon) < v^2 >$, $N(\varepsilon) < v_x^2 >$, $N(\varepsilon) < v_y^2 >$ and $N(\varepsilon) < v_z^2 >$ as functions of energy for the β-Ta $P4_2/mnm$ structure are shown in Figure 7.14.

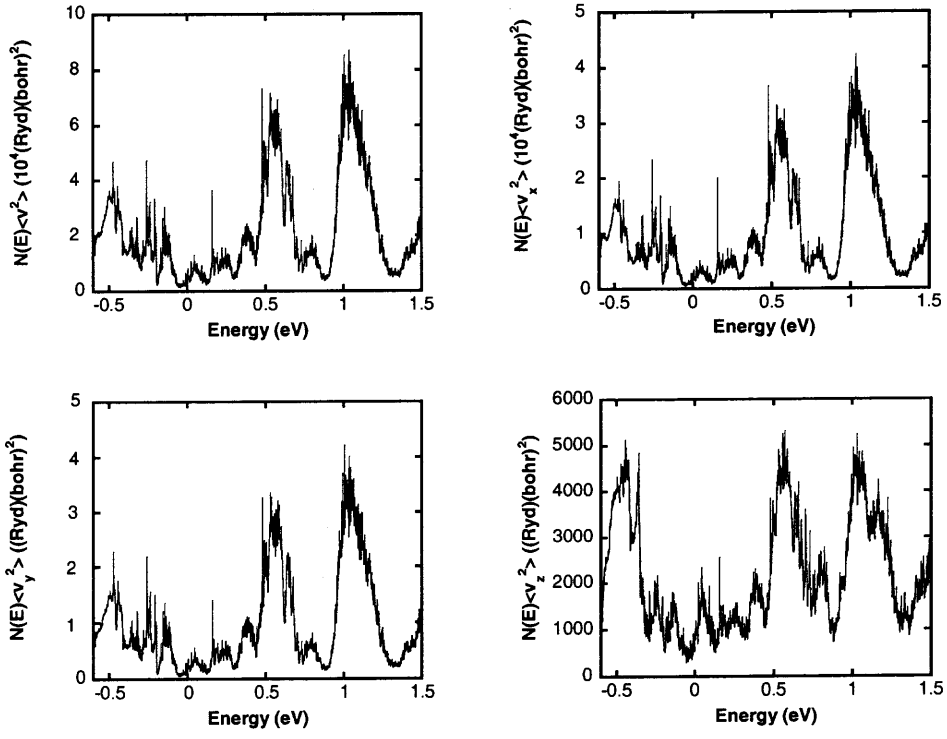


Figure 7.14 $N(\varepsilon) < v^2 >$, $N(\varepsilon) < v_x^2 >$, $N(\varepsilon) < v_y^2 >$ and $N(\varepsilon) < v_z^2 >$ as functions of energy for the $P4_2/mnm$ β-Ta structure. The Fermi energy is at 0 eV.

The calculation of $N(\varepsilon) < v^2 >$ is carried out using 252 k-points in the irreducible wedge of the Brillouin zone. A value for $N(\varepsilon_F) < v^2 >$ of 89.16Ry. bohr^2 is calculated from Figure 7.14 for β-Ta in the $P4_2/mnm$ structure. This yields a value of 5.01×10^8 cm/s for $\langle v^2 \rangle^{1/2}$, the square root of the mean squared Fermi velocity, for the $P4_2/mnm$ structure. This value and values of $\langle v_x^2 \rangle^{1/2}$, $\langle v_y^2 \rangle^{1/2}$ and $\langle v_z^2 \rangle^{1/2}$ for the bcc and $P4_2/mnm$ structures of Ta are listed in Table 7.4.

Table 7.4 Fermi Velocity and Plasma Frequency Values Calculated for $P4_2/mnm$ and Bcc Structures of Ta

	$\langle v^2 \rangle^{1/2}$ (10^7 cm/s)	$\langle v_x^2 \rangle^{1/2}$ (10^7 cm/s)	$\langle v_y^2 \rangle^{1/2}$ (10^7 cm/s)	$\langle v_z^2 \rangle^{1/2}$ (10^7 cm/s)	$\langle \omega_p^2 \rangle^{1/2}$ (s^{-1})	$\langle \omega_{p_x}^2 \rangle^{1/2}$ (s^{-1})	$\langle \omega_{p_y}^2 \rangle^{1/2}$ (s^{-1})	$\langle \omega_{p_z}^2 \rangle^{1/2}$ (s^{-1})
β-Ta ($P4_2/mnm$)	5.01	3.1	3.18	2.32	6.09	3.77	21.17	15.48
Bcc Ta	6.04	3.49	3.49	3.49	8.34	4.82	4.82	4.82

The mean free path l is given by $l = \langle v^2 \rangle^{1/2} \tau$, so the $P4_2/mnm$ structure will have an intrinsic mean free path similar to that in the bcc structure, for which a $\langle v^2 \rangle^{1/2}$ value of 6.04×10^7 cm/s is calculated. Assuming a spherical Fermi surface, the mean free path can be estimated using the relation [240]

$$\rho = \frac{3\pi^2 \hbar}{e^2 k_F^2 l}, \tag{7.4.2}$$

where k_F, the Fermi wave vector, is given by [178]

$$k_F = \sqrt{\frac{2m\varepsilon_F}{\hbar^2}}. \tag{7.4.3}$$

Using the measured resistivity of β-Ta at 4.2K from Figure 7.12 and an ε_F value of 0.893Ry, a value of ~ 2Å is obtained for the mean free path at low temperatures. The measured resistivity in Figure 7.12 changes by a maximum of +6% from 4.2K to 300K and is linear throughout this temperature range, so scattering by phonons can be ruled out as a dominant factor in determining the mean free path and hence the room temperature

resistivity. Furthermore, because there is such a small change in the measured resistivity of β-Ta between 4.2K – 300K, the mean free path at low temperatures will remain practically unchanged from its value at 300K, so the mean free path estimated at 4.2K can be used at 300K, without introducing much error.

7.4.1 Estimation of β-Ta Room Temperature Resistivity

Without renormalization of the electron-phonon coupling constant λ by the density of states at the Fermi level, $N(\varepsilon_F)$, a room temperature resistivity of 26.4μΩcm is obtained from Equation (6.2.2.27) for the $P4_2/mnm$ structure of β-Ta based on the calculated $N(\varepsilon_F)<v^2>$ value of 89.16Ry. bohr^2. Renormalizing the electron-phonon coupling constant according to the relation

$$1+\lambda = 3\gamma/2\pi^2 k_B N(\varepsilon_F), \tag{7.4.1.1}$$

where γ is the measured electronic specific-heat coefficient. β-Ta and bcc Ta have roughly the same value for the density (16.3 and 16.6 g cm $^{-3}$ respectively); so γ can be taken to be approximately the same for β-Ta and bcc Ta, without introducing appreciable error into calculations that depend on γ. The electron-phonon coupling constant λ_β for β-Ta can therefore be obtained from

$$\lambda_\beta = \frac{N_{bcc}(\varepsilon_F)}{N_\beta(\varepsilon_F)}(1+\lambda_{bcc})-1, \tag{7.4.1.2}$$

where $N_{bcc}(\varepsilon_F)$ is the total density of states at the Fermi energy for the bcc Ta structure (in states/eV.atom), $N_{\beta}(\varepsilon_F)$ is the total density of states at the Fermi energy for the $P4_2/mnm$ β-Ta structure and λ_{bcc} is the electron-phonon coupling constant for bcc Ta. Using a value of 0.83 [241] for λ_{bcc}, a value of 1.36 for λ_{β} is obtained from Equation (7.4.3). Using these values in Equation (6.2.2.27), a calculated room temperature resistivity value of 35.9μΩcm is obtained for the $P4_2/mnm$ structure of β-Ta. This calculated resistivity value is much lower than the reported values of the measured β-Ta resistivity, suggesting that other forms of scattering dominate the electron-phonon scattering contribution to the resistivity. It remains to determine which factor between grain boundary scattering and impurity scattering is more influential in determining the mean free path and the magnitude of the observed resistivity.

7.4.2 Importance of Grain Boundary and Impurity Scattering

β-Ta has a tendency to form in contaminated sputtering and vacuum evaporation environments, therefore it has not been ascertained whether the consistently high resistivity values measured for it are due to incorporated impurities, grain boundary effects or to intrinsic properties of the β-Ta crystal structure. Grain boundaries are bounding surfaces between crystals and as such are planar defects in the crystalline structure that reduce its periodicity. Grain boundary–electron scattering phenomena can contribute significantly to the electrical resistivity of a metallic sample when the mean free path of a conduction electron is comparable to the average grain size in the sample. Over the decades, grain sizes of various β-Ta samples have been measured by several authors using different experimental techniques and are listed in Table 7.5. The measured

grain sizes are at least an order of magnitude greater than the estimated low-temperature mean free path, therefore the mean free path in β-Ta is probably not limited by electrons scattering off grain boundaries.

Table 7.5 Experimentally Determined β-Ta Grain Sizes

Reference	β-Ta Grain Size (Å)	Grain Size Determination Method
Mills, 1965	350	TEM
Face, 1987	300	TEM
Sajovec, 1992	79	TEM
Kwon, 1997	90	HREM
Stavrev, 1997	900	XRD
Yohannan, 2001	1200	TEM

Source: A. Yohannan, *Characterization of α and β Phases of Tantalum Coatings*, Masters Thesis, New Jersey Institute of Technology (2001).

Thin films in general are able to dissolve more impurities than the corresponding bulk material. β-Ta can dissolve up to 22% oxygen by weight, resulting in a lattice parameter variation of up to 2% [115]. β-Ta films have been observed to be prepared exclusively when the argon used as a sputtering gas is mixed with oxygen [81]. The electrical resistivity of these films was also found to increase appreciably with an increase in oxygen concentration. Thin film deposition experiments [245] in an RF sputtering system have concluded that impurity atoms adsorbed on the substrate surface, or reactive gases present in the sputtering atmosphere, are necessary for the nucleation and growth of β-Ta. In fact one proposed structure [114] of β-Ta is believed to incorporate nitrogen impurities. An increase in ρ has been observed [88] for thin Ta

films evaporated onto sodium chloride, alumina, glass and fused quartz substrates as the deposited films were exposed to oxygen and nitrogen, however, the temperature dependence of ρ was unaffected. The residual resistance ratio $\rho_{300K}/\rho_{4.2K}$ of thin Ta films deposited by a getter sputtering system on MgO substrates has been observed [92] to increase from 30 to 1800 upon film annealing. The drop in $\rho_{300K}/\rho_{4.2K}$ was attributed to impurity removal. In addition, polycrystalline metals such as gold, platinum, rhenium, beryllium and tungsten with strong resistance to oxidation below 673K have been observed to nucleate the bcc phase of Ta in sputtering processes. Conversely, β-Ta seems to nucleate preferentially on polycrystalline metals susceptible to oxidation [110]. Due to these observations, impurities have long been believed to influence the electrical properties of β-Ta.

Several authors have prepared β-Ta in impurity-rich environments and have proposed [91, 108, 110, 113 – 115] that it is stabilized by these impurities. Other works [118] have suggested that impurities incorporated into the β-Ta crystal structure are the dominant factor responsible for its abnormally high values of measured electrical resistivity. The estimated mean free path of β-Ta is more in keeping with these views because it is of the same order of magnitude as the distance between a Ta atom and an interstitial impurity (~ 3Å). It is therefore highly likely that the high values of the electrical resistivity measured for β-Ta are not due to intrinsic properties of the β-Ta crystal structure, such as for example its highly fragmented Fermi surface, or to grain boundary effects, but are probably attributable to impurities incorporated into its crystal structure. This would also seem to support the view that impurities are necessary for the nucleation and growth of β-Ta films.

APPENDIX

TANTALUM THIN FILM PREPARATION DETAILS

The preparation details of Ta thin films produced by various authors are listed in tabular form on the following pages.

Reference	Deposition Method	Deposition Pressure (torr)	Deposition Rate (Å/min)	Substrate	Substrate Temp. (°C)	Film Structure	Resistivity (μΩ cm)	Details	Reactive Sputtering
Gerstenberg and Mayer, 1962	5 k V 0.25 mA/cm^2	1.5 x 10^{-2}	120	Glass		bcc	50		N_2
Altman, 1962	3-6 k V 0.22-0.87mA/cm^2	7 x 10^{-3}- 7 x 10^{-2}	77-430	Pyrex	400	bcc	33	ρα ~ 0.05 for films 180-4,540 Å thick, DC sputtering	
Gerstenberg and Calbick, 1964	5 k V 0.25 mA/cm^2	2 x 10^{-2}	120	Soft glass	400	bcc	50		O_2, N, CH_4
Read and Altman, 1965		2 x 10^{-2}				β	180-220	X-ray diffraction analysis Films 100-20,000 Å thick, DC sputtering	
Krikorian and Sneed, 1966	1-6 k V 0.098-0.65 mA/cm^2	5 x 10^{-3} 1.7 x 10^{-1}	50-300	7059*	400, 450	bcc	30-8 x 10^3		
Calbick, 1966	1.5 k V	7 x 10^{-2}-10^{-1}		Carbon coated 7059		Possibly β		DC sputtering	
Cook, 1967	4-6 k V 0.3-0.5 mA/cm^2	~ 10^{-2}	~ 200	7059		β	200-700		
Vratny, Vromen and Harendza-Harinxma, 1967	4-5 k V In line and batch systems	2-3 x 10^{-2}		7059		β	180-190		
Sosniak, Polito and Rozgonyi, 1967	Oil diff. pump Sputter Ion and Ti Sub. Sputter Ion and Ti Sub.	2 x 10^{-2} 2-8 x 10^{-2} 2-8 x 10^{-2}	120 35-80 35-80	7059 7059 7059	20-400 20-550 340-380	β+bcc β, bcc β	28-187 38-337 112-207	DC sputtering Up to 15% b.c.c. Phases separate or mixed	
Coyne and Tauber, 1968	4.5 kV 0.079-0.63 mA/cm^2	1.5-4.5 x 10^{-3}	10-64	Quartz				deposited in pure N_2	
Sosniak, 1968	Diffusion pump and turbomolecular pump system	1-2 x 10^{-2}	120	7059		β +bcc traces	140-180	Background gases monitored during DC sputtering	

*7059 – 7059 Corning glass

Reference	Deposition Method	Deposition Pressure (torr)	Deposition Rate (Å/min)	Substrate	Substrate Temp. (°C)	Film Structure	Resistivity (μΩ cm)	Details	Reactive Sputtering
Westwood and Waterhouse, 1971	3.5 k V 0.3 mA/cm^2	2 x 10^{-2}		7059	~ 250	β + bcc	~ 250	Proportion of β-Ta increased with flow rate of O_2-Ar mixture, DC sputtering	O_2
Schauer and Peters, 1974	0.5 kV, 2.5 A	7 x 10^{-4}	93	Ta_2O_5/Glass		β + bcc	30-180	RF sputtering/electrodeless ring discharge	
Thornton and Hoffman, 1976	0.35-1 kV	9.7-300 x 10^{-4}	600	Glass		bcc	100	DC magnetron sputtering	
Sato, Inoue and Sasaki, 1981	5 kV 1 kW	5 x 10^{-2} 5 x 10^{-3}	200 1000	Glazed ceramic Glazed ceramic		bcc + TaO_x β	150 190	DC diode sputtering DC magnetron sputtering	
Face and Prober, 1987	1.5 kV 35 mA/cm^2	1.5 x 10^{-4}	180	Si Nb(<30Å thick) Nb(>30Å thick)		β bcc+β bcc	150	Ion beam sputtering, rotating target	
Sajovec, 1992	1 kV 20 mA	2.3 x 10^{-4}	36	Si Nb(<30Å thick) Nb(>30Å thick)		β bcc+β bcc	200-220 40-55	Ion beam sputtering, rotating target	
Catania, Roy and Cuomo, 1993	0.47 kV, 0 V bias 0.47 kV, -100 V bias	3 x 10^{-4}	500	Si wafer		bcc β	20–28 170	DC magnetron sputtering	
Matson et al, 2000		3 x 10^{-3}		4340 steel	202-210 180-230	β bcc		Triode sputtering w/ Ar Triode sputtering w/ Ar, Xe, Kr	
Lee et al, 2001	0.05-1.5kV 250 mA	4 x 10^{-3}	3600	4340 steel	250	β + bcc		Triode sputtering w/ Kr	
Liu et al, 2001		1.8 x 10^{-2}	960	Si wafer		β	170	DC magnetron Ar gas, 200Å grain size	
Lee et al, 2002	10-150W	5-100 x 10^{-3}	10-60	Glass		β + bcc		360-2500Å thick Ar gas	

REFERENCES

[1] Tantalum-Niobium International Study Center, Bulletin **110**, (2002).

[2] K. Andersson, K. Reichert, and R. Wolf, Ullmann, Encyclopedia of Industrial Chemistry **A26**, 71 (1995).

[3] X. J. Yang and C. Pin, Anal. Chim. Acta **458**, 375 (2002).

[4] O. M. El-Hussaini and M. A. Mahdy, Hydrometallurgy **64**, 219 (2002).

[5] US Geological Survey, *Mineral Commodity Summaries* (2002).

[6] N. J. Welham, Int. J. Mineral. Proc. **61**, 145 (2001).

[7] C. Bird, Met. Bull. Monthly **5**, 48 (1993).

[8] M. E. Campderros and J. Marchese, J. Memb. Sci. **164**, 205 (2000).

[9] P. Senise and L. Sant'Agostino, Anal. Chim. Acta **22**, 296 (1960).

[10] A. Kononov, S. Kuznetsov, and E. Polakov, J. Alloys Comp. **218**, 173 (1995).

[11] W. A. Serjak, H. Seyeda, C. G. Cymorek, *Tantalum Availability-2000 and Beyond*, Technical Brochure, (H. C. Starck Inc., 2000); Tantalum-Niobium International Study Center Bulletin, **109** (2002).

[12] I. Gaballah, E. Allain, and M. Djona, Metall. Mat. Trans. **28B**, 359 (1997).

[13] C. Liyuan and Z. Haiyun, Metal Powder Report **55**, 30 (2000).

[14] S. Senderoff, G. W. Mellors, and W. J. Reinhardt, J. Electrochem. Soc. **112**, 840 (1965).

[15] T. Balaji, R. Govindaiah, M. K. Sharma, Y. Purushotam, A. Kumar, and T. L. Prakash, in press.

[16] J. Walter, H. Shioyama, and Y. Sawada, Carbon **37**, 41 (1999).

[17] K. Lehovec, J. Less-Common Metals **7**, 397 (1964).

[18] J. V. Grahn, P. E. Hellberg, and E. Olsson, J. Appl. Phys. **84** 1632 (1998).

[19] R. W. Buckman, Jr., AIP Conf. Proc. **271**, 93 (1993).

[20] I. Ayerdi, E. Castaño, A. García–Alonso, and F. J. Gracia, Sensors and Actuators A **46**, 218 (1995).

[21] W. T. Elwell and D. F. Wood, Anal. Chim. Acta **26**, 1 (1962).

[22] S. Mato, G. E. Thompson, P. Skeldon, K. Shimizu, H. Habazaki, and P. Masheder, Corrosion Science **43**, 993 (2001).

[23] K. Mimura, M. Uchikoshi, T. Kekesi, and M. Isshiki, Mat. Sci. Eng. A **334**, 127 (2002).

[24] W. C. Wetzel, J. A. C. Broekaert, and G. M. Hieftje, Spectrochimica Acta B: Atomic Spectroscopy, **57** 1009 (2002).

[25] S. L. Tong and W. W. Harrison, Spectrochimica Acta B **48**, 1237 (1993).

[26] H. Grabatin, V. Schlett, H. Stuke, and H. Weiss, J. Vac. Sci. Technol. A **3**, 2545 (1985).

[27] B. N. Chapman, *Glow Discharge Processes: Sputtering and Plasma Etching*, (Wiley, New York, 1980).

[28] M. J. Albom, Vacuum **13**, 332 (1963).

[29] R. L. Taylor and H. E. Haring, J. Electrochem. Soc. **103**, 611 (1956).

[30] C. Chaneliere, J. L. Autran, B. Balland, and R. A. B. Devine, Mat. Sci. Eng. Rep. **22**, 269 (1998).

[31] E. Atanassova, Microelectronics and Reliability **39**, 1185 (1999).

[32] S. Dueñas, H. Castán, J. Barbolla, R. R. Kola, and P. A. Sullivan, Microelectronics Reliability **40**, 659 (2000).

[33] T. Miyazaki, H. Kim, K. Kokubo, C. Ohtsuki, H. Kato, and T. Nakamura, Biomaterials **23**, 827 (2002).

[34] S. G. Steinemann, Injury **27**, S-C16 (1996).

[35] F. Macionczyk, B. Gerold, and R.Thull, Surf. Coat. Tech. **142**, 1084 (2001).

[36] H. Matsuno, A. Yokoyama, F. Watari, M. Uo, and T. Kawasaki, Biomaterials **22**, 1253 (2001).

[37] M. Niinomi, D. Kuroda, K. Fukunaga, M. Morinaga, Y. Kato, T. Yashiro, and A. Suzuki, Mat. Sci. Eng. A263 193 (1999).

[38] M. J. Ignatius, N. Sawhney, A. Gupta, B. M. Thibadeau, O. R. Monteiro, and I. G. Brown, J. Biomed. Mat. Res. **40**, 264 (1998).

[39] N. E. Promisel, Ed., *The Science and Technology of Selected Refractory Metals*, (Macmillan, New York, 1964).

[40] A. F. Tavassoli, J. Nucl. Mat. **302**, 73 (2002).

[41] J. R. DiStefano, J. Mater. Eng. **11**, 215 (1989).

[42] S. M. Cardonne, P. Kumar, C. A. Michaluk, and H. D. Schwartz, Int. J. Refract. Mat. Hard Met. **13**, 187 (1995).

[43] A. L. Purvis and B. M. Warnes, Surf. Coat. Tech. **133** 23, (2000).

[44] T. Ushikubo and K. Wada, Appl. Catalysis A **124**, 19 (1995).

[45] T. Ushikubo and K. Wada, J. Catalysis **148**, 138 (1994).

[46] T. Ushikubo, Catalysis Today **57**, 331 (2000).

[47] P. Moggi and G. Albanesi, Appl. Catalysis **68**, 285 (1991).

[48] T. Ushikubo and K. Wada, Appl. Catalysis **67**, 25 (1990).

[49] V. N. Kurlov, B. S. Redkin, and V. A. Tatarchenko, Izv. Akad. Nauk SSSR, Ser. Fiz. **52**, 128, (1988).

[50] Y. Furukawa, J. Cryst. Growth **197**, 889 (1999).

[51] D. R. Shearer and L. E. Rowe, Am. J. Roentgenology **148**, 1023 (1987).

[52] N. Rilinger, B. Hoberg, T. Pfeifer, R. Tomczak, J. Haberle, and J. M. Friedrich, Rontgenpraxis **46**, 313 (1993).

[53] L. Tien and C. Chou, D. Tsai, Ceram. Int. **26**, 57 (2000).

[54] C. Lu and C. Tsai, Mat. Sci. Eng. B **55**, 95 (1998).

[55] J. S. Kim, J. -W. Kim, C. I. Cheon, Y. -S. Kim, S. Nahm, and J. D. Byun, J. Euro. Ceram. Soc. **21**, 2599 (2001).

[56] B. Ahn and N. Kim, Mat. Res. Bull. **35**, 1677 (2000).

[57] A. E. Feiring, J. Fluorine Chem. **14**, 7 (1979).

[58] W. Lee, J. Lin, and C. Lee, Mat. Chem. Phys. **68**, 266 (2001).

[59] I. Ayerdi, E. Castaño, A. García–Alonso, and F. J. Gracia, Sensors and Actuators A **42**, 435 (1994).

[60] C. G. Sridhar, R. Chow, and G. Nocerino, Thin Solid Films **140**, 51 (1986).

[61] L. Holland, Ed., *Thin Film Microelectronics*, (Wiley, New York, 1965).

[62] Z. G. Pinsker, *Electron Diffraction*, (Butterworth, London, 1953).

[63] D. H. Kay, *Techniques for Electron Microscopy*, (Blackwell, Oxford, 1965).

[64] G. Binnig, C. F. Quate, and C. Gerber, Phys. Rev. Lett. **56**, 930 (1986).

[65] P. K. Hansma and J. Tersoff, J. Appl. Phys. **61**, 23 (1987).

[66] L. I. Maissel and M. H. Francombe, *An Introduction to Thin Films*, (Gordon and Breach, New York, 1973).

[67] A. Wagendristel and Y. Wang, *An Introduction to Physics and Technology of Thin Films*, (World Scientific, Singapore, 1994).

[68] K. Chen, M. Nielsen, E. J. Rymaszewski, and T. M. Lu, Mat. Chem. Phys. **49**, 42 (1997).

[69] K. Ohta, K. Yamada, R. Shimizu, and Y. Tarui, IEEE Trans. Elect. Dev. **29**, 368 (1992).

[70] T. Kato, T. Itoh, M. Taguchi, T. Nakamura, and H. Ishikawa, Symp. VLSI Technol., Digest of Technical Papers, IEEE, (1983).

[71] G. S. Oehrlein, J. Appl. Phys. **59**, 1587 (1986).

[72] C. Hasimoto, H. Oikawa, and N. Honma, IEEE Trans. Elect. Dev. **36**, 14 (1989).

[73] Y. Numasawa, S. Kamiyama, M. Zenke, and M. Sakamoto, IEDM Tech. Dig. **89**, 43 (1989).

[74] S. Zaima, T. Furuta, Y. Yasuda, and M. Iida, J. Electrochem. Soc. **137**, 1297 (1990).

[75] S. Zaima, T. Furuta, Y. Koide, and Y. Yasuda, J. Electrochem. Soc. **137**, 2876 (1990).

[76] H. Shinriki, IEEE Trans. Elect. Dev. **38**, 55 (1991).

[77] K. A. McKinley and N.P. Sandler, Thin Solid Films **290**, 440 (1996).

[78] M. T. Kief and W. F. Egelhoff, Jr., Phys. Rev. B **47**, 10785 (1994).

[79] L. M. Falicov, D. T. Pierce, S. D. Bader, R. Gronsky, K. B. Hathaway, H. J. Hopster, D. N. Lambeth, S. S. P. Parkin, G. Prinz, M. Salamon, I. K. Schuller, and R. H. Victora, J. Mater.Res. **5**, 1299 (1990).

[80] S. C. Erwin, S. H. Lee, and M. Scheffler, Phys. Rev. B **65**, 205422 (2002).

[81] W. D. Westwood and N. Waterhouse, J. App. Phys. **42**, 2946 (1971).

[82] S. Sato, T. Inoue, and S. Sasaki, Thin Solid Films **86,** 21 (1981).

[83] W. D. Westwood, N. Waterhouse, and P. S. Wilcox, *Tantalum Thin Films*, (Academic Press, London, 1975).

[84] H. D. Seghezzi and E. Gebhardt, Bergakademie **10**, 75 (1958).

[85] C. A. Neugebauer and M. B. Webb, J. Appl. Phys. **33**, 74 (1962).

[86] R. B. Marcus, J. Appl. Phys. **37**, 3121 (1966).

[87] N. J. Maskalick and C. W. Lewis, Trans 8th Natl. Vac. Symp. (1961).

[88] E. Fromm, G. Hofer, Zeits. Metallkunde **62**, 223 (1971).

[89] J. I. Budnick, Phys. Rev. **119**, 1578 (1960).

[90] F. Schrey, R. D. Mathis, and R. T. Payne, Thin Solid Films **5**, 29 (1970).

[91] D. W. Face and D. E. Prober, J. Vac. Sci. Technol. A **5**, 3408 (1987).

[92] J. H. Hauser and H. C. Theurer, Phys. Rev. **134**, A198 (1964).

[93] M. H. Read and C. Altman, Appl. Phys. Lett. **7**, 51 (1965).

[94] D. Gerstenberg and P. M. Hall, J. Electrochem. Soc. **111**, 936 (1964).

[95] N. Schwartz, W. A. Reed, P. Polash, and M. H. Read, Thin Solid Films **14**, 333 (1972).

[96] R. B. Marcus and S. Quigley, Thin Solid Films **2**, 467 (1968).

[97] P. N. Denbigh and R. B. Marcus, J. Appl. Phys. **37**, 4325 (1966).

[98] K. L. Chopra, M. R. Randlett, and R. H. Duff, Phil. Mag. **16**, 261 (1967).

[99] D. B. Dove, J. Appl. Phys. **35**, 2785 (1964).

[100] J. E. O'Neal, Thin Films **2**, 119 (1972).

[101] D. I. Lainer and V. A. Kholmyanskii, Sov. Phys. Crystall. **12**, 913 (1968).

[102] V. V. Klechkovskaya and V. I. Khitrova, Sov. Phys. Crystall. **13**, 428 (1968).

[103] D. Gerstenberg and C. J. Calbick, J. Appl. Phys. **35**, 402 (1964).

[104] V. I. Smirnova and B. F. Ormont, Dokl. Akad. Nauk. S. S. R. **96**, 557 (1954).

[105] M. S. Borushko, B. T. Boiko, P. A. Pantsheka, and S. W. Marunko, Wissen. Zeitsch. der Tech. Hochsch. Otto v. Guericke (1970).

[106] P. T. Moseley and C. J. Seabrook, Acta Cryst. B**29**, 1170 (1973).

[107] V. P. Belevskiy, M. V. Belous, V. G. Permyakov, and V. M. Yashnik, Fiz. Metal. Metalloved. **33** 564 (1972).

[108] A. Schauer and M. Roschy, Thin Solid Films **12**, 313 (1972).

[109] Y. Chouan and D. Collobert, J. App. Phys. **48**, 2274 (1977).

[110] L. G. Feinstein and R. D. Huttemann, Thin Solid Films **16**, 129 (1973).

[111] K. Hieber and N. M. Mayer, Thin Solid Films **90**, 43 (1982).

[112] S. Sato, Thin Solid Films **94**, 321 (1982).

[113] K. Hieber and E. Lautenbacher, Thin Solid Films **66**, 191 (1980).

[114] G. Das, Thin Solid Films **12**, 305 (1972).

[115] P. N. Baker, Thin Solid Films **14**, 3 (1972).

[116] R. D. Burbank, Appl. Cryst. **6**, 217 (1973).

[117] M. H. Read and D. H. Hensler, Thin Solid Films **10**, 123 (1972).

[118] P. Catania, R. A. Roy, and J. J. Cuomo, J. Appl. Phys. **74**, 1008 (1993).

[119] A. C. Lawson and C. E. Olsen, Acta Cryst. **B44,** 89 (1988).

[120] K. Kondo, M. Nakabayashi, K. Kawakami, T. Chijimatsu, M. Nakaishi, M. Yamada, M. Yamabe, and K. Sugishima, J. Vac. Sci. Technol. **A11**, 3067 (1993).

[121] J. Donohue and M. Einspahr, Acta Cryst. **B27,** 1740 (1971).

[122] J. Donohue, *The Structures of the Elements*, (Wiley, New York, 1974).

[123] C. W. Tucker and P. Senio, Acta Cryst. **6,** 753 (1953).

[124] H. M. Rietveld, Acta Cryst **2,** 65 (1969).

[125] R. B. Von Dreele, J. D. Jorgensen, and C. G. Windsor, Acta Cryst. **15,** 581 (1982).

[126] A. Arakcheeva, G. Chapuis, and V. Grinevitch, Acta. Cryst. B **58,** 1 (2002).

[127] P. Rao and G. Thomas, Scripta Metallurgica **4**, 243 (1970).

[128] S. Steeb and J. Renner, J. Less Common Metals **10**, 246 (1966).

[129] R. H. Geiss and K. R. Lawless, Scripta Metallurgica **4**, 245 (1970).

[130] H. G. Schneider and H. D. Langer, Wissen. Z. Ilmenau **14**, 165 (1968).

[131] W. B. Pearson, *A Handbook of Lattice Spacings and Structures of Metals and Alloys*, (Pergamon, Oxford, 1964).

[132] G. L. Miller, *Metallurgy of the Rarer Metals, No. 6. Tantalum and Niobium,* (Butterworths, London, 1962).

[133] D. D. Vvedensky, A. Zangwill, C. N. Luse, and M. R. Wilby, Phys. Rev. B **48**, 852, (1993).

[134] D. D Vvedensky, S. Clarke, K. J. Hugill, A. K. Myers-Beaghton, and M.R. Wilby *Kinetics of Ordering and Growth at Surfaces*, M.G. Lagally, Ed., (Plenum Press, New York, 1990).

[135] H. Brune, Surf. Sci. Rep. **31**, 121, (1998).

[136] J. P. Hirth and K. L. Moazed in, *Physics of Thin Films* **4**, G. Hass, R. E. Thun, Eds., (Academic Press, New York, 1967).

[137] B. Lewis and D. S. Campbell, J. Vac. Sci. Technol. **4**, 209 (1967).

[138] T. N. Rhodin in, *The Use of Thin Films in Physical Investigations*, J. C. Anderson, Ed., (Academic Press, London, 1966).

[139] D. D. Vvedensky, Comp. Mat. Sci. **6**,182 (1996).

[140] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos Rev. Mod. Phys. **64**, 1045, (1992).

[141] R. Wang and K. Fichthorn, Thin Solid Films **272**, 223 (1996).

[142] S. Rousset, S. Chiang, D. E. Fowler, and D. D. Chambliss, Phys. Rev. Lett. **69**, 3200 (1992).

[143] G. Vidali, Surf. Rev. Lett. **4**, 709 (1997).

[144] T. Shitara, D. D. Vvedensky, and M. R. Wilby, Phys. Rev. B **46**, 6815 (1992).

[145] E. Kaxiras, O. L. Alerhand, J. D. Joannopoulos, and G. W. Turner, Phys. Rev. Lett. **62**, 2484 (1989).

[146] C. V. Dharmadhikari, A. O. Ali, N. Suresh, D. M. Phase, S. M. Chaudhari, A. Gupta, and B. A. Dasannacharya, Mat. Sci. Eng. **B75**, 29 (2000).

[147] G. Vidali and H. Zeng, App. Surf. Sci. **92**, 11 (1996).

[148] H. Zeng and G. Vidali, Phys. Rev. Lett **75**, 582 (1994).

[149] G. Rosenfeld, R. Servaty, C. Teichert, B. Poelsema, and G. Comsa, Phys. Rev. Lett. **71**, 895, (1993).

[150] J. L. Menendez, P. Caro, and A. Cebollada, J. Cryst. Growth **192**, 164 (1998).

[151] G. Ehrlich and F. G. Hudda, J. Chem. Phys. **44**, 1039 (1966).

[152] R. L. Schwoebel and E. J. Shipsey, J. Appl. Phys. 37, 3682 (1966).

[153] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Appl. Phys. Lett. **40**, 178 (1982).

[154] Y. Tian, F. Jona, and P. M. Marcus, Phys. Rev. B **58**, 14 051 (1998).

[155] Y. Tian, F. Jona, and P. M. Marcus, Phys. Rev. B **59**, 12 286 (1999).

[156] J. Xu, A. J. Freeman, and T. Jarlborg, Phys. Rev. B **29**, 1250 (1984).

[157] M. Maoujoud, P. Kons, M. Jardinier-Offergeld, and F. Bouillon, Thin Solid Films **238**, 62 (1994).

[158] L. A. Clevenger, A. Mutscheller, J. M. E. Harper, C. Cabral, Jr., and K. Barmak, J. Appl. Phys. **72**, 4918 (1992).

[159] I. A. Weerasekera, S. I. Shah, D. V. Baxter, and K. M. Unruh, Appl. Phys. Lett. **64**, 3231 (1994).

[160] M. J. O'Keefe and J. T. Grant, J. Appl. Phys. **79**, 9134 (1996).

[161] S. Basavaiah and S. R. Pollack, J. App. Phys. **39**, 5548 (1968).

[162] W. K. Yeh, M. C. Chen, P. J. Wang, L. M. Liu, and M. S. Lin, Mater. Chem. Phys. **45**, 284 (1996).

[163] Y. G. Shen, Y. W. Mai, Q. C. Zhang, D. R. McKenzie, W. D. McFall, and W. E. McBride, J. App. Phys. **87**, 177 (2000).

[164] P. Petroff, T. T. Sheng, A. K. Sinha, G. A. Rozgonyi, and F. B. Alexander, J. App. Phys. **44**, 2545 (1973).

[165] W. R. Morcom, W. L. Worrell, H. G. Sell, and H. I. Kaplan, Metall. Trans. **5**, 155 (1974).

[166] G. Mannella and J. O. Hougen, J. Phys. Chem. **60**, 1148 (1956).

[167] M. J. O'Keefe, J. T. Grant, and J. S. Solomon, J. Elec. Mater. **24**, 961 (1995).

[168] E. C. Bain, Trans. AIME **70**, 25 (1924).

[169] P. M. Marcus and F. Jona, J. Phys. Condens. Matter **9**, 6241 (1997).

[170] R. Prasad, R. E. Somekh, and A. L. Greer, Adv-Phys. Metall., Sel. Pap. Proc. Int. Conf., 1994, p150.

[171] D. J. Thoma and J. H. Perepezko, Mater. Sci. Forum **179 - 181**, 769 (1995).

[172] D. J. Thoma, J. H. Perepezko, D. H. Plantz, and R. B. Schwarz, Mater. Sci. Eng.A, **A179 - A180**, 176 (1994).

[173] S. Cottenier, *Density Functional Theory and the Family of (L)APW Methods: A Step-by-Step Introduction*, (Instituut voor Kern Stralingsfysika, K. U. Leuven, Belgium), 2002, ISBN 90-807215-1-4 (to be found at http://www.wien2k.at/reg_user/textbooks).

[174] D. J. Singh, *Planewaves, Pseudopotentials and the LAPW Method*, (Kluwer Academic Publishers, Massachusetts, 1994).

[175] J. P. Perdew and W. Yue, Phys. Rev. B **33**, 8800 (1986).

[176] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[177] C. Stampfl, W. Mannstadt, R. Asahi, and A. J. Freeman, Phys. Rev. B **63**, 155106 (2001).

[178] N. W. Ashcroft and N. D. Mermin, *Solid State Physics,* (Holt, Rinehart and Winston, Philadelphia, 1976).

[179] R. Yu, D. Singh, and H. Krakauer, Phys. Rev. B **43**, 6411 (1991).

[180] B. Kohler, S. Wilke, M. Scheffler, R. Kouba, and C. Ambrosch-Draxl, Comput. Phys. Commun. **94**, 31 (1996).

[181] C. G. Broyden, J. E. Dennis, and J. J. Moré, J. Inst. Maths. Appl. **12**, 223 (1973).

[182] K. W. Brodlie, in: *The State of the Art in Numerical Analysis*, ed. D.A.H. Jacobs (Academic Press, London, 1977).

[183] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, (Prentice-Hall, Englewoods Cliffs, 1983).

[184] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, (Cambridge University Press, Cambridge, 1992).

[185] P. Novak, *Calculation of Spin-Orbit Coupling* (unpublished).

[186] P. Blaha, K. Schwarz, and J. Luitz, WIEN97, *A Full Potential Linearized Augmented Plane Wave Package for Calculating Crystal Properties*, (Karlheinz Schwarz, Techn. Univ. Wien, Vienna 1999). ISBN 3-9501031-0-4

[187] J. F. Nye, *Physical Properties of Crystals, their Representation by Tensors and Matrices*, (Clarendon Press, Oxford, 1957).

[188] T. Charpin, *A Package for Calculating Elastic Properties of Cubic Phases using WIEN*, (unpublished).

[189] F. D. Murnaghan, *Finite Deformation of an Elastic Solid*, (John Wiley and Sons, New York, 1951).

[190] F. Birch, J. Phys. Chem. Solids **38**, 175 (1977).

[191] P. M. Marcus and P. Alippi, Phys. Rev. B **57**, 1971 (1998).

[192] P. Alippi, P. M. Marcus, and M. Scheffler, Phys. Rev. Lett. 78, 3892 (1997).

[193] P. E. Blöchl, O. Jepsen, and O. K. Anderson, Phys. Rev. B **49**, 16223 (1994).

[194] R. W. G. Wyckoff, *Crystal Structures*, (John Wiley and Sons, New York, 1963); G.V. Samsonov (Ed.) in *Handbook of the physicochemical properties of the elements*, (IFI-Plenum, New York, 1968).

[195] J. M. Wills, O. Eriksson, P. Söderlind, and A. M. Boring, Phys. Rev. Lett. **68**, 2802 (1992).

[196] G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: a Handbook*, (MIT Press, Cambridge, 1971).

[197] J. H. Xu and A. J. Freeman, Phys. Rev. B **40**, 11927 (1989).

[198] W. G. Burgers, Physica **1**, 561 (1935).

[199] R. M. Wentzcovitch and H. Krakauer, Phys. Rev. B **42**, 4563 (1990).

[200] V. P. Dmitriev, Yu. M. Gufan, and P. Tolédano, Phys. Rev. B **44**, 7248 (1991).

[201] L. M. Hsiung and D. H. Lassila, Scripta Materialia **38**, 1371 (1998).

[202] L. M. Hsiung and D. H. Lassila, Scripta Materialia **39**, 603 (1998).

[203] G. B. Grad, P. Blaha, J. Luitz, K. Schwarz, A. F. Guillermet, and S. J. Sferco, Phys. Rev. B **62**, 12743 (2000).

[204] R. Ahuja, P. Söderlind, J. Trygg, J. Melsen, and J. M. Wills, Phys. Rev. B **50**, 14690 (1994).

[205] P. Söderlind and J. A. Moriarty, Phys. Rev. B **57**, 10340 (1998).

[206] G. W. Fernando, R. E. Watson, M. Weinert, Y. J. Wang, and J. W. Davenport, Phys. Rev. B **41**, 11813 (1990).

[207] H. Brune, K. Bromann, H. Röder, K. Kern, J. Jacobsen, P. Stoltze, K. Jacobsen, and J. Nørskov, Phys. Rev. B **52**, R14 380 (1995).

[208] C. Ratsch, A. P. Seitsonen, and M. Scheffler, Phys. Rev. B **55**, 6750 (1997).

[209] C. Ratsch and M. Scheffler, Phys. Rev. B **58**, 13163 (1998).

[210] F. Bloch, Z. Phys. **52**, 555 (1928).

[211] J. M. Ziman, Proc. Roy. Soc (London) **A252**, 63 (1959).

[212] J. M. Ziman, *Electrons and Phonons: the Theory of Transport Phenomena in Solids*, (Oxford University Press, New York, 2001).

[213] P. B. Allen, Phys. Rev. B **3**, 305 (1971).

[214] J. Yamashita and S. Asano, Prog. Theor. Phys. **51**, 317 (1974).

[215] B. N. Harmon and S. K. Sinha, Phys. Rev. B **16**, 3919 (1977).

[216] W. H. Butler, H. G. Smith, and N. Wakabayashi, Phys. Rev. Lett. **39**, 1004 (1977).

[217] P. B. Allen, Phys. Rev. B **17**, 3725 (1978).

[218] C. Kittel, *Introduction to Solid State Physics*, (John Wiley and Sons, New York, 1996).

[219] N. Wiser, Contemp. Phys. **25**, 211 (1984).

[220] J. S. Dugdale and Z. S. Basinski, Phys. Rev. **157**, 552 (1967).

[221] F. Y. Fradin, Phys. Rev. Lett. **33**, 158 (1974).

[222] F. J. Pinski, P. B. Allen, and W. H. Butler, Phys. Rev. B **23**, 5080 (1981).

[223] M. J. Laubitz, C. R. Leavens, and R. Taylor, Phys. Rev. Lett. **39**, 225 (1977).

[224] N. Tralshawala, J. F. Zasadzinski, L. Coffey, W. Gai, M. Romalis, Q. Huang, R. Vaglio, and K. E. Gray, Phys. Rev. B **51**, 3812 (1995).

[225] W. L. McMillan, Phys. Rev. **167**, 331 (1968).

[226] P. B. Allen, T. P. Beaulac, F. S. Khan, W. H. Butler, F. J. Pinski, and J. C. Swihart, Phys. Rev. B **34**, 4331 (1986).

[227] E. L. Wolf, *Principles of Electronic Tunneling Spectroscopy*, (Oxford University Press, New York, 1985).

[228] E. N. Adams, Phys. Rev. **98**, 947 (1955).

[229] M. H. Cohen, Phil. Mag. **3**, 762 (1958).

[230] J. Neve, B.Sundqvist, and O. Rapp, Phys. Rev. B **28**, 629 (1983).

[231] C. M. Soukoulis and D. A. Papaconstantopoulos, Phys. Rev. B **26**, 3673 (1982).

[232] B. Chakraborty, W. E. Pickett, and P. B. Allen, Phys. Rev. B **14**, 3227 (1976).

[233] L. F. Mattheiss, Phys. Rev. B **1**, 373 (1970).

[234] P. D. Desai, T. K. Chu, H. M. James, and C. Y. Ho, J. Phys. Chem. Ref. Data **13**,1069 (1984).

[235] T. P. Beaulac, P. B. Allen, and F. J. Pinski, Phys. Rev. B **26**, 1549 (1982).

[236] V. H. Crespi and M. L. Cohen, Solid State Commun. **81**, 187 (1992).

[237] E. L. Wolf, T. P. Chen, and D. M. Burnell. Phys. Rev. B **31**, 6096 (1985).

[238] R. A. Roy, J. J. Cuomo, and D. S. Yee, J. Vac. Sci. Technol. **A 6**, 1621 (1988).

[239] J. E. Nestell, Jr., K. J. Scoles, R. W. Christy, J. Appl. Phys. **53**, 8993 (1982).

[240] M. Calandra and O. Gunnarsson, Phys. Rev. B **66**, 205105 (2002).

[241] S.Y. Savrasov, D. Y. Savrasov, Phys. Rev. B **54**, 16487 (1996).

[242] M. Stavrev, D. Fischer, C. Wenzel, K. Drescher, N. Mattern, Thin Solid Films **307**, 79 (1997).

[243] D. Mills, J. Can. Cer. Soc. **35**, 48 (1965).

[244] A. Yohannan, *Characterization of α and β Phases of Tantalum Coatings*, Masters Thesis, New Jersey Institute of Technology (2001).

[245] A. Schauer, W. Peters, Thin Solid Films **27**, 95 (1975).

[246] K-W. Kwon, C. Ryu, R. Sinclair, Appl. Phys. Lett. **71**, 3069 (1997).

[247] A. Kokalj, J. Mol. Graphics Modeling **17**, 176 (1999).

[248] J. Sosniak, W. J. Polito, G. A. Rozgonyi, J. Appl. Phys. **38**, 3041 (1967).

[249] F. Vratny, B. H. Vromen, A. J. Harndza-Horinxma, Electrochemical Tech. **5** 283 (1967).

[250] E. Krikorian, R. J. Sneed, J. Appl. Phys. **37**, 3674 (1966).

[251] J. Sosniak, J. Vac. Sci. Technol. **4**, 87 (1967).

[252] H. C. Cook, J. Vac. Sci. Technol. **4**, 80 (1967).

[253] J. Sosniak, J. Appl. Phys. **39**, 4157 (1968).

[254] D. Gerstenberg, E. H. Mayer, Proc. Elec. Comp. Conf., 1962, p57.

[255] C. Altman, *Trans. 9th Nat. Vac. Symp*, (Macmillan, New York, 1962).

[256] C. J. Calbick, *6th Int. Congr. Electron Microscopy*, (Maruzen, Tokyo, 1966).