Band Engineering of Advanced Materials for Semiconductor Devices

Jiaqi Chen

Christ's College

Electrical Engineering Division, Department of Engineering

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

March 2023

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit of 65,000 words for the Engineering Degree Committee.

Jiaqi Chen

March 2023

Band Engineering of Advanced Materials for Semiconductor Devices

Jiaqi Chen

ABSTRACT

As the downscaling of metal–oxide–semiconductor field-effect transistors (MOSFETs) continues, the short-channel effects (SCEs) and contact resistance severely degrade device performance. It is crucial to understand the physics at various interfacial regions of MOSFETs to provide guidance for overcoming these limitations. In this thesis, the representative materials and their contacts employed in the development of MOSFETs are studied using density functional calculations, with an emphasis on understanding the electronic behaviours of metal–semiconductor junctions and heterojunctions.

The thesis studies the structural, electronic, and optical properties of nine polymorphs of SiO_2, employing both the traditional generalised gradient approximation (GGA) and the state-of-the-art screened exchange (sX) functional. Calculations using the sX functional accurately reproduce the experimental band gap values, whereas GGA is more effective in describing the optical properties.

The advanced sX method and the more efficient GGA + U scheme are applied to several important oxides (ZnO, CdO, SrO, and MgO) to address the underestimated band gaps of oxides by the traditional GGA functional. The GGA + U scheme is further applied to calculate the Schottky barrier heights (SBHs) at various metal–oxide interfaces. The metal-induced gap states (MIGS) model is demonstrated to be a reliable simplified approach for predicting the pinning effect.

A similar investigation involving high-κ CaF_2 is carried out, which confirms the accuracy of the sX method in characterising wide band gap materials. Moreover, the computed electronic properties of Si–CaF_2 and metal–CaF_2 interfaces obtained using the GGA + U scheme are consistent with the MIGS predictions.

Studies are also conducted on p-type and ambipolar monolayer WSe_2 contacting with various metals. Through a rational design approach, weakly pinned, low-resistance metal–WSe_2 contacts are achieved, offering potential applications in 2D semiconductor devices.

Dedication

to my loving parents, cherished friends, and inspiring mentors

for their love, support, and wisdom

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Prof. John Robertson, for his unwavering support, guidance, and encouragement throughout my research journey. His extensive knowledge, expertise, and dedication have been invaluable in shaping my understanding and growth in this field. As an exceptional mentor and a great source of inspiration, his passion for research and commitment to excellence have motivated me to challenge myself and strive for the best in my work.

I would like to extend my gratitude to Dr. Zhaofu Zhang and Prof. Yuzheng Guo for their invaluable guidance in teaching me computational simulation methods and essential research skills. Additionally, I would like to thank the other members of our group, Dr. Haichang Lu, Dr. Guandong Bai, and Xuewei Zhang, for their support and camaraderie during my time in Cambridge.

I also wish to acknowledge my friends Yinglun Hong, Tianhui Jiang, Quanyao Lin, Yi Lin, Yuling Xiong, Kangning Zhang, and Zhihan Zhang for their unwavering support and encouragement. A special mention goes to my boyfriend, Derek Lu, whose love, patience, and understanding have been a constant source of strength, making even the most challenging moments bearable.

Lastly, I must express my deepest appreciation to my parents, whose unconditional love and care have been the pillars that supported me throughout my life. I love them dearly, and I would not have made it this far without their unwavering belief in me. Their sacrifices, encouragement, and guidance have shaped me into the person I am today, and I am forever grateful for their presence in my life.

LIST OF PUBLICATIONS

[1] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Revisiting the Electronic and Optical Properties of SiO_2 Polymorphs by Hybrid Functional Calculations**. *Journal of Applied Physics* **133**, 044101 (2023) (Selected as Editor's Pick).

[2] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Metal Contacts with Moire Interfaces on WSe_2 for Ambipolar Applications**. *Applied Physics Letters* **121**, 051602 (2022)

[3] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Electronic Properties of CaF_2 Bulk and Interfaces**. *Journal of Applied Physics* **131**, 215302 (2022)

[4] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Schottky Barrier Heights of Defect-Free Metal/ZnO, CdO, MgO, and SrO Interfaces**. *Journal of Applied Physics* **129**, 175304 (2021)

[5] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Schottky Barrier Height at Metal/ZnO Interface: A First-principles Study**. *Microelectronic Engineering* **216**,111056 (2019)

[6] Zhaofu Zhang, Jiaqi Chen, Yuzheng Guo, and John Robertson. **Band Alignment Calculation of Dielectric Films on VO_2**. *Microelectronic Engineering* **216**,111057 (2019)

Conference Presentations

[1] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Ambipolar Contacts at WSe_2/Metal via Moire Interfaces**, *53rd IEEE Semiconductor Interface Specialists Conference (SISC)*, San Diego, USA, 2022 (Oral Presentation) (Winner of the 2022 SISC Ed Nicollian Award for Best Student Paper).

[2] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Electronic Structure of CaF_2**, *European Materials Research Society (E-MRS) Spring Meeting*, Virtual, 2022 (Oral Presentation)

[3] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **CaF_2 as High-κ Dielectric for Si Devices**, *8th Joint International EuroSOI Workshop and International Conference on Ultimate Integration on Silicon (EuroSOI-ULIS)*, Udine, Italy, 2022 (Poster Presentation)

[4] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Schottky Barrier Heights and Tunnelling at Metal/Oxide Interfaces**, *52nd IEEE Semiconductor Interface Specialists Conference (SISC)*, Virtual, 2021 (Poster Presentation)

[5] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Band Alignment Models of Semiconductor/ZnO Interfaces**, *50th IEEE Semiconductor Interface Specialists Conference (SISC)*, San Diego, USA, 2019 (Poster Presentation)

[6] Jiaqi Chen, Zhaofu Zhang, Yuzheng Guo, and John Robertson. **Schottky Barrier Height at Metal–ZnO (110) Non-Polar Interface**, *21st Conference on Insulating Films on Semiconductors (INFOS)*, Cambridge, UK, 2019 (Poster Presentation)

TABLE OF CONTENTS

List of Abbreviations..X

Chapter 1 Introduction..1

1.1 Overview of Semiconductor Devices..1

1.2 MOSFET and Scaling..3

1.2.1 Fundamentals of MOSFET..3

1.2.2 Scaling Challenges in MOSFET Technology..4

1.2.3 Advances in MOSFET Scaling..5

1.3 Metal–Semiconductor Contacts..8

1.3.1 Ohmic Contacts and Schottky Contacts..8

1.3.2 Schottky Barrier Height..10

1.3.3 Fermi Level Pinning: the MIGS Model..11

1.4 Heterojunctions..14

1.4.1 Heterojunction Band Alignment..14

1.4.2 Band Offset: Electron Affinity Rule and CNL Model..17

1.5 Motivation and Thesis Outline..19

Chapter 2 Simulation Theory..21

2.1 Elementary Quantum Mechanics..21

2.1.1 The Schrödinger Equation..21

2.1.2 The Hartree–Fock Approximation..23

2.2 Density Functional Theory..25

2.2.1 The Electron Density..25

2.2.2 The Hohenberg–Kohn Theorems..26

2.2.3 The Kohn–Sham Approach..27

2.3 Exchange-correlation Functionals..30

2.3.1 Local Density Approximation..30

2.3.2 Generalised Gradient Approximations..31

2.3.3 Orbital-Dependent Functionals: LDA + U..31

2.3.4 Hybrid Functionals..33

Chapter 3 Revisiting the Electronic and Optical Properties of SiO_2 Polymorphs by Hybrid Functional Calculations **37**

3.1 Introduction 37

3.2 Calculation Methods 40

3.3 Results 41

3.3.1 Structural Properties of SiO_2 41

3.3.2 Electronic Properties of SiO_2 46

3.3.3 Optical Properties of SiO_2 50

3.4 Discussion 54

3.5 Conclusions 58

Chapter 4: Schottky Barrier Heights of Defect-Free Metal–ZnO, CdO, MgO and SrO Interfaces **59**

4.1 Introduction 59

4.2 Calculation Methods 62

4.3 Results and Discussion 65

4.3.1 Band Structures of Oxides 65

4.3.2 Electronic Properties at the Metal–Oxide Interfaces 71

4.4 Conclusions 80

Chapter 5: Electronic Properties of CaF_2 Bulk and Interfaces **82**

5.1 Introduction 82

5.2 Calculation Methods 85

5.3 Results and Discussion 85

5.3.1 Band Structures of CaF_2 85

5.3.2 Electronic Properties at the Si–CaF_2 Interfaces 90

5.3.3 Electronic Properties at the Metal–CaF_2 Interfaces 95

5.4 Conclusions 97

Chapter 6: Metal Contacts with Moiré Interfaces on WSe_2 for Ambipolar Applications **98**

6.1 Introduction 98

6.2 Calculation Methods 102

6.3 Results and Discussion 102

6.3.1 Interface Modelling 102

6.3.2 Interface Stability and Bonding 107

6.3.3 Schottky Barrier Heights and Interfacial States 112

6.4 Conclusions 121

Chapter 7: Conclusions and Outlook **122**

References **124**

LIST OF ABBREVIATIONS

ACBN0	Agapito–Curtarolo–Buongiorno Nardelli
B88	Becke88
BCC	body-centred cubic
BJT	bipolar junction transistor
BTBT	band-to-band tunnelling
CB	conduction band
CBM	conduction band minimum
CBO	conduction band offset
CMOS	complementary metal–oxide–semiconductor
CNL	charge neutrality level
CVD	chemical vapour deposition
DFT	density functional theory
DIBL	drain-induced barrier lowering
EA	electron affinity
EOT	equivalent oxide thickness
f. u.	formula unit
FCC	face-centred cubic
FET	field-effect transistor
FinFET	fin field-effect transistor
FLAPW	full-potential linearised augmented plane wave
FLP	Fermi level pinning
GAA	gate-all-around
GAAFET	gate-all-around field-effect transistor
GGA	generalised gradient approximation
h-BN	hexagonal boron nitride
HCP	hexagonal closest packed
HEG	homogeneous electron gas
HF	Hartree–Fock
HFET	heterostructure field-effect transistor
HSE	Heyd–Scuseria–Ernzerhof
IC	integrated circuit
IP	ionisation potential
LDA	local density approximation

LYP	Lee–Yang–Parr
MBE	molecular beam epitaxy
MCA	magnetocrystalline anisotropy
MD	molecular dynamics
MESFET	metal–semiconductor field-effect transistor
MIGS	metal-induced gap states
MIS	metal–insulator–semiconductor
MOS	metal–oxide–semiconductor
MOSFET	metal–oxide–semiconductor field-effect transistor
MTJ	magnetic tunnel junction
OLCAO	orthogonalised linear combinations of atomic orbitals
PBE	Perdew–Burke–Ernzerhof
PDMS	polydimethylsiloxane
PDOS	partial density of states
PMA	perpendicular magnetic anisotropy
poly-Si	polycrystalline silicon
RHEED	reflection high-energy electron diffraction
SBH	Schottky barrier height
SCE	short-channel effect
SCF	self-consistent field
SOC	spin–orbit coupling
SPD	Schottky photodiode
SSDT	Schottky barrier source/drain transistor
STT-MRAM	spin-transfer torque magnetic random-access memory
sX	screened exchange
TB	tight binding
TDOS	total density of states
TEM	transmission electron microscopy
TFET	tunnel field-effect transistor
TMD	transition metal dichalcogenide
VB	valence band
VBM	valence band maximum
VBO	valence band offset
vdW	van der Waals
XC	exchange-correlation

CHAPTER 1 INTRODUCTION

1.1 Overview of Semiconductor Devices

Since the discovery of the first point-contact transistor in 1947 by Bardeen and Bratain [1] at the Bell Laboratories, there have been significant advances in the semiconductor device field. The development of integrated circuits (ICs) in the late 1950s [2] allowed multiple transistors and other electronic components to be fabricated on a single piece of semiconductor material, which greatly increases the complexity and functionality of electronic devices.

Nowadays, semiconductor devices find their applications in numerous products, spanning fields like communication, computing, clean energy, and countless other areas. The significant advancements in semiconductor technology have resulted in high-performance devices that revolutionise daily life and work environments. Currently, there are 74 major semiconductor devices and over 200 device variations associated with them [3]. All these devices can be constructed with four basic building blocks: metal–semiconductor junction, p–n junction, heterojunction, and metal–oxide–semiconductor (MOS) structure (Fig. 1.1).

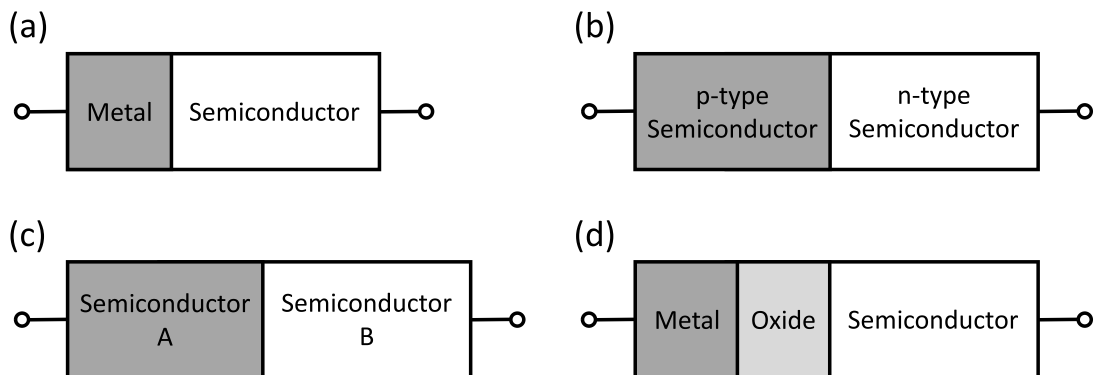


Figure 1.1 Basic building blocks of semiconductor devices. (a) Metal–semiconductor junction, (b) p–n junction, (c) heterojunction, and (d) MOS structure.

The metal–semiconductor junction [Fig. 1.1(a)] is the close contact between a metal and a semiconductor. The metal–semiconductor interface can either be rectifying and forms a Schottky barrier, or non-rectifying as an Ohmic contact. Typically, an Ohmic contact is preferred to facilitate the easy conduction between the active region of a transistor and the external circuit. However, in certain cases, a Schottky barrier can

be beneficial, a notable example is its application as the gate for metal–semiconductor field-effect transistors (MESFETs).

The p–n junction [Fig. 1.1(b)] is formed between a p-type semiconductor (containing an excess of holes) and an n-type semiconductor (containing an excess of electrons). The junction between the p-type and n-type semiconductors can become depleted of charge carriers and lead to a non-conductive region. The p–n junction is a crucial component in various important semiconductor devices. For instance, combining two p–n junctions in either an n–p–n or p–n–p configuration produces a bipolar junction transistor (BJT), which is capable of amplification or switching.

The heterojunction [Fig. 1.1(c)] is the interface formed between two dissimilar semiconductors with different band gap energies. The presence of a potential barrier at the interface of a heterojunction can be utilised to control the properties of electronic devices. Depending on the alignment of the energy bands, heterojunctions are utilised in the fabrication of various semiconductor devices, such as heterostructure field-effect transistors (HFETs), which hold promise for high-frequency applications [4].

The MOS structure [Fig. 1.1(d)] is a sandwich structure that consists of a thin layer of oxide (usually SiO_2) in the middle between a metal and a semiconductor layer (usually Si). When a voltage is applied to the metal, an electric field is created in the oxide layer, which then controls the flow of charge carriers (electrons or holes) in the semiconductor. The MOS structure is utilised as the gate of the metal–oxide–semiconductor field-effect transistor (MOSFET), which is one of the most important transistors used in advanced ICs.

1.2 MOSFET and Scaling

1.2.1 Fundamentals of MOSFET

MOSFET is the most widely used device in both memory circuits and microprocessors, offering several advantages over other device types, including well-established fabrication technologies, excellent scaling characteristics, and the ability to form complementary metal–oxide–semiconductor (CMOS) circuits. Fig. 1.2 displays a 3D representation of an n-channel MOSFET, with some of the basic device parameters labelled. The n-type MOSFET is a four-terminal device constructed using a p-type semiconductor substrate, with two heavily n-doped regions that serve as the source and the drain. The gate, made of either metal or poly-crystalline material, covers the area between the source and the drain and is separated from the semiconductor by the gate oxide.

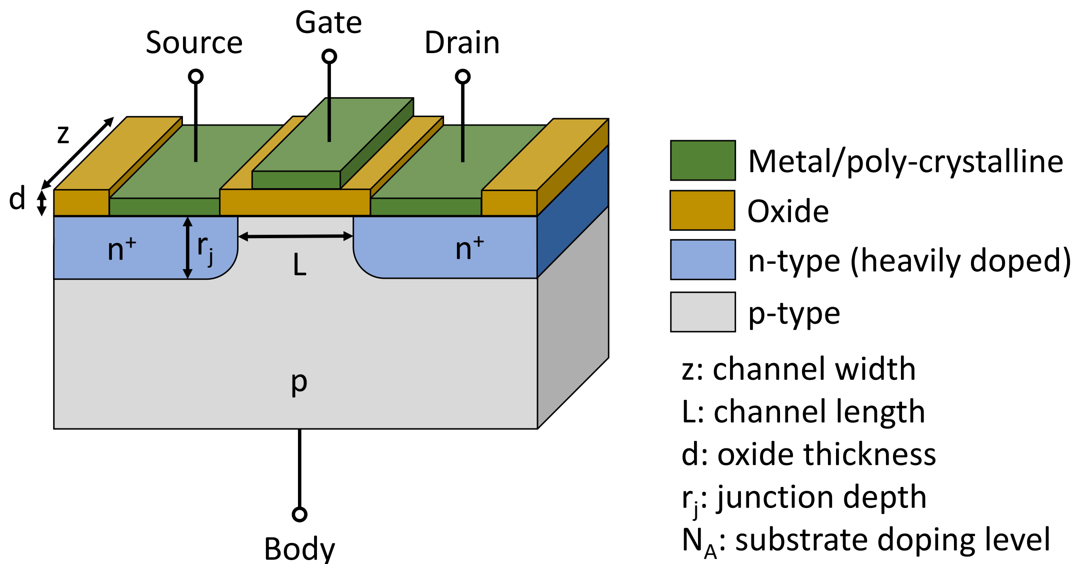


Figure 1.2 Perspective view of an n-type MOSFET. The basic device parameters are labelled.

The voltage applied to the gate controls the flow of charge carriers between the source and the drain in a MOSFET. In the case of an n-type MOSFET, when a positive voltage is applied to the gate electrode, it creates a depletion layer by repelling the positively charged holes away from the interface between the gate insulator and the semiconductor. This process exposes a carrier-free region near the interface where immobile, negatively charged acceptor ions reside. If the gate voltage is sufficiently

high, an inversion layer forms adjacent to the interface between the insulator and the semiconductor, characterised by a high concentration of negative charge carriers.

One of the most important parameters in MOSFET design is the threshold voltage, which determines the minimum voltage required to activate a MOSFET and initiate conduction. The threshold voltage is influenced by several factors such as the doping level of the substrate, the choice of gate oxide material, the oxide thickness, as well as the dimensions of the channel. Precise control of these factors can effectively regulate the threshold voltage and improve the performance of MOSFETs.

The drain voltage is another significant parameter in controlling the performance of MOSFETs, which affects the current flow in conducting channel. Initially, the drain current increases linearly with the applied drain-to-source voltage. However, if the drain voltage is increased beyond the saturation value, the depletion region around the reverse-biased drain junction will expand towards the source, which leads to current saturation.

1.2.2 Scaling Challenges in MOSFET Technology

The first MOSFET was created in 1960 by Kahng and Atalla [5] using a thermally oxidised silicon substrate with a 20 μm channel length and a 100 nm gate oxide thickness. The technological revolution in MOSFETs scaling can be traced back to the 1970s when Intel introduced its first single-chip microprocessor, the "Intel 4004". The processor featured a 10 μm technology node and held 2300 transistors. For decades, the gate length of MOSFETs in ICs has been aggressively scaled down, driven by the need for higher packing density, lower power consumption, and improved device performance. As a result, the number of transistors per unit area has doubled every 18 months, consistent with the prediction of Moore's law [2]. While Moore's law continues to provide a framework for technology development, the trend of downscaling is now facing significant challenges and slowing down due to physical limitations and increasing manufacturing complexity.

The short-channel effects (SCEs) are ubiquitous and significant challenges that arise with MOSFETs scaling. The SCEs occur when the channel length decreases to a similar order of magnitude as the depletion layer widths of the source and drain junctions. The most predominant SCEs are the drain-induced barrier lowering (DIBL) and the punch-through effect [6,7]. Other typical SCEs include surface scattering, velocity saturation, impact ionisation, and hot electrons [8]. The occurrence of SCEs

results in modifications to the threshold voltage, which is known as threshold voltage roll-off. The charge-sharing model [9] provides a simplified explanation for this phenomenon. SCEs can lead to increased leakage current, loss of gate control, mobility degradation, and unpredictable device behaviours, causing severe deterioration of device performance.

To mitigate the effects of SCEs, reductions in the dimensions of not only the gate length and width, but also other parameters such as oxide thickness, junction depth, and substrate doping, are necessary. The constant field scaling approach [10] provides an effective method for achieving this goal. By scaling down all dimensions and voltages by a specific factor, the internal electric field remains at the same level. Although this scheme can lead to reduced power consumption and improved device performance, the reliability of the gate oxide becomes a greater concern as the oxide thickness continues to decrease. Once the thickness of the SiO_2 used in the traditional MOSFET is reduced below 2 nm, the gate leakage current caused by direct electron tunnelling through the oxide exceeds the maximum tolerable value (1 A/cm^2 at 1 V) [11]. The direct tunnelling of carriers is a critical factor that reduces the device performance and limits MOSFET scaling.

1.2.3 Advances in MOSFET Scaling

The introduction of cutting-edge materials and the advances in device structure engineering have helped improve device performance while surpassing anticipated limits in MOSFETs scaling.

The use of high-κ dielectrics as gate oxide materials marked a breakthrough in MOSFETs scaling, due to their substantially higher relative dielectric constant (κ) in comparison to SiO_2 (κ = 3.9). This enabled the attainment of the same capacitance between the gate and the channel with a greater physical thickness of the gate materials, effectively reducing the tunnelling-induced leakage current. The gate capacitance of a MOSFET can be defined as

$$C = \frac{\kappa \varepsilon_0 A}{t}, \tag{1.1}$$

where C is the capacitance, κ is the dielectric constant of the oxide, ε_0 is the absolute dielectric constant of the vacuum, A is the surface area of the oxide, and t is the thickness of the oxide. In the design and fabrication of devices using high-κ dielectrics, the concept of equivalent oxide thickness (EOT) is introduced for convenience. EOT

represents the thickness of a SiO_2 layer that would yield the same level of electrical performance as the actual oxide layer in a MOSFET device. It is defined as

$$EOT = \frac{3.9}{\kappa_{high-\kappa}} t_{high-\kappa}, \tag{1.2}$$

where $\kappa_{high-\kappa}$ and $t_{high-\kappa}$ are the dielectric constant and thickness of the high-κ dielectrics, respectively. Here 3.9 is the dielectric constant of SiO_2. With the introduction of high-κ dielectrics, a more aggressive reduction in the MOSFET dimensions has been pursued to further enhance transistor performance. Some promising high-κ materials include HfO_2, Al_2O_3, and ZrO_2 [12–14]. It is important to note that ideal high-κ materials should also have a large band gap to prevent charge injection from the electrodes. However, a trade-off exists between the band gap and dielectric constant of well-known oxides, which typically possess either a large band gap (4–8 eV) or a high dielectric constant (4–20) [15]. To date, ideal dielectric materials that exhibit both features simultaneously are yet to be discovered.

With the advent of high-κ materials, using polycrystalline silicon (poly-Si) as the gate electrode has become less advantageous because some high-κ oxides can have an undesired reaction with poly-Si [12]. Besides, the low carrier density of poly-Si can lead to the development of a depletion region in proximity to the gate oxide interface when a gate voltage is applied, which increases the effective EOT. To resolve this issue, metal gate electrodes are applied, as they have virtually no depletion, no boron penetration, and extremely low sheet resistance [16]. The gate metals are selected primarily based on their work functions. For example, the optimisation of CMOS requires a work function of approximately 5.1 eV near the Si valence band (VB) for p-type MOSFET and approximately 4.0 eV near the Si conduction band (CB) for n-type MOSFET [12]. Using different metals with distinct work functions for the p-type MOSFET and n-type MOSFET would increase the complexity of integration. As an alternative approach, the same metal system can be used, but with different doping or alloy compositions to adjust the work function. However, this approach only covers a limited range of work functions [12].

In general, there is an ongoing need to improve the high-κ materials by achieving lower EOT and larger band gap, as well as developing appropriate metal gate stacks.

In addition to the development of gate materials, there is also active research into high-mobility channel materials. The superior electronic properties and high quality of the Si–SiO_2 interface have made silicon the dominant semiconductor material for decades. With the replacement of SiO_2 by high-κ oxides, it is natural to explore the use of other high-mobility channel materials such as Ge and III–V semiconductors [17,18]. However, as these 3D semiconductors are thinned down, the roughness of their surface can cause carrier scattering, which in turn degrades channel mobility [19].

The discovery of 2D materials has provided a promising solution to the problem. One important class of 2D materials is the transition metal dichalcogenide (TMD). Their monolayers are atomically thin and free of dangling bonds, which suppress the SCEs and carrier scattering effect. The structural defects that typically impact the performance of conventional MOSFETs also do not apply to TMDs. Furthermore, TMDs have decent band gaps of approximately 1 to 2 eV [20], which enable a good switch-off and a sufficient on-off ratio. Meanwhile, their relatively high carrier effective mass reduces the direct source-to-drain tunnelling current. Despite these significant benefits of 2D TMDs in addressing SCEs, the high contact resistance that arises when they interact with metals can limit the maximum driving current of 2D devices. Directly depositing metal electrodes onto 2D TMDs can partially destroy the intralayer covalent bonds and introduce defects. This results in Fermi level pinning (FLP), formation of Schottky contacts, and consequently, a larger contact resistance [21]. Therefore, a comprehensive understanding of the metal–TMD interfaces is essential to achieve Ohmic contacts with low contact resistance.

Alongside the exploration of new MOSFET materials, novel device structures are proposed to overcome the scaling limitations. The adoption of non-planar structures, such as fin field-effect transistors (FinFETs), has been a major achievement that enables scaling beyond the limits of planar transistors. Alternative device architectures like tunnel field-effect transistors (TFETs) [22] and gate-all-around field-effect transistors (GAAFETs) [23] are also being investigated to push the downscaling limit even further and can be seen as successors to FinFETs.

1.3 Metal–Semiconductor Contacts

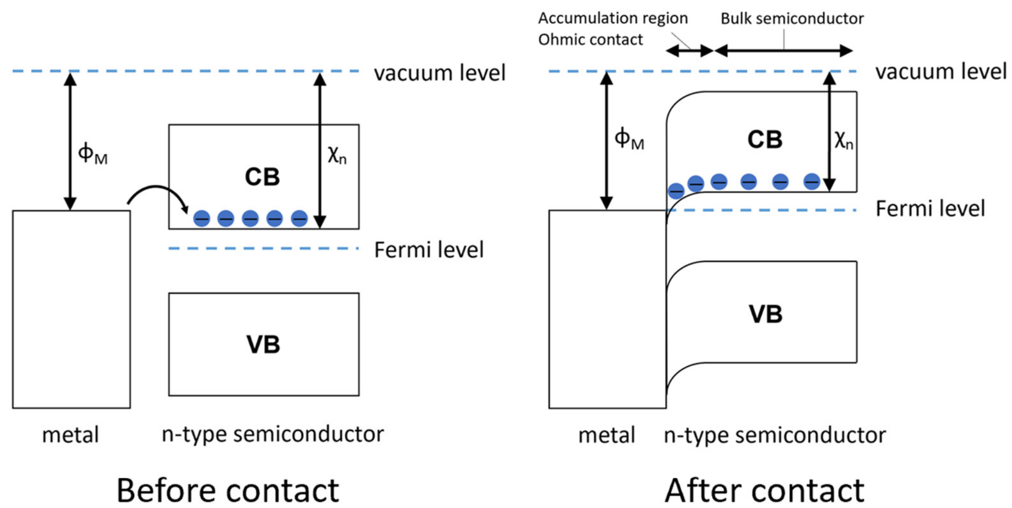
The electronic properties of metal–semiconductor contacts are critical in influencing MOSFET performance. The development of new contact materials and interfacial configurations is necessary to address the challenges of MOSFET scaling. Hence, a comprehensive understanding of the microscopic mechanisms involved in forming the metal–semiconductor interfaces is crucial for effective device design. This section will introduce the fundamental concepts in the physics of the metal–semiconductor interface, with an emphasis on barrier properties and theoretical methods for determining barrier height.

1.3.1 Ohmic Contacts and Schottky Contacts

When metal and semiconductor come into contact, the resulting junction can take two forms depending on their respective work functions. In the case of an n-type semiconductor, if the metal work function is lower than the semiconductor Fermi level, or if the potential barrier formed is significantly lower than the thermal energy, electrons will flow from the metal into the semiconductor side, as illustrated in Fig. 1.3(a). This results in the formation of a non-rectifying Ohmic contact, which exhibits a linear relationship between current and voltage, consistent with Ohm's law and has modest resistance in both directions. Ohmic contacts are crucial to the active region between a semiconductor device and the external circuit, as they allow charges to flow easily in both forward and reverse directions.

On the other hand, when the metal has a higher work function than the n-type semiconductor, a rectifying Schottky contact is formed due to the charge transfer at the interface, which creates a depletion layer at the semiconductor side. The barrier that prevents charge transfer from the metal towards the semiconductor is called the Schottky barrier. The Schottky barrier source/drain transistor (SSDT) has been suggested as an efficient approach to mitigating the DIBL effect in short-channel MOSFET [24].

(a) Ohmic contact



(b) Schottky contact

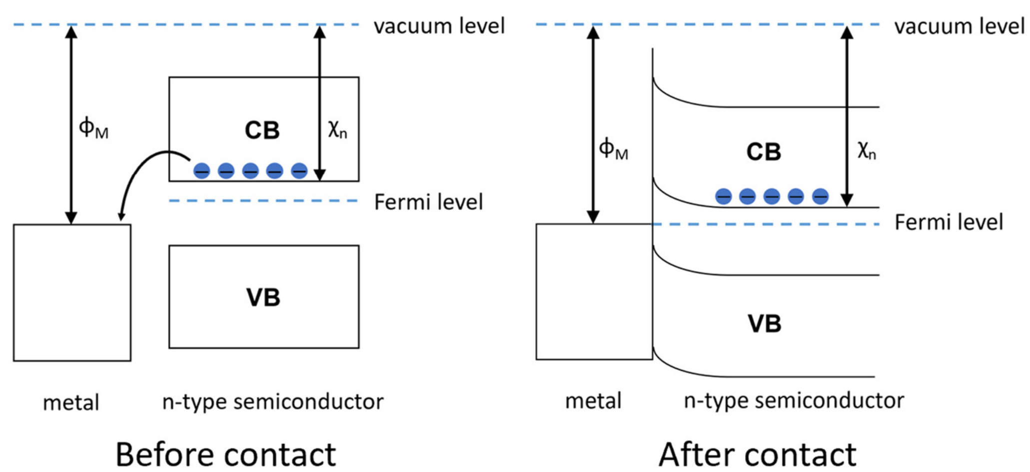


Figure 1.3 The formation of (a) Ohmic contact and (b) Schottky contact. When metal and semiconductor are brought together, the alignment of their Fermi energies leads to band bending.

1.3.2 Schottky Barrier Height

The Schottky barrier height (SBH) is a critical characteristic of metal–semiconductor contacts and has been the focus of extensive experimental and theoretical research. A well-understood formation mechanism is necessary for controlling SBH and minimising contact resistance. However, effective techniques for achieving desired SBH remain limited, and conflicting models exist for predicting SBH.

The ideal SBH can be described by the Schottky–Mott rule [25,26]:

$$\phi_n = \phi_M - \chi_S, \tag{1.3}$$

where ϕ_n is the n-type SBH, ϕ_M is the metal work function, and χ_S is the electron affinity (EA) of the semiconductor. The concept is based on naturally aligning the bands of metal and semiconductor while maintaining their respective energy distance from the vacuum level unchanged at the interface. However, the Schottky–Mott rule has limited validity because it assumes no electronic or atomic rearrangement during interface formation. Many experimental studies have found that the barrier height has a weak dependence on the work function [27,28].

In practice, the behaviours of SBH are influenced by induced states within the energy gap of the semiconductor during the metal–semiconductor contact formation. These chargeable states can accommodate electrons donated by the metal and effectively screen the semiconductor from metal details. There are several types of induced states inside the energy gap of the semiconductor, which can be classified as intrinsic or extrinsic. The intrinsic states typically include surface states induced by the termination of the semiconductor, where the wave function decays exponentially into both the vacuum and the semiconductor side [Fig. 1.4(a)]; and metal-induced gap states (MIGS), which occur due to the metal wave function tailing into the semiconductor band gap [Fig. 1.4(b)]. The extrinsic states are usually caused by interface imperfections, such as grain boundaries, impurities, and improper contact with metals [29].

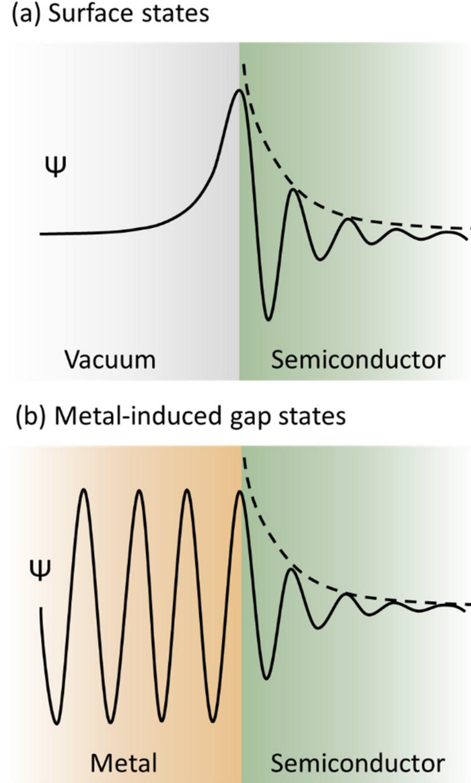


Figure 1.4 Schematic diagram of the intrinsic gap states induced in the semiconductor. (a) illustrates surface states induced by semiconductor termination, with the wave function decaying into both the vacuum and the semiconductor. (b) shows MIGS as tails of the metal Bloch states decaying into the semiconductor band gap.

1.3.3 Fermi Level Pinning: the MIGS Model

The weak dependence of SBH on metal work function is related to the FLP effect induced by the gap states. Taking this phenomenon into consideration, SBH shows a linear dependence on metal work function, which is modulated by a pinning factor S [30,31]:

$$\phi_n = S(\phi_M - \phi_S) + (\phi_S - \chi_S), \tag{1.4}$$

where S is the FLP factor which varies between 0 and 1, and ϕ_s is the reference pinning energy of the semiconductor. The parameter S represents the ability of the semiconductor to shield against external effects arising from the metal. When S = 1, there is no charge transfer at the interface, and the pinning effect is absent. This condition is referred to as the Schottky limit. Under this circumstance, the SBH is determined by the difference between the metal work function and the semiconductor EA, following the Schottky–Mott rule. Conversely, when S = 0, a significant amount

of charge transfer occurs, driven by the energy difference between the metal work function and the reference energy level, leading to a pronounced pinning effect. This is described as the Bardeen limit. The pinning factor S can be calculated using the linear response model of Cowley and Sze [32]:

$$S = \frac{1}{1 + \frac{e^2 N\lambda}{\varepsilon\varepsilon_0}}, \tag{1.5}$$

where e is the electronic charge, N is the density of the interface states per unit area, λ is their decay length into the semiconductor, and ε is the dielectric constant of the semiconductor at the interface region.

Numerous theoretical models have been suggested to estimate the strength of pinning at the interface and determine the SBH. Most of these models rely on assessing the effectiveness of different types of induced states in accounting for FLP. Among the various assumptions, the MIGS model [33–40], which states that the metal-induced interface states have a dominant role in determining the SBH, is believed to be a reliable approach. The model has successfully predicted the chemical trends of S for a wide range of materials [39,40]. It provides a promising description of the interfacial electronic properties because the concentration of MIGS is relatively high in most cases, and MIGS tend to penetrate deeply into the semiconductor [41]. These factors make MIGS a major contributor to the dipole formation at the metal–semiconductor interface, particularly when compared to other types of induced interface states. Besides, MIGS have been observed in the experiment [42].

Tersoff [36,37] put forward the concept that dipole formation at the interface will pull the Fermi level closer to the charge neutrality level (CNL), which corresponds to a state of zero net charge density at the interface (as shown in Fig. 1.5). Tersoff calculated the CNL of MIGS and obtained good agreement between the theoretical values and experimental results. Furthermore, he suggested that the pinning factor S is inversely proportional to the optical dielectric constant ε_∞. These findings provide strong support for the validity of the MIGS model. Later, Mönch [38,39] derived an empirical formula for the pinning factor S by incorporating the optical dielectric constant into the basic MIGS model:

$$S = \frac{1}{1 + 0.1(\varepsilon_\infty - 1)^2}. \tag{1.6}$$

According to Mönch, the dependence of the pinning factor on the optical dielectric constant arises because calculations of the Nλ in Eq. (1.5) are determined by the average band gap energy of the semiconductor, which is related to the electronic polarisability ($\varepsilon_{\infty}-1$) of the semiconductor.

Figure 1.5 Schematic diagram of MIGS and CNL at the metal–semiconductor contact. The metal Fermi level tends to align with the CNL of the semiconductor as a result of charge transfer.

1.4 Heterojunctions

In MOSFETs, the heterojunction interfaces between the source/drain and channel regions are essential to ensure efficient carrier injection and extraction, which is crucial for device performance. This section will cover the fundamentals of heterojunctions, including various types of interfaces, mechanisms of charge transport, and basic models for determining the band offset.

1.4.1 Heterojunction Band Alignment

Heterojunctions can be categorised into three types based on the alignment of energy levels: type-I (straddling gap), type-II (staggered gap), and type-III (broken gap), as shown in Fig. 1.6. Each type exhibits distinct electronic properties that can be exploited to enhance the performance of semiconductor devices.

Type-I heterojunctions have the CB and VB edges of semiconductor A higher and lower than the corresponding band edges of semiconductor B, respectively. As a result, electrons and holes tend to accumulate at the CB and VB of semiconductor B, leading to stronger carrier recombination. Type-I heterojunctions have been observed to be a common type of contact between 2D materials [43,44].

Type-II heterojunctions have the CB and VB edges of semiconductor A both higher than the corresponding band edges of semiconductor B. For this type of energy level alignment, electrons will migrate to semiconductor B, while holes will transfer to semiconductor A. The separation of charge carriers into different layers can increase their lifetime and is desirable for photovoltaic applications [45,46]. Type-II heterojunctions are intensively studied and often considered the most effective among conventional heterojunctions. In addition to their numerous applications in photovoltaics, type-II heterojunctions have also been applied in TFETs using material systems such as strained Si/strained Ge [47] and III–V materials [48].

Type-III heterojunctions are similar to type-II heterojunctions but represent a more extreme case where the band gaps of the two semiconductors do not overlap. These heterojunctions are less common than the previous two cases. The non-overlapping band gaps make charge transport across the heterostructures via band-to-band tunnelling (BTBT), making type-III heterojunctions suitable for TFETs with high-speed operation and low-power consumption [49]. Recently, there have been theoretical and experimental reports on the design of TFETs based on van der Waals (vdW)

heterojunctions using 2D materials with type-III alignment, to achieve strain-free heterointerface and better channel control [43,49,50].

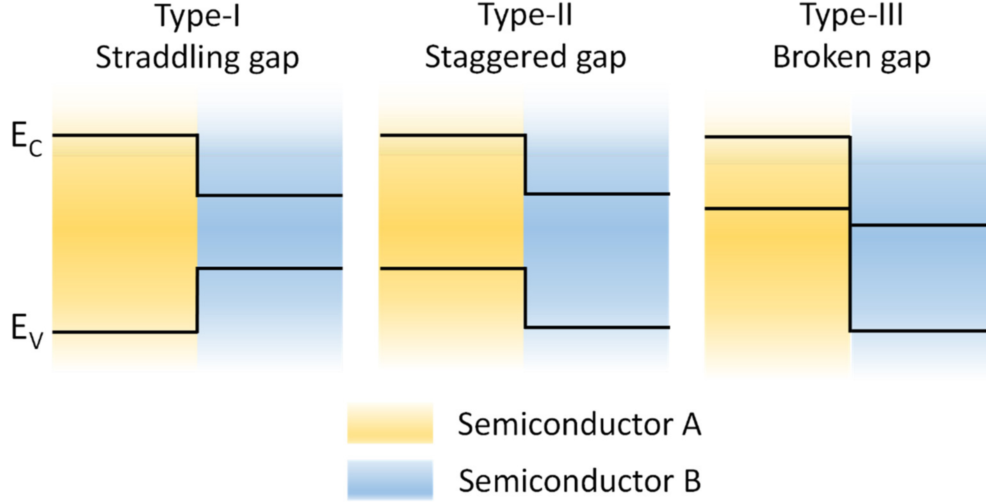


Figure 1.6 Different types of heterojunction band alignment. In type-I, the band gap of semiconductor B is completely contained within the band gap of semiconductor A. In type-II, the CB and VB of semiconductor B are both lower than the corresponding bands of semiconductor A. In type-III, there is no overlap between the band gaps of the two semiconductors.

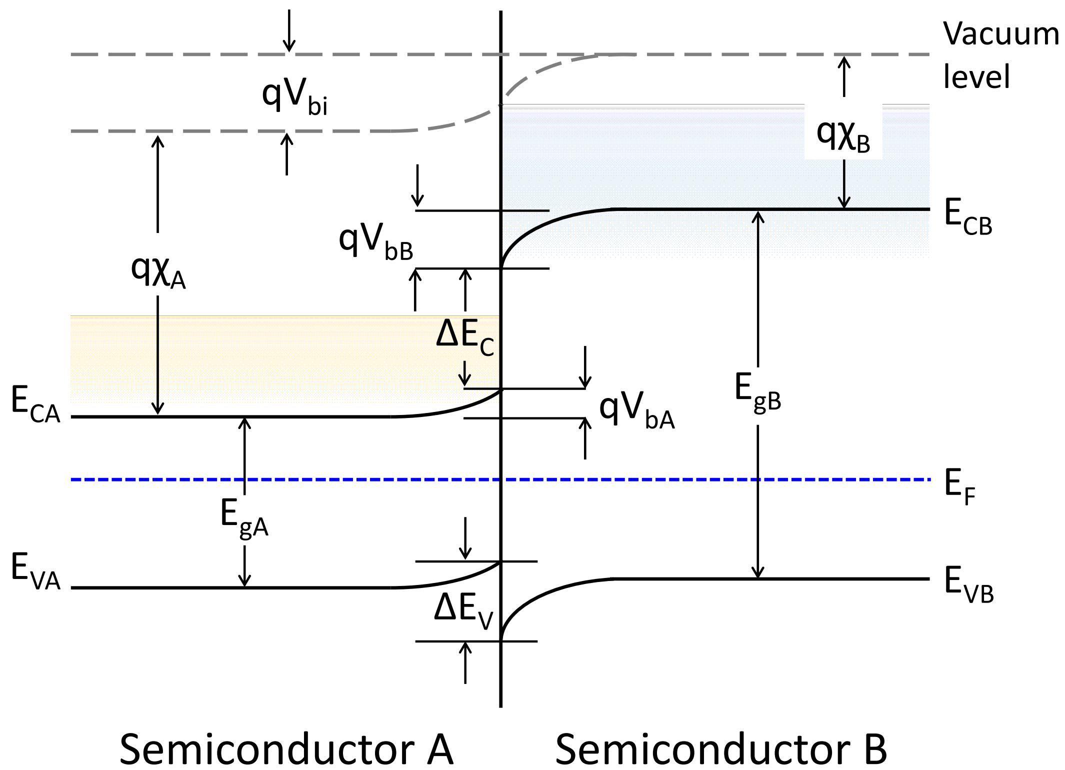
Figure 1.7 shows a schematic band diagram of a type-I heterojunction at equilibrium, to provide a better understanding of the interface electrostatics. The difference between the band gaps is distributed between the VB discontinuity ΔE_V, also known as valence band offset (VBO), and the CB discontinuity ΔE_C, also known as conduction band offset (CBO):

$$E_{gB} - E_{gA} = \Delta E_C + \Delta E_V. \tag{1.7}$$

where E_{gA} and E_{gB} are the band gap values of semiconductor A and semiconductor B, respectively. The difference in energy levels between the two semiconductors leads to charge transfer across the interface, resulting in the formation of an electric field and subsequent band bending. The potential difference across the junction, also known as the built-in potential, can be described using the band bending:

$$V_{bi} = V_{bA} + V_{bB}, \tag{1.8}$$

where V_{bi} is the built-in potential, V_{bA} and V_{bB} are the band bending values of semiconductor A and semiconductor B, respectively.



χ: electron affinity
E_C: conduction band minimum
E_V: valence band maximum
E_F: Fermi level
ΔE_C: conduction band offset
ΔE_V: valence band offset
V_{bi}: built-in voltage
V_b: band bending

Figure 1.7 Schematic band diagram for a type-I heterojunction at equilibrium, with the main parameters labelled.

1.4.2 Band Offset: Electron Affinity Rule and CNL Model

The band offsets at the interfaces are critical parameters that affect the electronic properties of heterojunctions. The VBO and CBO values determine the direction of charge transfer between the two semiconductors, leading to different carrier dynamics and recombination rates. Through precise control of the band offsets, researchers can design heterojunctions with tailored electronic properties for specific applications.

Several models have been proposed to predict the band offsets, including Anderson's EA model [51], model-solid approach [52,53], atom-centred Wigner–Seitz cells [54], and CNL model [37,55]. One common feature of these well-known theories is that they associate a reference energy level with each semiconductor. The reference level is considered an intrinsic property of the bulk material, and the band offsets can be predicted by aligning the reference levels of semiconductors. The models with the fixed internal reference level assumption usually use the concept of an interface dipole, although the magnitude and contribution of the dipole can vary widely between different models [55,56].

The simplest model for calculating the band offsets is the EA model, which states that the vacuum levels of two semiconductors on either side of the heterojunction should be aligned at the same energy. For example, consider semiconductors A and B, where A has a higher EA, the CBO of their heterojunction is given by:

$$\Delta E_C = \chi_A - \chi_B, \tag{1.9}$$

where χ_A and χ_B are the EA values of semiconductor A and semiconductor B, respectively. The EA model is closely related to the Schottky–Mott rule of the metal–semiconductor contacts. It is an idealised model that assumes the semiconductors at the heterojunction behave as they would in the limit of a large vacuum separation. In most cases, the EA rule fails to accurately predict the real band offsets, because the model attempts to use the bulk properties of semiconductors to describe the interface properties, while ignoring the reconstruction of the atomic and electronic structure at the interface.

The CNL model provides a more practical solution for predicting band offsets. The model is based on the idea that the interface states induced during the formation of heterojunction can accommodate the charge transfer and give rise to the pinning effect. The band alignment is described by matching the CNLs of the semiconductors, modified by a pinning factor S:

$$\Delta E_C = (\chi_A - \phi_{CNL,A}) - (\chi_B - \phi_{CNL,B}) + S(\phi_{CNL,A} - \phi_{CNL,B}), \quad (1.10)$$

where $\phi_{CNL,A}$ and $\phi_{CNL,B}$ are the CNLs of semiconductor A and semiconductor B, respectively. Note that the pinning factor S is for the semiconductor with a wider band gap in the heterojunction. When S = 1, there is no charge transfer across the junction and CBO equals the energy difference between the EAs of the two semiconductors, as stated by the EA model. When S = 0, there is substantial charge transfer between the semiconductors, which sets up a dipole. The surface dipole aligns the CNLs with each other, and the heterojunction is at the strong pinning condition.

The CNL model is an analogy to the MIGS model used to describe metal–semiconductor contacts. In the case of a metal–semiconductor interface, the induced gap states can be regarded as the metal wave function decaying into the semiconductor region. Similarly, for a semiconductor heterojunction, MIGS can be considered as the wave function of the narrower band gap semiconductor tailing into the band gap of the other semiconductor [37], as shown in Fig. 1.8. It is worth noting that for the semiconductor with a smaller band gap, the states observed within its band gap are introduced by defects [57,58] and are not considered as MIGS because there are no corresponding extended states by the wider gap semiconductor.

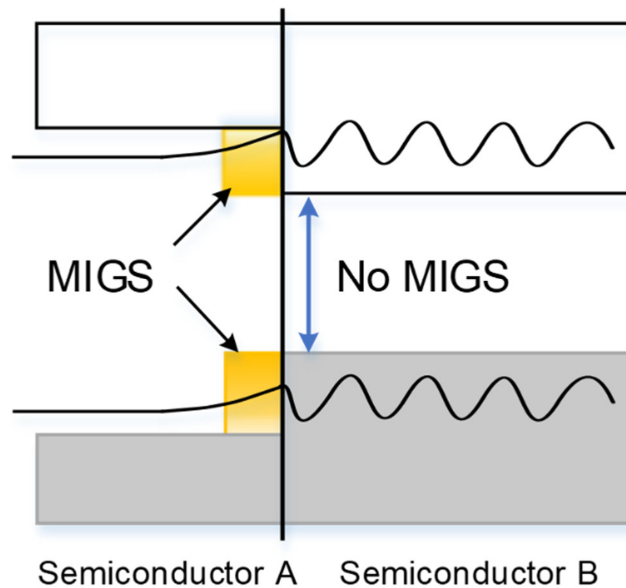


Figure 1.8 Schematic diagram of MIGS at the semiconductor heterojunction interface. The interface states only exist within the band gap of semiconductor A, where there is a band gap overlap with semiconductor B.

1.5 Motivation and Thesis Outline

The primary objective of this thesis is to investigate representative materials and their contacts used in the development of MOSFETs. A comprehensive treatment of conventional materials in the semiconductor industry is provided, while also addressing gaps in understanding the recent novel materials. The theories of SBH at metal–semiconductor junctions and the band alignment at semiconductor heterojunctions are studied to understand the physics of forming various contacts. Specifically, critical tests are performed based on the MIGS and CNL models using density functional calculations, to provide a clear theoretical explanation in atomic scale.

The thesis contains 7 chapters and is organised as follows:

Chapter 1 provides a discussion on the physics and characteristics of major semiconductor devices, with a particular emphasis on MOSFETs due to their importance in contemporary electronic applications. A brief historical review is given regarding the MOSFETs scaling trends, challenges, and solutions. This chapter also demonstrates the electronic properties of the metal–semiconductor junctions and the semiconductor heterojunctions, which are essential components in building MOSFETs.

Chapter 2 introduces the theoretical background of the simulation method applied in this study. The chapter begins with a brief introduction to density functional theory (DFT), which is a powerful theoretical tool widely used in materials science and condensed matter physics. The chapter focuses on the basic concepts underlying DFT, including the Hohenberg–Kohn theorems and the Kohn–Sham equations. The chapter then delves into the developments in exchange-correlation (XC) functionals, which are essential components of DFT calculations. The limitations of the local density approximation (LDA) and the generalised gradient approximations (GGAs) are discussed, and the need for more accurate XC functionals is highlighted.

Chapter 3 presents a comprehensive investigation into the structural, electronic, and optical properties of nine different polymorphs of SiO_2. The study uses both the traditional GGA scheme and the state-of-the-art screened exchange (sX) functional to access the strengths and weaknesses of different functionals employed in DFT calculations. The research aims to develop a profound understanding of the industrially important Si–SiO_2 interface at the microscopic level.

Chapter 4 investigates the defect-free interfaces between metal and various oxides (ZnO, CdO, MgO, and SrO). The GGA + U scheme is applied to address the

widely recognised band gap underestimation of oxides by the GGA functional where necessary. The study models oxides with different terminations in contact with metals and evaluates the electronic properties, such as SBH, of the metal–oxide interfaces. The aim is to describe the pinning effect at the interface and test the applicability of the MIGS model.

Chapter 5 delves into the electronic properties of CaF_2, a potential candidate for gate insulators, due to its ultra-wide band gap and relatively high dielectric constant. The research explores the band structures of CaF_2 using both the sX functional and the GGA + U scheme. The GGA + U method is further employed to examine Si–CaF_2 and metal–CaF_2 interfaces with different CaF_2 terminations. The investigation seeks to assess the accuracy of sX in characterising wide band gap materials and extend the application of the GGA + U scheme to fluorides and associated contacts.

Chapter 6 studies the metal contacts on the 2D TMDs. The chapter begins by summarising the recent achievements in metal–TMD contacts. Building upon previous work on the rational design for metals on n-type MoS_2, this study extends the calculations to the p-type and ambipolar WSe_2. The research aims to provide valuable theoretical insights for constructing low-resistance metal–TMD contacts and guide the design of novel 2D semiconductor devices.

Chapter 7 summarises the main achievements of this thesis and provides an outlook for future work.

CHAPTER 2 SIMULATION THEORY

First-principles calculations have emerged as a powerful and versatile tool for predicting material properties, owing to advances in computing technology and algorithms. By definition, these methods derive physical properties directly from the fundamental principles of quantum mechanics without the need for experimental data or empirical models. First-principles calculations provide robust and accurate insights into the properties of materials at the atomic and electronic levels, helping to accelerate the discovery and development of new materials with desirable properties.

This chapter describes the most widely used first-principles approach today—DFT. The chapter begins with an introduction to elementary concepts from basic quantum mechanics, focusing on the classical Hartree–Fock (HF) approximation. The use of electron density as a fundamental variable is discussed, and the development of DFT is introduced, from the seminal theorems of Hohenberg and Kohn to the practical approach by Kohn and Sham. The chapter concludes with an overview of the currently popular approximate XC functionals.

2.1 Elementary Quantum Mechanics

2.1.1 The Schrödinger Equation

The understanding of the electronic structure of materials roots firmly in solving the fundamental time-independent, non-relativistic Schrödinger equation:

$$H\Psi = E\Psi, \tag{2.1}$$

where H is the Hamiltonian operator, Ψ is a set of eigenstates of the Hamiltonian, and E is the associated eigenvalue of each eigenstate. The detailed definition of the Hamiltonian depends on the physical system being described by the Schrödinger equation. For a many-body system, where multiple electrons interact with multiple nuclei, the Hamiltonian operator can be represented as:

$$\hat{H} = -\sum_{I}\frac{\hbar^2}{2M_I}\nabla_I{}^2 + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|\overrightarrow{R_I}-\overrightarrow{R_J}|} - \sum_{i.I}\frac{Z_I e^2}{|\overrightarrow{r_i}-\overrightarrow{R_I}|} - \frac{\hbar^2}{2m}\sum_{i}\nabla_i{}^2 + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\overrightarrow{r_i}-\overrightarrow{r_j}|}, \tag{2.2}$$

where electron-related parameters are denoted by lower subscripts (charge e, mass m, and location $\vec{r}$), while nucleus-related parameters are denoted by upper subscripts (charge Z, mass M, and location $\vec{R}$). $\hbar$ is the reduced Planck constant. The five terms in Eq. (2.2), in order, represent the kinetic energy of nuclei, the repulsion between nuclei, the attraction between nuclei and electrons, the kinetic energy of electrons, and the repulsion between electrons, respectively.

Solving the many-body problem exactly is notoriously difficult due to the large number of particles involved and the complexity of their interactions. The first simplification made here is to separate the motion of the nuclei and electrons, known as the Born–Oppenheimer approximation. This approximation assumes that the nuclei are much more massive than the electrons, which allows them to be considered stationary. In this case, the kinetic energy of the nuclei is zero, and the potential energy due to the repulsion between nuclei is merely a constant. Therefore, the Hamiltonian operator given in Eq. (2.2) can be reduced to the Hamiltonian of electrons, in which the positions of the nuclei are parameters:

$$\hat{H} = \hat{T} + \hat{V}_{ext} + \hat{V}_{int} + E_{II}. \tag{2.3}$$

With the adoption of Hartree atomic units, the terms can be further simplified. $\hat{T}$ is the kinetic energy operator for electrons,

$$\hat{T} = -\frac{1}{2}\sum_{i} \nabla_i{}^2; \tag{2.4}$$

$\hat{V}_{ext}$ is the external potential imposed by the nuclei,

$$\hat{V}_{ext} = -\sum_{i.I} V_I(|\vec{r_i} - \vec{R_I}|); \tag{2.5}$$

$\hat{V}_{int}$ is the electron–electron interaction,

$$\hat{V}_{int} = \frac{1}{2}\sum_{i \neq j} \frac{1}{|\vec{r_i} - \vec{r_j}|}, \tag{2.6}$$

and E_{II} is the nucleus–nucleus interaction and other terms that contribute to the total energy of the system but are not relevant to describing the electrons. Eq. (2.3) reveals that the specific parts of the Hamilton operator that depend on the target system are the number of electrons in the system and the external potential. The solution of the simplified Hamiltonian yields the electronic wave functions and the energy of the

electrons for a given nuclear configuration. By solving the electronic wave function for a range of $\vec{R}$, the potential energy curve along with the nuclei move can be obtained.

Unfortunately, apart from a few simple systems, such as a single particle in a box or a hydrogen atom, there is no strategy to solve the Schrödinger equation exactly, and various approximation methods must be used. One of the most powerful and widely used approximation methods is the variational principle, which provides a way to obtain an approximate solution to the Schrödinger equation by minimising the energy of a trial wave function, which gives an estimate of the ground-state energy and wave function of the system. Recall that the time-independent expression for any observable is an expectation value of an operator $\hat{O}$, which involves an integral over all coordinates:

$$\langle \hat{O} \rangle = \frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \tag{2.7}$$

The variational principle states that the energy E_{trial} computed with Eq. (2.7) as the expectation value of the Hamilton operator $\widehat{H}$ from the normalised guessed Ψ_{trial} will be an upper bound to the ground-state energy E_0 with wave function Ψ_0:

$$\langle \Psi_{trial} | \widehat{H} | \Psi_{trial} \rangle = E_{trial} \geq E_0 = \langle \Psi_0 | \widehat{H} | \Psi_0 \rangle. \tag{2.8}$$

2.1.2 The Hartree–Fock Approximation

It is not feasible to search through all acceptable wave functions to find the ground-state energy, as the number of possible wave functions for a given system is typically too large to explore exhaustively. Instead, a subset of wave functions must be defined that offers a reasonable approximation to the exact wave function, while still being manageable in practice. One typical example is the HF approximation, which satisfies the antisymmetry principle (Pauli exclusion principle) when describing the wave function, using the Slater determinant:

$$\phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r_1}, \sigma_1) & \phi_1(\vec{r_2}, \sigma_2) & \cdots & \phi_1(\vec{r_N}, \sigma_N) \\ \phi_2(\vec{r_1}, \sigma_1) & \phi_2(\vec{r_2}, \sigma_2) & & \phi_2(\vec{r_N}, \sigma_N) \\ & \vdots & & \vdots \\ \phi_N(\vec{r_1}, \sigma_1) & \phi_N(\vec{r_2}, \sigma_2) & \cdots & \phi_N(\vec{r_N}, \sigma_N) \end{vmatrix}, \tag{2.9}$$

where the one electron functions $\phi_i(\vec{r_j}, \sigma_j)$ are called spin–orbitals, which are the products of a spatial function $\psi_i^{\sigma}(\vec{r_j})$ and a spin function $\alpha_i(\sigma_j)$. The expectation value of Eq. (2.3) with the wave function described by Eq. (2.9) is given by:

$$\langle\phi|\hat{H}|\phi\rangle = \sum_{i.\sigma}\int d\vec{r}\psi_i^{\sigma*}(\vec{r})\left[-\frac{1}{2}\nabla^2 + V_{ext}(\vec{r})\right]\psi_i^{\sigma}(\vec{r}) + E_{II}$$

$$+\frac{1}{2}\sum_{i,j,\sigma_i,\sigma_j}\int d\vec{r}d\vec{r'}\psi_i^{\sigma_i*}(\vec{r})\psi_j^{\sigma_j*}(\vec{r'})\frac{1}{|\vec{r}-\vec{r'}|}\psi_i^{\sigma_i}(\vec{r})\psi_j^{\sigma_j}(\vec{r'})$$

$$-\frac{1}{2}\sum_{i,j,\sigma}\int d\vec{r}d\vec{r'}\psi_i^{\sigma*}(\vec{r})\psi_j^{\sigma*}(\vec{r'})\frac{1}{|\vec{r}-\vec{r'}|}\psi_j^{\sigma}(\vec{r})\psi_i^{\sigma}(\vec{r'}). \quad (2.10)$$

The first term is the kinetic energy and the electron–nucleus attraction, which involves a sum over orbitals, whereas the third and fourth terms are the Coulomb and exchange integrals among electrons, which are double sums. The eigenenergy from Eq. (2.10) is a functional of the spin–orbitals. The Lagrange multipliers ε_i are introduced to maintain the orthonormality through minimisation and have their physical interpretation as orbital energies. The resulting equations represent the HF equations:

$$\left[-\frac{1}{2}\nabla^2 + V_{ext}(\vec{r}) + \sum_{j,\sigma_j}\int d\vec{r'}\psi_j^{\sigma_j*}(\vec{r'})\psi_j^{\sigma_j}(\vec{r'})\frac{1}{|\vec{r}-\vec{r'}|}\right]\psi_i^{\sigma}(\vec{r})$$

$$-\sum_j\int d\vec{r'}\psi_j^{\sigma*}(\vec{r'})\psi_i^{\sigma}(\vec{r'})\frac{1}{|\vec{r}-\vec{r'}|}\psi_j^{\sigma}(\vec{r}) = \varepsilon_i^{\sigma}\psi_i^{\sigma}(\vec{r}). \quad (2.11)$$

It is worth noting that the exchange term is summed over all orbitals of the same spin including the self-interaction condition when i = j, which cancels the unphysical self-term included in the Coulomb integral. The HF scheme elegantly solved the self-interaction problem, which is the major obstacle in the density functional approaches discussed in the next section.

The HF equation is a pseudoeigenvalue problem that must be solved iteratively because the orbitals are derived from their effective potential. The specific procedure for solving the HF equation is called the self-consistent field (SCF) method. The SCF method begins by introducing a guessed set of orbitals, which are used to solve the HF equations. The resulting new set of orbitals is then used in the next iteration of the HF calculation. This process is repeated until the input and output orbitals converge to a self-consistent set. The use of a finite basis set of orbitals is necessary for the SCF method, as the number of molecular orbitals is infinite and cannot be solved exactly. The choice of basis set is important for obtaining accurate results, and a variety of basis sets with different levels of complexity and accuracy are available for different types of systems.

One of the most significant limitations of the HF approximation is that it cannot fully account for electron correlation effects, due to the approximation of describing the many-electron wave function using a single Slater determinant. The electron correlation is mainly caused by the instantaneous repulsion of the electrons which is not covered by the effective HF potential. The electrons are getting too close because the electrostatic interactions are treated in an average manner under the HF scheme. Therefore, the HF energy is often higher than the true ground-state energy of the system, and the difference is called the correlation energy.

The HF scheme is known for being computationally demanding, particularly when trying to account for the effects of electron correlation. Addressing these effects requires additional computational resources and the use of more sophisticated methods such as second-order perturbation theory or coupled cluster theory. These methods can offer a more accurate description of the electronic structure of the system. In addition, the calculation of the two-electron Coulomb and exchange integrals in the HF method can be time-consuming. The computational demands increase significantly for larger systems since the number of integrals that need to be calculated scales quadratically with the number of basis sets used.

2.2 Density Functional Theory

2.2.1 The Electron Density

The conventional methods in quantum mechanics, like the HF approximation discussed in Sec. 2.1.2, employ the wave function Ψ as a central component, as it provides access to all information about a specific state of a target system. However, the wave function is a complex quantity that depends on three spatial variables and one spin variable for each of the N electrons, making it difficult to handle systems with many atoms and electrons. Additionally, the wave function cannot be directly observed experimentally.

The fundamental tenet of DFT is that the properties of a many-electron system can be considered unique functionals of the ground-state electron density of the system $n_0(\vec{r})$, which is a scalar function of the position $\vec{r}$. As stated by Eq. (2.7), any observable is an expectation value of an operator $\hat{O}$. The density of particles $n(\vec{r})$, is given by the expectation value of the density operator $\hat{n}(\vec{r}) = \sum_{i=1,N} \delta(\vec{r} - \vec{r_i})$,

$$n(\vec{r}) = \frac{\langle \Psi | \hat{n}(\vec{r}) | \Psi \rangle}{\langle \Psi | \Psi \rangle} = N \frac{\int d^3r_2 \cdots d^3r_N \sum_{\sigma_1} |\Psi(\vec{r}, \vec{r_2}, \vec{r_3}, \cdots, \vec{r_N})|^2}{\int d^3r_1 d^3r_2 \cdots d^3r_N |\Psi(\vec{r_1}, \vec{r_2}, \vec{r_3}, \cdots, \vec{r_N})|^2}. \quad (2.12)$$

DFT offers several significant advantages by using electron density as its primary variable. Firstly, this approach simplifies the 4N variables present in the wave function down to only three spatial variables. Secondly, the electron density is measurable experimentally, allowing for direct comparisons between theoretical predictions and experimental results. Additionally, since electron density is a unique property of a system, it enables the creation of universal functionals applicable to a wide array of systems. Lastly, DFT can handle large and complex systems, enhancing its practical utility.

2.2.2 The Hohenberg–Kohn Theorems

The concept of density has been an aspect of quantum mechanics since its early days, as evident in the Thomas–Fermi model [59,60], which is considered of historical interest. The decisive step towards a full-fledged DFT was taken by two theorems proved by Hohenberg and Kohn in 1964 [61]. The first one states that the external potential $\hat{V}_{ext}$ of a many-electron system is uniquely determined by its ground-state electron density $n_0(\vec{r})$. As discussed in Sec. 2.1.1, the Hamiltonian can be described by the external potential and the total number of electrons N, which can be calculated from the electron density by integration over space. Therefore, an important corollary of the first Hohenberg–Kohn theorem is that the Hamilton operator can be fully determined by the ground-state electron density, which in turn characterises the wave functions of all states (ground and excited). This theorem is of paramount importance as it asserts that, with just the ground-state electron density, all properties of a system can be accurately described. However, the theorem does not give a prescription for the actual functional, which is addressed by the second Hohenberg–Kohn theorem.

The second theorem states that the ground-state energy of a system can be obtained by minimising the energy functional with respect to the electron density, and the density that minimises the functional is the exact ground-state density. As all properties can be uniquely identified using $n(\vec{r})$, they can be regarded as functionals of $n(\vec{r})$. Consequently, the energy functional $E_{HK}[n]$ can be expressed as follows:

$$E_{HK}[n] = T[n] + E_{int}[n] + \int d^3r\, V_{ext}(\vec{r}) n(\vec{r}) + E_{II}. \quad (2.13)$$

The terms in order are the kinetic energies of electrons, the interaction energy between electrons, the external potential imposed by the nuclei, and the interaction energy of the nuclei. The energy expression can be conveniently divided into two components: system-dependent terms, which encompass the external potential and nuclear interactions; and universally valid terms, which cover all internal energies. The system-independent parts are combined into a new functional $F_{HK}[n]$:

$$F_{HK}[n] = T[n] + E_{int}[n]. \quad (2.14)$$

The theorem can be easily proven with the variational principle [Eq. (2.8)]. Consider a trial electron density $n_1(\vec{r})$ corresponds to external potential $V_{ext1}(\vec{r})$, the energy functional equals the expectation value of the Hamilton operator with the wave function Ψ_1 can be expressed as:

$$E_{HK}[n_1] = \langle \Psi_1 | \widehat{H_1} | \Psi_1 \rangle = E_1 \geq E_0 = \langle \Psi_0 | \widehat{H} | \Psi_0 \rangle. \quad (2.15)$$

The equality holds only when the trial density $n_1(\vec{r})$ equals the true ground-state density $n_0(\vec{r})$ with Ψ_0. In summary, if the functional $F_{HK}[n]$ is known, minimising the total energy of the system concerning variations in the density function $n(\vec{r})$ would yield the exact ground state density and energy. Note that the functional only gets access to the ground state properties and cannot be utilised to determine energies and properties of electronically excited states. While the Hohenberg–Kohn theorems provide the foundation for DFT by establishing the existence of a universal functional that relates the ground-state energy to the electron density, they do not offer a practical way to compute the functional itself. Besides, it is not possible to accurately calculate the kinetic energy of a system solely from the electron density, without resorting to an expression in terms of a set of wave functions.

2.2.3 The Kohn–Sham Approach

In 1965, Kohn and Sham [62] developed a ground-breaking method to simplify the analysis of electron interactions within many-body systems. Their approach involved replacing the complex interacting electron system with a fictitious non-interacting system, governed by an effective potential. This innovation facilitated the use of independent-particle methods for accurate calculations of many-body system properties and enabled the development of highly successful approximate formulations.

The resulting computational efficiency allowed large-scale electronic structure calculations in various materials.

The Kohn–Sham method assumes that the ground-state electron density of the non-interacting system is equivalent to that of the original interacting system. This assumption yields a set of independent-particle equations for the non-interacting system, which are effectively solvable and encompass the challenging many-body terms within an XC functional dependent on the electron density. By solving these equations, one can determine the ground state density and energy of the original interacting system, providing valuable insights into its electronic structure.

The Kohn–Sham construction of the system starts by defining an auxiliary Hamiltonian for an auxiliary independent-particle system using the Hartree atomic units:

$$\hat{H}_{aux}^{\sigma} = -\frac{1}{2}\nabla^2 + V^{\sigma}(\vec{r}). \tag{2.16}$$

For a system with N independent electrons that follows the auxiliary Hamiltonian, the ground state has one electron in each of the N^{σ} orbitals ψ_i^{σ} with the lowest eigenenergies ϵ_i^{σ}. The electron density is defined as:

$$n(\vec{r}) = \sum_{\sigma} n(\vec{r},\sigma) = \sum_{\sigma}\sum_{i=1}^{N^{\sigma}} |\psi_i^{\sigma}(\vec{r})|^2. \tag{2.17}$$

The Hohenberg–Kohn expression [Eq. (2.13)] can be rewritten with the Kohn–Sham approach:

$$E_{KS} = T_s[n] + E_{Hartree}[n] + \int d^3r\, V_{ext}(\vec{r})n(\vec{r}) + E_{II} + E_{XC}[n], \tag{2.18}$$

where T_s is the independent-particle kinetic energy, expressed as a functional of the orbital:

$$T_s = -\frac{1}{2}\sum_{\sigma}\sum_{i=1}^{N^{\sigma}} \langle\psi_i^{\sigma}|\nabla^2|\psi_i^{\sigma}\rangle = \frac{1}{2}\sum_{\sigma}\sum_{i=1}^{N^{\sigma}} \int d^3r\, |\nabla\psi_i^{\sigma}(\vec{r})|^2, \tag{2.19}$$

$E_{Hartree}$ is the Coulomb interaction energy by treating the electron density $n(\vec{r})$ as a classical charge density, i.e., the self-interaction energy of $n(\vec{r})$:

$$E_{Hartree}[n] = \frac{1}{2}\int d^3r d^3r' \frac{n(\vec{r})n\left(\vec{r'}\right)}{\left|\vec{r} - \vec{r'}\right|}, \tag{2.20}$$

and E_{XC} is the XC energy that includes all many-body effects. The XC energy corrects for the unphysical self-interaction by introducing non-classical effects that arise from the interactions between electrons, such as the exchange and correlation energies. As discussed in Sec. 2.1.2, the exchange energy arises from the antisymmetric nature of the wave function, while the correlation energy arises from the repulsion between electrons. Combining the Eq. (2.13) of Hohenberg–Kohn and Eq. (2.18) of Kohn–Sham, E_{XC} can be expressed as:

$$E_{XC}[n] = T[n] - T_s[n] + E_{int}[n] - E_{Hartree}[n]. \tag{2.21}$$

The first two terms account for the difference between the kinetic energy of the interacting and non-interacting systems, while the last two terms represent the non-classical electrostatic contributions. Eq. (2.21) reveals that the XC term includes not only the self-interaction correction that contributes to the potential energy but also a portion that contributes to the kinetic energy.

Derived from Eq. (2.18), the Kohn–Sham equation is expressed as:

$$(H_{KS}^{\sigma} - \varepsilon_i^{\sigma})\psi_i^{\sigma}(\vec{r}) = 0, \tag{2.22}$$

where ε_i^{σ} are the eigenvalues and H_{KS}^{σ} is the effective Hamiltonian:

$$\begin{aligned} H_{KS}^{\sigma}(\vec{r}) &= -\frac{1}{2}\nabla^2 + \frac{\delta E_{ext}}{\delta n(\vec{r},\sigma)} + \frac{\delta E_{Hartree}}{\delta n(\vec{r},\sigma)} + \frac{\delta E_{XC}}{\delta n(\vec{r},\sigma)} \\ &= -\frac{1}{2}\nabla^2 + V_{ext}(\vec{r}) + V_{Hartree}(\vec{r}) + V_{XC}^{\sigma}(\vec{r}). \end{aligned} \tag{2.23}$$

The Kohn–Sham equations can be solved through a self-consistent iterative process. In simple terms, an initial electron density $n(\vec{r})$ is proposed, and the Kohn–Sham equations are solved using this trial density to obtain the single-electron wave function $\Psi_i(\vec{r})$. Subsequently, the electron density derived from this wave function is compared to the trial density. Once the calculated electron density $n_{KS}(\vec{r})$ matches the trial density $n(\vec{r})$, the ground-state electron density has been achieved.

It is important to note that the Kohn–Sham approach itself involves no approximations. If the precise forms of the XC energy $E_{XC}[n]$ and potential $V_{XC}^{\sigma}(\vec{r})$ are known, then the exact ground state density and the energy of the interacting system can be computed. As a result, the primary objective of contemporary DFT is to continuously improve the approximations for these two critical quantities.

2.3 Exchange-correlation Functionals

An accurate description of the XC functional is essential for predicting the electronic structure, energetics, and other properties of materials. Despite the immense success of DFT, the exact form of the XC functional remains elusive, prompting researchers to explore a broad range of approximations in search of improved accuracy and computational efficiency.

2.3.1 Local Density Approximation

The simplest system to derive the XC functional is based on the idea of a hypothetical homogeneous electron gas (HEG): the electrons are evenly spread and attain a constant density at all points in space. The uniform electron gas system is crucial in DFT, as it represents the only case where XC functionals can be derived with high accuracy. This makes it an essential benchmark system for testing new theoretical and computational methods in electronic structure calculations.

Under this circumstance, the LDA is proposed, in which the XC energy density at each point is assumed to be the same as that of a uniform electron gas with the same density:

$$\begin{aligned} E_{XC}^{LDA}[n] &= \int d^3 r n(\vec{r}) \epsilon_{XC}^{hom}[n(\vec{r})] \\ &= \int d^3 r n(\vec{r}) \{\epsilon_x^{hom}[n(\vec{r})] + \epsilon_c^{hom}[n(\vec{r})]\}. \end{aligned} \tag{2.24}$$

where $\epsilon_{XC}^{hom}[n(\vec{r})]$ is the XC energy per particle of a uniform electron gas of density $n(\vec{r})$, it can be further separated in the exchange part $\epsilon_x^{hom}[n(\vec{r})]$ and correlation part $\epsilon_c^{hom}[n(\vec{r})]$. The exchange energy is given by:

$$\epsilon_x^{hom}[n(\vec{r})] = -\frac{3}{4}\sqrt[3]{\frac{3n(\vec{r})}{\pi}}, \tag{2.25}$$

and the correlation energy has been calculated with great accuracy by Ceperly and Alder [63] using Monte Carlo methods.

Despite its simplicity and computational efficiency, the constant electron density assumption is not the case for most of the real materials, which typically have rapidly varying densities. The LDA approximation exhibits significant limitations for

materials with strong energy gradients, such as insulators and semiconductors, where the band gap is substantially underestimated.

2.3.2 Generalised Gradient Approximations

The LDA has served as a basis for the development of many functionals, with the most well-known being the class of GGAs. GGAs have marked improvement over the LDA for many cases and are widely used to provide the required accuracy for DFT calculations.

The GGA approaches use not only the electron density $n(\vec{r})$ at a particular position $\vec{r}$, but also information about the gradient of the charge density $\nabla n(\vec{r})$, to account for the non-homogeneity of the true electron density. It refers to a collection of methods, classified according to the different ways in which the gradient is incorporated. The PW91 [64] and Perdew–Burke–Ernzerhof (PBE) [65] are the most widely used functionals. The GGA functionals can be written in general form:

$$\begin{aligned} E_{XC}^{GGA}[n] &= \int d^3 r n(\vec{r}) \epsilon_{XC}[n(\vec{r}), |\nabla n(\vec{r})|] \\ &= \int d^3 r n(\vec{r}) \epsilon_x^{hom}[n(\vec{r})] F_{XC}[n(\vec{r}), |\nabla n(\vec{r})|], \end{aligned} \tag{2.26}$$

where $F_{XC}[n(\vec{r}), |\nabla n(\vec{r})|]$ is the enhancement factor whose explicit form defines the exchange and correlation contribution to the energy. Although GGA functionals have significantly improved the accuracy of DFT calculations compared to LDA, the approach is semi-local and still neglects the long-range character of electron–electron interaction. The deficiency of the traditional LDA and GGA approaches is particularly pronounced in materials characterised by localised electrons and strong interactions, such as transition metal oxides, rare earth elements, and their compounds.

2.3.3 Orbital-Dependent Functionals: LDA + U

To address the limitations of traditional DFT functionals, researchers have developed various techniques to extend the functional approach by incorporating effects considered significant from a physical perspective. One such method is the LDA + U approach [66–68], which enhances standard DFT calculations with an additional orbital-dependent interaction inspired by the Hubbard model [69,70].

Although referred to as LDA + U, the method can encompass LDA or GGA calculations.

The multi-band Hubbard model postulates that strongly correlated d or f electrons are subject to on-site quasi-atomic interactions. This model introduces the crucial Hubbard parameter U, which represents the Coulomb energy cost of placing two electrons at the same site. In the LDA + U approach, electrons are divided into two subsets: delocalised electrons (s and p electrons), for which the LDA/GGA description is adequate, and localised electrons (d and f electrons), which require consideration of Coulomb interaction. By applying an additional on-site Coulomb interaction U to highly localised orbitals, the localised nature of electrons can be more accurately described. Within the LDA + U framework, the total energy of a system can be expressed as:

$$E_{LDA+U}[n] = E_{LDA}[n] + E_{Hub}\left[\{n_{mm'}^{I\sigma}\}\right] - E_{dc}[\{n^{I\sigma}\}]. \quad (2.27)$$

In this equation, $E_{LDA}[n]$ represents the energy functional within the standard DFT approximation, while $E_{Hub}\left[\{n_{mm'}^{I\sigma}\}\right]$ accounts for electron–electron interaction and is modelled using the Hubbard Hamiltonian. The variable $\{n_{mm'}^{I\sigma}\}$ represents the occupation numbers of the localised orbitals, where n indicates the occupation number, I corresponds to the atomic site index, and m refers to the state index. To avoid duplication of the interaction energy already captured in E_{Hub}, the double-counting term $E_{dc}[\{n^{I\sigma}\}]$ is subtracted from $E_{LDA}[n]$. The term $E_{dc}[\{n^{I\sigma}\}]$ represents the contribution of correlated electrons using a mean-field approximation to E_{Hub}.

The LDA + U method enhances the representation of electronic structures and energetics in materials where conventional DFT methods might fall short, particularly in transition metal oxides and materials with localised 3d and 4f electron states [71]. This method is favoured for its simplicity and relatively low computational cost. However, it is crucial to accurately determine the U terms, as using an excessively large U value could introduce artificial effects or discontinuities in the electronic structure that do not reflect the actual physics of the materials. The inaccuracies could lead to misleading conclusions or predictions regarding the behaviours of materials.

2.3.4 Hybrid Functionals

Hybrid functionals have emerged as a powerful alternative to traditional LDA and GGA methods, offering enhanced accuracy in predicting material properties. These functionals blend a fraction of orbital-dependent HF exchange with an explicit density functional, resulting in a more refined representation of XC effects and improved property predictions.

The construction of hybrid functionals relies on the concept of adiabatic connection, which enables a continuous connection between the non-interacting Kohn–Sham system and the fully interacting electron system. This is achieved by introducing a coupling strength parameter λ, scaling the electron–electron interaction from 0 (non-interacting) to 1 (fully interacting).

The coupling constant integration formula calculates the XC energy along the adiabatic connection by integrating non-classical contributions to the total energy concerning λ from 0 to 1. In the context of hybrid functionals, the two limits of this integration provide crucial information. At $\lambda = 0$, the non-interacting system has only exchange contributions, with the exchange energy exactly calculated using the HF formalism. At $\lambda = 1$, the fully interacting system contains both exchange and electron correlation contributions, with the true XC energy approximated using traditional DFT functionals such as LDA or GGA. The energies for intermediate λ values are the only remaining information needed, and the suitable description of the path between the limits constitutes the key difference between various hybrid functionals.

The simplest form of hybrid functional is the half-and-half combination proposed by Becke [72], which assumes a linear dependence of potential energy on λ:

$$E_{XC}^{HH} = \frac{1}{2}E_{x}^{HF} + \frac{1}{2}E_{XC}^{DFA}, \tag{2.28}$$

where DFA refers to either LDA or GGA functional. Later, Stephens *et al.* [73] introduced B3LYP:

$$E_{XC}^{B3LYP} = (1-a)E_{XC}^{LSDA} + aE_{x}^{HF} + bE_{x}^{B88} + cE_{c}^{LYP} + (1-c)E_{c}^{LSDA}, \tag{2.29}$$

with empirically adjusted coefficients (a = 0.20, b = 0.72, and c = 0.81). This three-parameter functional combines HF exchange with Becke88 (B88) exchange and Lee–Yang–Parr (LYP) correlation. B3LYP has been widely used and benchmarked against

various experimental and high-level quantum chemistry results, demonstrating its applicability to numerous material systems.

Perdew *et al.* [74] developed PBE0, another well-known hybrid functional, which derives its form from theoretical coupling-constant integration:

$$E_{XC}^{PBE0} = \frac{1}{4}E_x^{HF} + \frac{3}{4}E_x^{PBE} + E_c^{PBE}. \tag{2.30}$$

This functional blends HF exchange with PBE exchange in a 1:3 ratio, based on theoretical considerations, and includes PBE correlation.

Several reasons necessitate the exploration beyond the conventional hybrid functionals, which employ a static fraction of HF exchange. In its current form, any functional possessing a non-zero fraction of HF exchange cannot be applied to metallic systems. The long-range part of the exchange should be entirely discarded as it yields non-physical outcomes at the Fermi surface. Furthermore, it has been noted that a greater fraction of HF tends to enhance the gaps for materials with larger gaps, while a smaller fraction of HF does the same for materials with smaller gaps [75].

More recently, range-separated hybrid functionals have been developed. These functionals separate short-range and long-range exchange contributions, providing a more balanced description of various chemical systems. A convenient division into short- and long-range parts can be expressed as follows:

$$\frac{1}{r} = \frac{1 - erf(\eta r)}{r} + \frac{erf(\eta r)}{r}, \tag{2.31}$$

where erf is the error function and η is the range parameter. The general formulation of range-separated hybrid functionals is given by:

$$E_{XC} = E_{XC}^{DFT} + a\left(E_x^{HF,SR} - E_X^{DFT,SR}\right) + b\left(E_x^{HF,LR} - E_X^{DFT,LR}\right). \tag{2.32}$$

Here, DFT represents any of the local or semi-local density functionals.

In solid-state calculations, a widely adopted option is the Heyd–Scuseria–Ernzerhof (HSE) functional [76,77]. This functional explicitly excludes the long-range HF term, thereby setting b = 0 in Eq. (2.32). Consequently, it assumes the form:

$$E_{XC} = E_{XC}^{PBE} + \frac{1}{4}\left(E_x^{HF,SR} - E_X^{PBE,SR}\right). \tag{2.33}$$

In this expression, a is selected as $\frac{1}{4}$, which is the same as in Eq. (2.30). One of the significant advantages of this formulation is its suitability for metallic systems. As the

parameter η approaches infinity, it simplifies to the PBE version of GGA; as η approaches zero, it becomes the PBE0 hybrid.

Following the development of the HSE functional, an analogous separation of the short- and long-range parts of the sX was proposed by Bylander and Kleinman [78]. This approach uses a screened, non-local exchange potential mixed with the LDA. Despite its utility in band structure calculations [79–81], it was not extensively applied due to the computational costs associated with a plane wave basis. This situation changed with the calculation of the Hellman–Feynman stress terms for a plane wave basis, facilitating full structural energy minimisation and making sX more accessible for a range of calculations [82,83]. The sX functional is related to the HF method but is an implicit functional of the density, and part of the XC operator is non-local. Therefore, the Kohn–Sham orbitals are solutions of a non-local Schrödinger equation as follows:

$$-\frac{1}{2}\nabla^2\psi_i(\vec{r}) + V_{loc}(\vec{r})\psi_i(\vec{r}) + \int d\overrightarrow{r'}V_{nl}^{XC}\left(\vec{r},\overrightarrow{r'}\right)\psi_i\left(\overrightarrow{r'}\right) = \varepsilon_i\psi_i(\vec{r}). \tag{2.34}$$

The local potential term $V_{loc}(\vec{r})$ contains the Hartree potential, the local part of the XC potential, and the external pseudopotential terms in standard plane wave calculations. The term $V_{nl}^{XC}\left(\vec{r},\overrightarrow{r'}\right)$ represents the non-local part of the XC potential, which is similar to the HF potential but incorporates the correlation effects by screening the long-range interactions of exchange using an exponentially decaying factor.

In the sX method, the XC energy is divided into a local part and a non-local part:

$$E_{XC} = E_{XC}^{LOC} + E_{XC}^{NL}. \tag{2.35}$$

The local part E_{XC}^{LOC} is defined in a very similar way to the LDA, and the XC energy per particle is given by:

$$\epsilon_{XC}^{LOC}(\vec{r}) = \epsilon_{XC}^{hom}[n(\vec{r})] - \epsilon_{XC}^{NLhom}[n(\vec{r})], \tag{2.36}$$

where $\epsilon_{XC}^{NLhom}[n(\vec{r})]$ is the non-local XC energy per particle of a uniform electron gas of density $n(\vec{r})$. The non-local contribution to the total energy is given by:

$$E_{XC}^{NL} = -\frac{1}{2}\sum_{ij,kq}\iint d\vec{r}d\overrightarrow{r'}\frac{\psi_{ik}^*(\vec{r})\psi_{ik}\left(\overrightarrow{r'}\right)\psi_{jq}^*\left(\overrightarrow{r'}\right)\psi_{jq}(\vec{r})}{\left|\vec{r}-\overrightarrow{r'}\right|}e^{-k_s\left|\vec{r}-\overrightarrow{r'}\right|}, \tag{2.37}$$

where k and q are the k-points, and k_s is the reciprocal screening length.

However, range-separated hybrid functionals typically require more computational resources than traditional hybrid functionals like B3LYP and PBE0, making them less suitable for large-scale calculations or highly complex systems. Furthermore, the performance of range-separated hybrid functionals depends on the choice of parameters, such as the range-separation parameter, which can affect the accuracy of the functionals for different systems. This introduces an additional layer of complexity when selecting the most suitable functional for a specific problem.

Chapter 3 Revisiting the Electronic and Optical Properties of SiO_2 Polymorphs by Hybrid Functional Calculations

The structural, electronic, and optical properties of nine different silica polymorphs are investigated using both the traditional GGA scheme and the hybrid functional. The analysis includes the 4:2-coordinated systems of quartz, tridymite, cristobalite, coesite, and keatite, as well as the 6:3-coordinated stishovite. Phases with similar crystal structures tend to share comparable electronic and optical characteristics, while stishovite presents different properties. For the 4:2-coordinated polymorphs, strong linear relationships are established between the band gap and the static dielectric constant vs the unit cell volume, respectively. Further investigation of their refractive indices also reveals a notable linearity with the corresponding densities. Calculations using the sX functional effectively reproduce the experimental band gap values of silica, while GGA is superior in determining overall trends of optical properties.

3.1 Introduction

Silicon dioxide (SiO_2), or silica, is one of the most important materials for semiconductor industries benefiting from its sharp interface with Si and ease of synthesis [84]. Although SiO_2 is amorphous and Si is crystalline, the Si–SiO_2 interface retains its abruptness and smoothness, especially in the widely used (001)-orientated interface [85]. As microelectronics continue to scale down, 3D device structures such as the FinFET have been introduced to address the SCEs [86]. The development of non-planar devices involves high Miller index facets, resulting in more complex network topologies [87]. Therefore, a deeper understanding of the Si–SiO_2 interface at the microscopic level acquires enhanced technological importance.

Models interpreting the interface properties have been proposed, based on the assumption that crystalline SiO_2 acts as a thin transition layer [88–90]. However, SiO_2 has numerous polymorphic forms, and the appropriate candidate for such an epitaxial

oxide remains controversial: Ourmazd *et al.* [88] and Pasquarello *et al.* [89] suggested that the interface involves a tridymite monolayer, while Tersoff *et al.* [90] constructed an ordered interface between crystalline Si and amorphous SiO_2. The selection of silica polymorphs can result in different interface behaviours. Therefore, a systematic understanding of various SiO_2 polymorphs can offer valuable insights for describing Si–SiO_2 interfaces and may provide theoretical guidance for potential phases in constructing FinFETs. This knowledge can also be crucial for modelling recent gate-all-around (GAA) structures.

Most silica polymorphs are found abundant in nature, while some can be synthesised under laboratory conditions. The most common crystalline polymorphs include α- and β-quartz, α- and β-tridymite, α- and β-cristobalite, coesite, keatite, and stishovite. Except for stishovite, all these polymorphs are constructed from SiO_4 tetrahedra with minimal bond length and bond angle distortion. Each oxygen atom is shared between two tetrahedra, and in most cases, the O–Si–O angle is close to the ideal tetrahedral value of 109.5°. Under increased pressures, structural transitions have been observed within these polymorphs [91,92]. Stishovite is a high-pressure silica polymorph with a rutile structure, where each silicon atom is octahedrally coordinated by six oxygen atoms [93]. Phases denser than stishovite have also been proposed at higher pressures, exhibiting characteristics of high density, low compressibility, and a high elastic ratio [94].

Various theoretical methodologies have been employed to investigate the properties of SiO_2. Classical molecular dynamics (MD) simulations that take into account interatomic potentials formulated from both empirical [95] and non-empirical [91] methods were utilised to explore the impact of pressure on diverse silica polymorphs. Quantum mechanical calculations have also been conducted for studying silica polymorphs. The semi-empirical tight binding (TB) approach was used to explore defects in vitreous SiO_2 [96], while *ab initio* HF calculations were utilised for the examination of structural and electronic properties of various polymorphic forms of silica [97,98].

Studies leveraging DFT with varying XC functionals have gained considerable attention recently. Implementing the LDA, Ching *et al.* [99,100] employed the orthogonalised linear combinations of atomic orbitals (OLCAO) method for computing the electronic and optical properties of eight different silica polymorphs; Park *et al.* [94] investigated high-pressure phases of SiO2 using the full-potential linearised augmented

plane wave (FLAPW) method; Chelikowsky *et al.* [101,102] assessed structural properties of nine silica polymorphs with the pseudopotential method. For studies employing the GGA functional, Hamann [103] utilised the PW91 combined with pseudopotentials to examine the phase transition of SiO2; similar research was reported using the PBE form of GGA within the FLAPW method [104]. Most recently, a comprehensive study by Demuth *et al.* [105] explored the crystal structures of eight silica polytypes employing both LDA and GGA functionals within the framework of pseudopotential methods.

While local or semi-local functionals typically underestimate the band gap, methodologies such as the GW approximation [106,107] and related quasiparticle-based calculations can address these band gap discrepancies, though they necessitate considerable computational resources. Hybrid functionals like the HSE [76,107,108] have been proposed as solutions to the limitations of LDA and GGA functionals. Nevertheless, the default 25% fraction of HF exchange incorporated in the HSE functional has been found to be inadequate. This often necessitates an adjustment of the mixing parameter to align with experimental band gaps [108]. However, the treatment of the HF exchange fraction as a modifiable parameter during calculations may compromise the first principles character of these calculations.

An analogous range-separated hybrid functional is the sX functional [83]. In the sX method, the XC potential includes a non-local term that resembles the HF potential but incorporates correlation effects through screening of long-range exchange interactions. This is accomplished with an exponential decay factor dependent on electron separation. The functional also aims to preserve the exact XC energy for the HEG, which requires an additional local contribution to the total XC energy, parametrised using Perdew's expression for the LDA. The sX functional was designed for broad applicability across various systems without needing empirical adjustments. To date, no detailed investigation of silica polymorphs using the sX scheme has been conducted.

In this study, calculations on the nine most common polymorphs of SiO_2 are performed using both the traditional GGA and the advanced sX functional, aiming to provide an enhanced and systematic understanding of their structural, electronic, and optical properties. Excitons become increasingly important for poorly screened systems like silica and can be examined using Bethe–Salpeter methods [109]. However, the

focus here is on how the polymorphs might influence SiO_2 in devices, where the screening by their metallic systems counters most exciton effects.

3.2 Calculation Methods

The calculations employ the plane wave pseudopotential method within the CASTEP code [110], utilising norm-conserving pseudopotentials and a plane wave cutoff energy of 680 eV. Oxygen pseudopotentials are generated using the OPIUM method [111] to maintain a large band gap and minimise the introduction of anomalous features during the calculation. The band structures and optical properties of the silica polymorphs' primitive cell are studied using both the PBE version of the GGA and the sX hybrid functional. The calculations are converged to a residual energy difference of 10^{-6} eV and a force tolerance of 0.02 eV/A°.

This work considers nine silica polymorphs, including α- and β-quartz, α- and β-tridymite, α- and β-cristobalite, coesite, keatite, and stishovite. Some polymorph structures, such as β-tridymite, β-cristobalite, and coesite, remain controversial.

The idealised β-tridymite exhibits a space-group symmetry of $P6_3/mmc$ [112]. An alternative model with an orthorhombic structure ($Cc2m$) has the Si atoms located on the same position as the ideal hexagonal structure, but with split oxygen positions [113,114]. Due to the energy similarity of these two structures, the idealised form of β-tridymite is studied in this work.

The structure of β-cristobalite is subject to more scrutiny [115,116]. The majority of the literature uses the ideal cubic structure with space group $Fd3m$ due to its structural simplicity. However, the structure has the Si–O–Si bond angle (180°) and the Si–O bond length (1.54 Å) differing significantly from the values of other silica polytypes, which can be considered incorrect [116]. The present study focuses on the tetragonal structure with $I\bar{4}2d$ symmetry, which yields bond distances and angles in the correct range [116]. The ideal β-cristobalite is also studied as a comparison to other theoretical works.

The space-group symmetry of coesite has been described as either $C2/c$ [117] or $P2_1/c$ [118]. As the two structures are essentially energetically degenerate, the higher symmetry $C2/c$ is studied.

The atomic structures of the nine silica polymorphs are depicted in Fig. 3.1. Detailed structural information, including space group, lattice constant, volume per

formula unit (f. u.), and density, is summarised in Table 3.1. For all 4:2-coordinated SiO_2 systems, the connected network of Si–O bonds is believed to contain irregular rings of order $n \geq 4$, where n represents the number of Si atoms in a ring. The ring structures of the polymorphs, excluding the 6:3-coordinated stishovite, have been included, as the ring order was suspected to be one of the controlling parameters for the electronic and optical properties [119].

3.3 Results

3.3.1 Structural Properties of SiO_2

Quartz is one of the most studied silica polymorphs. It is thermodynamically stable under ambient conditions and throughout most conditions for rock crystallisation at the Earth's surface, making quartz the most abundant silica form. The low- and high-temperature forms of quartz, namely, α-quartz and β-quartz, are very similar in their structures. Both α-quartz and β-quartz possess hexagonal cells containing three SiO_2 molecular units. The Si–O bonds form a closed ring of six Si–O units. Their network structures are based on helical arrangements of SiO_4 tetrahedra, with the corners of neighbouring tetrahedrons spirally connected. In comparison to β-quartz, which has both sixfold and threefold screw-rotation axes, the α-quartz structure is a distorted version with only threefold screw-rotation axes. α-quartz has a rather regular and rigid network structure, serving as the paradigm for all silicate compounds.

The calculations predict the equilibrium volume of quartz with a slight overestimation of ~2%, compared to the experimental results [120,121]. The equilibrium volume is lower in the case of α-quartz. The total energy of α-quartz is also lower than that of β-quartz, with the structural energy difference being $\Delta E = 4.20$ meV per SiO_2 unit. These findings indicate that the thermodynamic driving potential for breaking the symmetry of β-quartz is stabilised at an expanded volume.

Tridymite and cristobalite are high-temperature, low-pressure polymorphs of silica. It is worth noting that the idealised structures of β-tridymite and β-cristobalite are closely related. Their basic building blocks consist of SiO_4 tetrahedra sheets arranged in hexagonal rings, with individual tetrahedra alternately pointing up and down. These sheets are packed in either a two-layer structure (β-tridymite) or a three-layer structure (β-cristobalite). Rings in successive sheets are directly superposed,

resulting in large open pipes formed perpendicular to the sheets and related by mirror symmetry. The α-tridymite and α-cristobalite are distortional derivatives of their idealised β-forms, occurring at lower temperatures and exhibiting lower symmetry. The relatively open tetrahedral framework of tridymite and cristobalite contributes to their lower density in comparison with quartz.

The calculations for the structural parameters of α-tridymite align with experimental findings [122]. However, the equilibrium volume for β-tridymite is 9.6% larger than the value deduced from the experimental lattice parameters (45.39 $Å^3/SiO_2$) [112]. On the other hand, calculated internal parameters, such as bond lengths and bond angles, are nearly identical to experimental values and essentially independent of volume. A similar overestimation of the equilibrium volume was previously observed for β-tridymite in a study using LDA [105].

Regarding the controversial β-cristobalite, the results indicate that the equilibrium volume of the $I\bar{4}2d$ phase is consistent with the experimental value (45.33 $Å^3/SiO_2$) [116], with an accuracy of 1%, while the value for the idealised $Fd3m$ phase is larger than the experiment by 9.8%. The $I\bar{4}2d$ structure also exhibits lower total energy and is energetically more favourable than the idealised structure.

Comparing the calculated results of β-tridymite and the idealised β-cristobalite, it is revealed that their equilibrium volume differs by only 0.01 $Å^3/SiO_2$, and the energy difference between the two polymorphs is within 2.5 meV/SiO_2. These findings further support their structural similarities.

The calculations of tridymite indicate that α-tridymite has a smaller equilibrium volume than the β-form and is lower in total energy, similar to the case of quartz. In contrast, α-cristobalite has a larger equilibrium volume than the tetragonal β-cristobalite, accompanied by a higher total energy. Overall, the formation energies of quartz, tridymite, and cristobalite are very close, with up to ~30 meV difference. The most stable form is β-cristobalite, and the least stable one is β-quartz.

Coesite is a metastable high-pressure polymorph of silica. Its structure contains four-membered rings of SiO_4 tetrahedra, which are aligned approximately parallel to (101) or (110) and further arranged into chains. Coesite has a relatively higher density than other 4:2-coordinated SiO_2. The calculated structural details are in accord with the available experimental data [117]. Additionally, coesite has a significantly higher total

energy (~100 meV) than that of α-quartz, confirming its metastable nature as a phase of SiO_2.

Keatite is a polymorph of silica with a density intermediate between those of α-cristobalite and β-quartz. It can be synthesised experimentally using the hydrothermal method. The crystallisation is favoured at moderate temperature (200–400 °C) and pressure (2–3 kbar) [123]. The natural occurrence of keatite has also been reported [124]. The keatite network consists of fivefold, sevenfold, and eightfold rings formed by corner-sharing SiO_4 tetrahedra. The calculated equilibrium volume of keatite differs from the experimental value by 2.3% [123] and the calculated unit cell energy is between that of α-quartz and β-quartz.

Stishovite is another rare polymorph formed at high pressure. Unlike the above-mentioned silica polymorphs, stishovite resembles the rutile (TiO_2) structure and is 6:3-coordinated: each Si atom is bonded to 6 O atoms in an octahedral configuration, and every O atom is coordinated by 3 Si atoms. The space group is $P4_2/mnm$, with the unit cell containing two SiO_2 molecular units. Stishovite is an extremely dense form of silica. The calculation predicts the equilibrium volume with a slight overestimation of 2.4% from the experiment [125]. The calculated density of stishovite is approximately 47% higher than that of coesite, which is the densest form of the 4:2-coordinated SiO_2 system. Importantly, there is a pronounced energy difference between stishovite and α-quartz (ΔE = 720 meV/SiO_2), which is nearly six times larger than the ΔE between coesite and α-quartz—it is appreciably less stable than other tetrahedrally coordinated SiO_2 polymorphs.

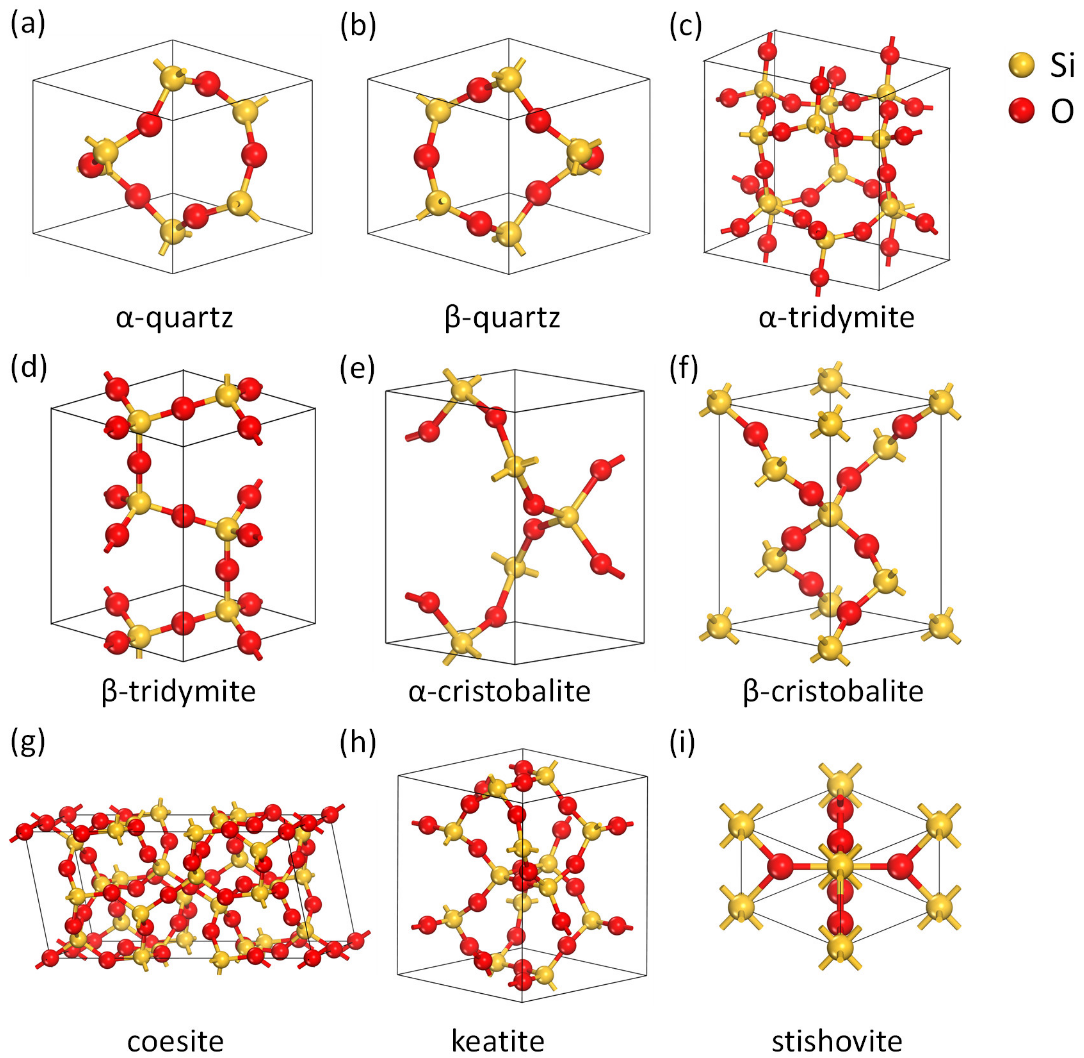


Figure 3.1 Atomic structures of (a) α-quartz, (b) β-quartz, (c) α-tridymite, (d) β-tridymite, (e) α-cristobalite, (f) β-cristobalite, (g) coesite, (h) keatite, and (i) stishovite in their conventional cells. The demonstrated cell of β-cristobalite in (f) is the tetragonal structure with $I\bar{4}2d$ *symmetry. Silicon and oxygen atoms are represented by yellow and red spheres, respectively.*

Table 3.1 Calculated structural information of nine SiO_2 polymorphs using the GGA functional. The space group, lattice constant, ring structure, volume per SiO_2 unit, and density are listed. Energy differences between various polymorphs and α-quartz are also summarised, with positive values indicating a higher systematic energy than α-quartz. Additionally, the band gap, static dielectric constant, and refractive index examined with both GGA and sX functionals are included.

Structural properties										
	Quartz		Tridymite		Cristobalite			Coesite	Keatite	Stishovite
	α-	β-	α-	β-	α-	β-				
Crystal structure	Hexagonal	Hexagonal	Orthorhombic	Hexagonal	Tetragonal	Tetragonal	Cubic	Monoclinic	Tetragonal	Tetragonal
Space group	$P3_121$	$P6_222$	$C222_1$	$P6_3/mmc$	$P4_12_12$	$I\bar{4}2d$	$Fd3m$	$C2/c$	$P4_32_12$	$P4_2/mnm$
Lattice constant (Å)	a = 4.94 c = 5.43	a = 5.03 c = 5.51	a = 8.84 b = 5.02 c = 8.31	a = 5.20 c = 8.49	a = 4.97 c = 6.92	a = 5.02 c = 7.30	a = 7.36	a = 7.14 b = 12.35 c = 7.16	a = 7.48 c = 8.77	a = 4.19 c = 2.65
Ring structure	6	6	6	6	6	6	6	4	5, 7, 8	…
f. u./cell	3	3	8	4	4	8	8	16	12	2
Volume (A^3/f. u.)	38.51	40.35	46.02	49.78	42.78	45.89	49.79	34.26	40.95	23.32
Density (g/cm^3)	2.61	2.47	2.17	2.00	2.33	2.17	2.00	2.91	2.44	4.28
ΔE (meV/f. u.)	0	4.20	−25.54	−3.35	−19.53	−33.43	-0.89	99.89	1.78	730.35

Electrical and optical properties											
		Quartz		Tridymite		Cristobalite			Coesite	Keatite	Stishovite
		α-	β-	α-	β-	α-	β-				
							$I\bar{4}2d$	$Fd3m$			
Band gap (eV)	GGA	5.98	6.01	5.75	5.77	5.84	5.90	5.57	5.96	5.91	5.52
	sX	9.17	9.08	8.86	8.78	8.97	9.02	8.82	9.16	9.06	8.67
$\varepsilon_1(0)$	GGA	2.24	2.18	2.03	1.91	2.13	2.02	1.95	2.41	2.18	3.55
	sX	1.68	1.64	1.56	1.50	1.62	1.48	1.54	1.75	1.65	2.43
Refractive index	GGA	1.50	1.48	1.42	1.38	1.46	1.42	1.40	1.55	1.48	1.88
	sX	1.30	1.28	1.25	1.23	1.27	1.21	1.24	1.32	1.28	1.56

3.3.2 Electronic Properties of SiO_2

The band structures for nine silica polymorphs calculated with different XC functionals are shown in Fig. 3.2 (GGA) and Fig. 3.3 (sX). The band structures are displayed along the high symmetry lines of the Brillouin zone, with a focus more on the bands with energies near the band gap region. The main features of the band structures are independent of the correction by hybrid functional. For all 4:2-coordinated SiO_2 systems, the occupied VBs consist of two parts separated by a sizable gap. The upper VB is primarily composed of the two non-bonding O 2p orbitals, and the lower VB consists of hybridised O 2p, Si 3s, and Si 3p states. The top of VB is very flat with a large hole effective mass. The semi-core bands located deep in energy correspond to the O 2s orbital. The bottom of the CBs is derived from significant admixtures of the antibonding orbitals O 2s, O 2p, Si 3s, and Si 3p, with the lowest CB exhibiting a strong dispersion. On the other hand, the 6:3-coordinated stishovite has a very different electronic structure. The most noticeable change is the vanishing of the pseudogap that separates the upper and lower VBs. Stishovite has a wide single VB that is derived from the hybridisation between O 2p and Si valence states. It also has a core-like band of predominately O 2s characteristics like the other tetrahedrally coordinated silicas.

The sX functional effectively rectifies the band gap underestimation exhibited by GGA, resulting in an approximate increase of 3 eV for all silica phases. Other band structure features also vary with different calculation schemes. For SiO_2 polymorphs with tetrahedral coordination, GGA calculates the upper VB to start from the valence band maximum (VBM), with an average bandwidth of 3.3 eV. The lower VB ranges from approximately −5 to −10 eV, with an average bandwidth of 4.6 eV. The semi-core bands are located at about −17 to −19 eV, with an average bandwidth of 2.2 eV. In the case of sX calculations, the top VBs correspond to lower energies and have increased bandwidth: the upper part of the VB locates from 0 eV to −3.6 eV on average; the lower VB ranges from approximately −5.5 to −11 eV with an average bandwidth of 5 eV. The semi-core bands are located deeper (−22 to −24 eV) with similar bandwidth as the GGA case. For the octahedrally coordinated stishovite, the sX scheme obtains a single VB 1.2 eV wider than the GGA. The sX semi-core band is found to be 4 eV lower in energy and slightly wider (~0.3 eV) than the GGA result. The detailed band structures obtained from the sX calculation are further elaborated below.

In the case of α-quartz, several theoretical studies have been performed: Chelikowsky and Schlüter [126] found an indirect band gap of 9.2 eV from M to Γ, while Ching and co-workers [99] found an indirect band gap of 8.8 eV from K to Γ. The current calculation reports an indirect band gap of 9.17 eV from M to Γ. Note that the top of the VB at K is only 0.001 eV below that of M. Between K and M, there is another high symmetry point A with a similar VBM energy (−0.096 eV). The difference between these high symmetry points is sufficiently small that their relative order may change. The sX result is in good agreement with the experimental band gap of 9.65 eV [127], considering that the band gap calculated with GGA is only 5.98 eV. The calculations also show that α-quartz has an upper VB with a 3.56 eV width and a wider lower VB of 5.35 eV, separated by a small pseudogap of 1.59 eV.

The sX calculation for β-quartz indicates an indirect band gap of 9.08 eV from A to Γ. The band structure of β-quartz is comparable to that of α-quartz as a result of their closely related crystal structures. Their upper VBs have a similar width, while β-quartz has a wider lower VB of 5.49 eV, which is also the widest among all the 4:2-coordinated SiO_2.

The β-tridymite also has a hexagonal cell, similar to the case of quartz, but with four molecular units per cell and a lower density. Its band structure is, therefore, quite different from quartz. According to the sX result, β-tridymite has an indirect band gap of 8.78 eV from M to Γ. It has the smallest band gap value compared to other polymorphs discussed in this work. The upper and lower VBs of β-tridymite are separated by a more pronounced pseudogap of 2.61 eV.

The other tetrahedrally coordinated silicas all have direct band gaps at Γ. Notably, the band structures of coesite and keatite are distinctively more complicated than those of other SiO_2 polycrystals. This can be attributed to their complex crystal structures: coesite has a low-symmetry monoclinic cell containing 16 SiO_2 units, and keatite has a tetragonal cell consisting of 12 SiO_2 units. The upper and lower VBs of coesite and keatite have roughly equal widths, while in the case of other silica polymorphs, the lower VB is significantly wider.

The particular case of stishovite with a different coordinate system has a direct band gap of 8.67 eV at Γ, which is the smallest of all SiO_2 polymorphs. It has a single VB of 11.9 eV in width, and the top of the VB has a comparatively small effective mass, indicating more covalent bonding compared to the other tetrahedrally coordinated

phases of silica. Stishovite also has a wide core-like band of 4.8 eV located at approximately −24.5 eV, which is almost twice the width of the other polymorphs.

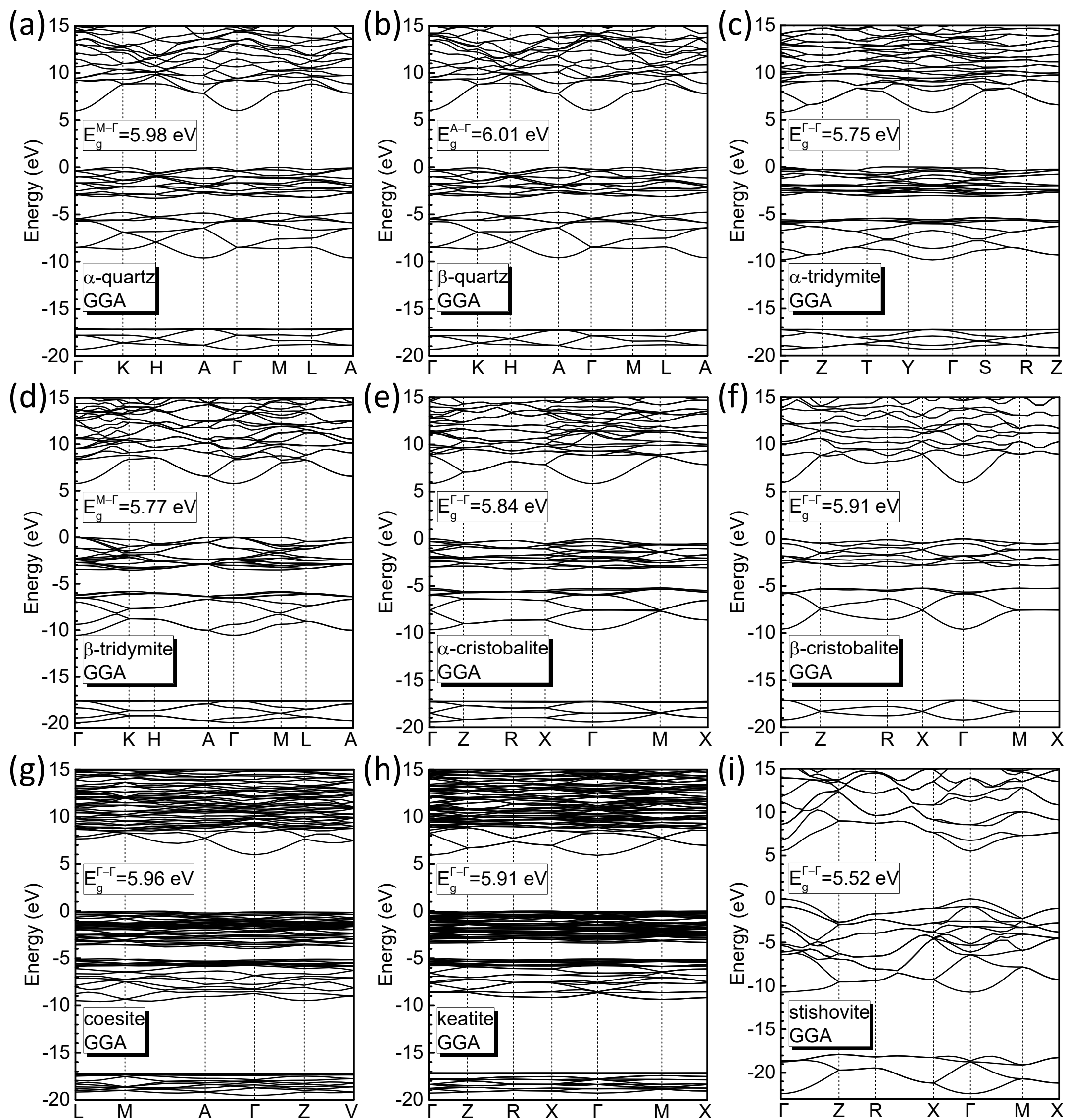


Figure 3.2 Calculated band structures along the high symmetry lines of the Brillouin zone for (a) α-quartz, (b) β-quartz, (c) α-tridymite, (d) β-tridymite, (e) α-cristobalite, (f) β-cristobalite, (g) coesite, (h) keatite, and (i) stishovite using the GGA functional. The VBM is set at energy zero. The band structure of β-cristobalite shown in (f) has $I\bar{4}2d$ *symmetry.*

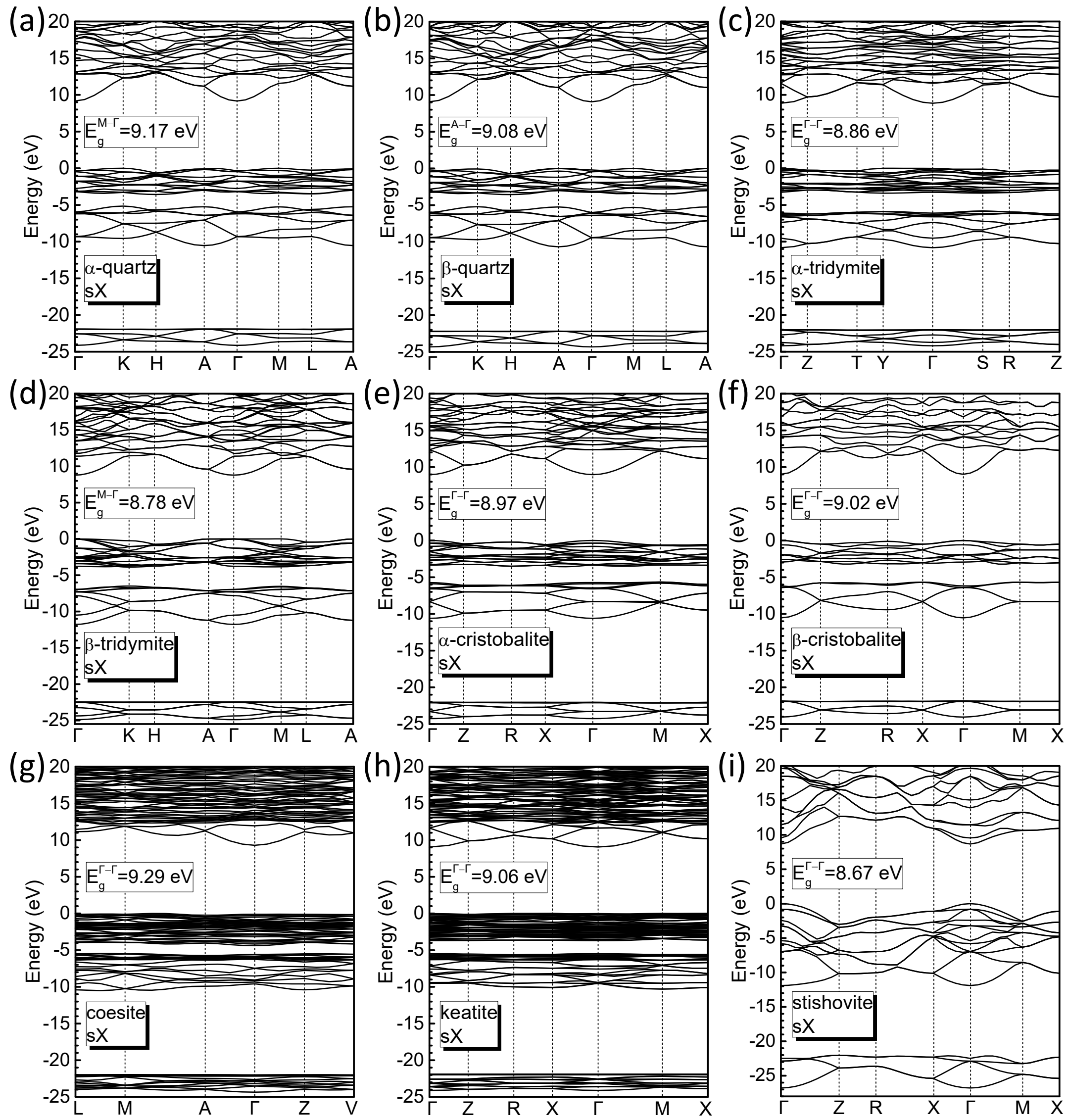


Figure 3.3 Calculated band structures along the high symmetry lines of the Brillouin zone for (a) α-quartz, (b) β-quartz, (c) α-tridymite, (d) β-tridymite, (e) α-cristobalite, (f) β-cristobalite, (g) coesite, (h) keatite, and (i) stishovite using the sX functional. The VBM is set at energy zero. The band structure of β-cristobalite shown in (f) has $I\bar{4}2d$ *symmetry.*

3.3.3 Optical Properties of SiO_2

Despite numerous experimental and theoretical studies on the structural and electronic properties of the silica system, systematic investigations into their optical properties are still limited. In this work, the optical properties of the nine silica polymorphs are reported in the form of the frequency-dependent dielectric function. These dielectric functions serve as remarkably good indicators of how materials respond to external electric fields. The dielectric constant shares a close relationship with optical characteristics such as refractive index and optical loss. Furthermore, it can be associated with other properties of the material, such as polarisability and density, as demonstrated by the Clausius–Mossotti relation [128].

The complex function is defined as $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$, where ω is the frequency of light. The imaginary part $\varepsilon_2(\omega)$ is deduced by directly evaluating the momentum matrix elements between the occupied and unoccupied electronic states. It represents the dissipation of energy and is associated with the absorption spectrum. The real part $\varepsilon_1(\omega)$ is linked to the imaginary part by a Kramers–Kronig transform and indicates the ability of the material to respond to an applied electric field through polarisation. The zero-frequency limit of $\varepsilon_1(\omega)$ yields the static dielectric constant $\varepsilon_1(0)$, which is another physical quantity of fundamental importance. The static dielectric constant indicates the maximum possible dielectric response to an applied field.

Some of the silica polymorphs considered in this work are uniaxial anisotropic crystals of high symmetry; therefore, the optical properties will also display some anisotropy. For the non-cubic birefringent crystal structures, the dielectric constant can be divided into two components: one parallel and the other perpendicular to the z-axis. A previous theoretical study found that the difference between these two components is generally small for some silica polymorphs [129]. Therefore, in this work, the focus is on the z-direction parallel component of the dielectric constant by taking the [001] polarisation of the electromagnetic field into account.

The real and imaginary parts of the dielectric functions calculated for different silica polymorphs are represented in Fig. 3.4 (GGA) and Fig. 3.5 (sX) as functions of photon energy. All phases of silica have the $\varepsilon_2(\omega)$ spectra located near the UV region, with negligible absorption at low energies, consistent with their insulating nature. The imaginary parts of the dielectric functions exhibit an onset, corresponding to the first

allowed direct transition at the fundamental band gap. The formation of peaks is due to interband transitions from VB to CB. The width of the non-zero part of $\varepsilon_2(\omega)$ is directly linked to the width of the CB: a more extended absorption spectrum is associated with a wider CB. On the other hand, the $\varepsilon_1(\omega)$ curves start with low static dielectric constants (approximately 2–4), indicating that SiO_2 polymorphs have very few available polarisation mechanisms. The real part of the dielectric constant gradually increases with the increase of the photon energy and reaches its maximum. With the continuous increase of photon energy, $\varepsilon_1(\omega)$ starts to decrease and undergoes several oscillations, each of which relates to an absorption peak in the corresponding $\varepsilon_2(\omega)$ spectra.

For the 4:2-coordinated SiO_2, the imaginary part $\varepsilon_2(\omega)$ has two main broad absorption regions, followed by some relatively small peaks, which are due to the two separated VBs. The two forms of quartz have closely related crystal and electronic structures; therefore, both the position and the amplitude of the dominant peaks of their spectra are greatly matched. Similarities of dielectric curves are also found between tridymite and cristobalite, as well as coesite and keatite since they share alike electronic features. The spectra of the 6:3-coordinated stishovite are distinctively different from the other polymorphs. The most notable feature is the relatively high amplitude of both real and imaginary parts of the dielectric constant. Stishovite has a tighter and sharper absorption band, which is compatible with its wide single VB. Besides, the first dominant peak of $\varepsilon_1(\omega)$ shifts toward a higher energy region compared to the other silica phases.

The shape of the spectra remains almost unchanged with different functionals. Two main effects of the hybrid scheme in the dielectric function can be observed: the amplitude of the spectra is appreciably reduced; the curves are shifted toward higher energies. In particular, the static dielectric constant $\varepsilon_1(0)$ is lowered by an average of 0.53 for tetrahedrally coordinated phases, and the threshold energies of $\varepsilon_2(\omega)$ are increased by ~3 eV, with a broadening of the absorption spectra. The $\varepsilon_1(0)$ difference is more pronounced in the case of stishovite, with a decrease of 1.1, alongside a similar $\varepsilon_2(\omega)$ energy shift as the other polymorphs. These findings are a consequence of the sX functional correcting the band gap underestimation by the GGA functional.

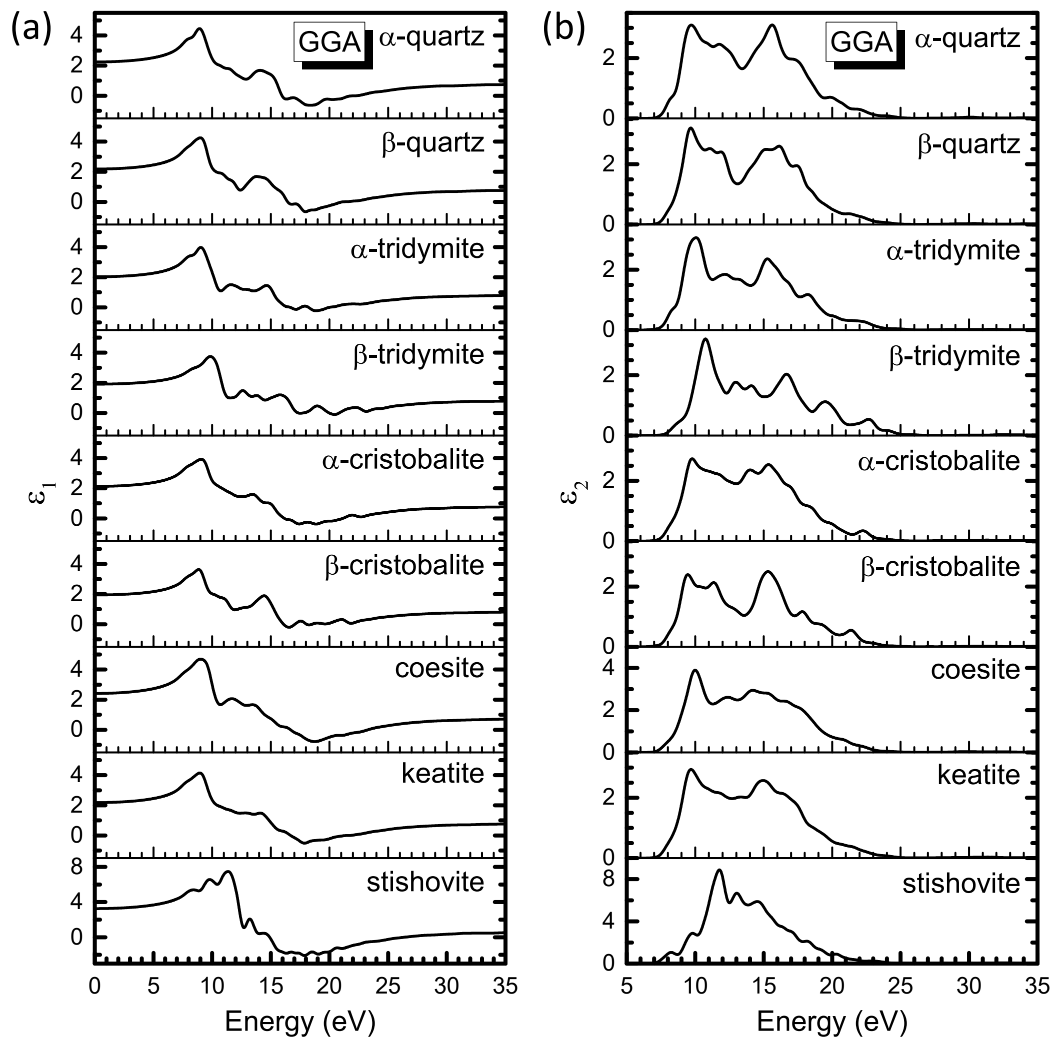


Figure 3.4 Calculated frequency-dependent dielectric functions of nine polymorphs using the GGA functional. The real part of the dielectric functions $\varepsilon_1(\omega)$ is summarised in (a), and the imaginary part $\varepsilon_2(\omega)$ is summarised in (b). The dielectric functions of β-cristobalite correspond to the $I\bar{4}2d$ symmetry.

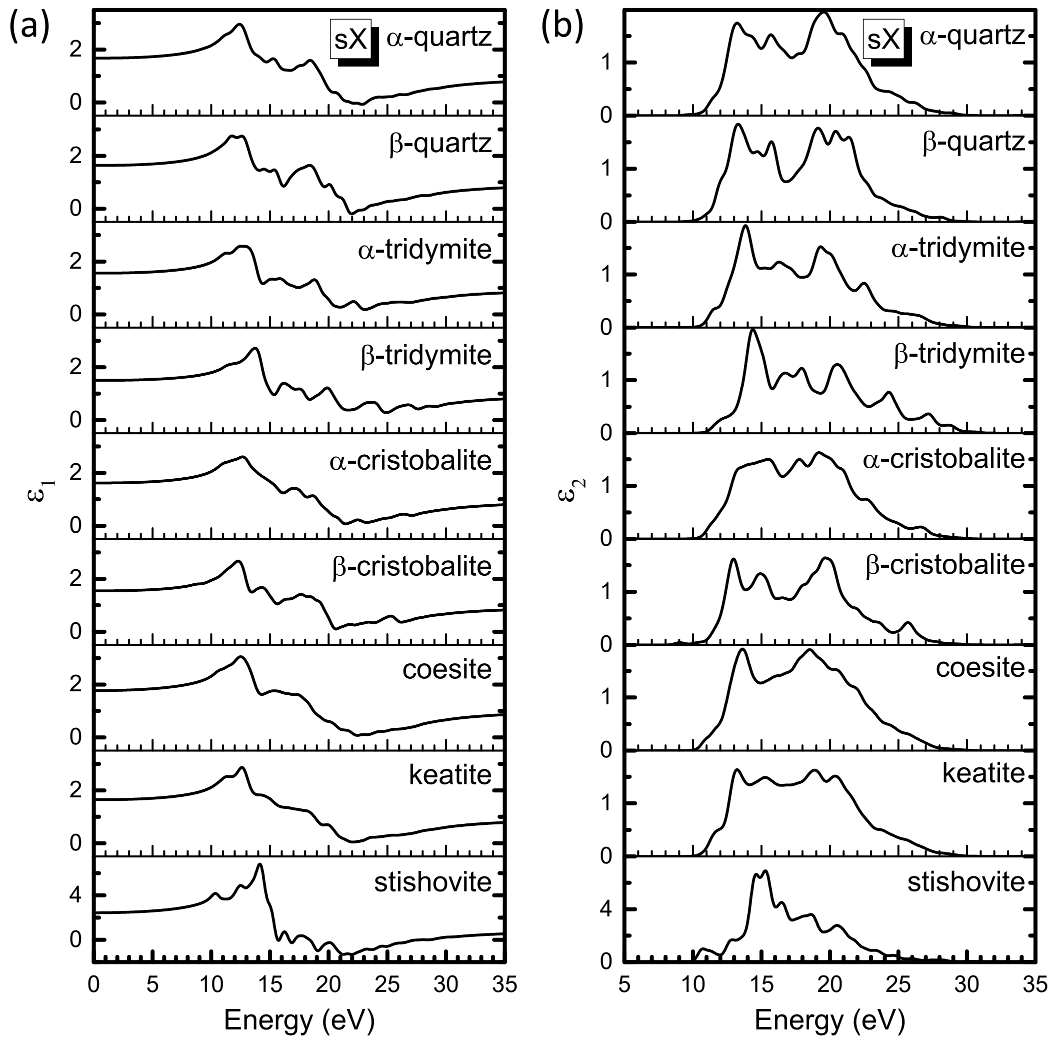


Figure 3.5 Calculated frequency-dependent dielectric functions of nine polymorphs using the sX functional. The real part of the dielectric functions $\varepsilon_1(\omega)$ is summarised in (a), and the imaginary part $\varepsilon_2(\omega)$ is summarised in (b). The dielectric functions of β-cristobalite correspond to the $I\bar{4}2d$ symmetry.

3.4 Discussion

The simulated results for the structural, electronic, and optical properties of nine SiO_2 polycrystals have been presented. Since the chemical composition of polymorphs is identical, the variety of their physical and optical properties is dominated by the network organisation. The calculation has revealed a significant correlation between the band gap E_g, the static dielectric constant $\varepsilon_1(0)$, and the corresponding silica density.

In Figs. 3.6(a) and 3.6(c), the E_g values calculated with different functionals are plotted against the relative volume V/V_0 of polymorphs, where the crystal volume of α-quartz is set as the standard V_0. There is a clear inverse relationship between the band gap and volume. Only stishovite shows the opposite dependence. It is worth noting that the two structures of β-cristobalite have distinct band gaps: the cubic structure with higher equilibrium volume shows a remarked E_g drop of 0.34 eV (GGA) and 0.39 eV (sX) compared to the tetragonal crystal. These results further support the strong relationship between the band gap and the atomic network. The application of hybrid functional yields an increased slope value, accompanied by an upward shift of the band gap.

The unique electronic behaviours of stishovite, compared to other 4:2-coordinated silica polymorphs, can be attributed to the higher coordination number of Si in stishovite. The increase in coordination from four to six allows the Si 3d orbitals to become involved in bonding, which results in a profound alteration of the electronic structure and results in a reduced band gap relative to other silica polytypes. The involvement of additional orbitals in the bonding process can also account for the distinction of the pseudogap in the VB, and concurrently, the broadening of the VB in stishovite.

For silica polymorphs following a 4:2-coordination system, the correlation between the band gap and the equilibrium volume can be attributed to enhanced interaction and overlap of the orbitals as the materials become more close-packed. An enhanced overlap tends to strengthen the bonding, thereby reducing the energy of the bonding states (i.e., the VB). This process implies a greater separation between the VB and the CB, thereby contributing to an overall expansion of the band gap.

Furthermore, there is an apparent linear dependence between the static dielectric constant and density. The results are shown in Figs. 3.6(b) (GGA) and 3.6(d) (sX). The $\varepsilon_1(0)$ values of different polymorphs exhibit less scattering with the varied volume—

there is a clearer linear correlation compared to the case of the band gap. Stishovite is omitted from the correlation analysis and lies far above the fitting line. The calculation again reveals a noticeable difference between the static dielectric constants of cubic and tetragonal β-cristobalite. The calculation based on the sX functional demonstrates a downward shift of the $\varepsilon_1(0)$ with a smaller slope.

It is worth noting that the traditional GGA method, although suffering from the band gap problem, can adequately describe the optical properties of SiO_2. On the other hand, while the hybrid functional can effectively increase the band gap for all silica polymorphs and reproduce experimental results, it tends to underestimate the dielectric constants, deviating from the experiment.

Generally, the DFT calculation with the semi-local GGA functional tends to overestimate the dielectric function, particularly for insulators [130]. However, silica polymorphs have been observed as an exception. Previous theoretical works calculated the high-frequency dielectric constant of α-quartz with different GGAs and found the value to be slightly underestimated [131,132]. Ramos *et al.* [133] studied the optical properties of β-cristobalite and found the dielectric function to reproduce the experimental value in the framework of LDA. They also reported a decrease in the dielectric constant when quasiparticle effects were included.

The calculated results of this work align with observations in earlier studies: the calculated dielectric constants using the GGA functional are in reasonable agreement with experiments, accompanied by slight underestimation; the hybrid functional predicts a higher band gap and inevitably worsens the agreement with the experiment, leading to underestimated dielectric constants by 20–30% with respect to the GGA values.

The findings of this work are influenced by the constraints of the current formalism. The GGA functional employed in this work does not take the excitonic effects into account, i.e., it ignores the fact that there is a virtual hole and electron interaction. However, silica as a wide band gap insulator exhibits weak electronic screening and large excitonic binding energy, which potentially makes the exciton effect significant. The exclusion of excitonic effects and the self-interaction error by the GGA functional systematically cancel each other [134], leading to the calculation of macroscopic optical dielectric constants that align closely with experimental reference values.

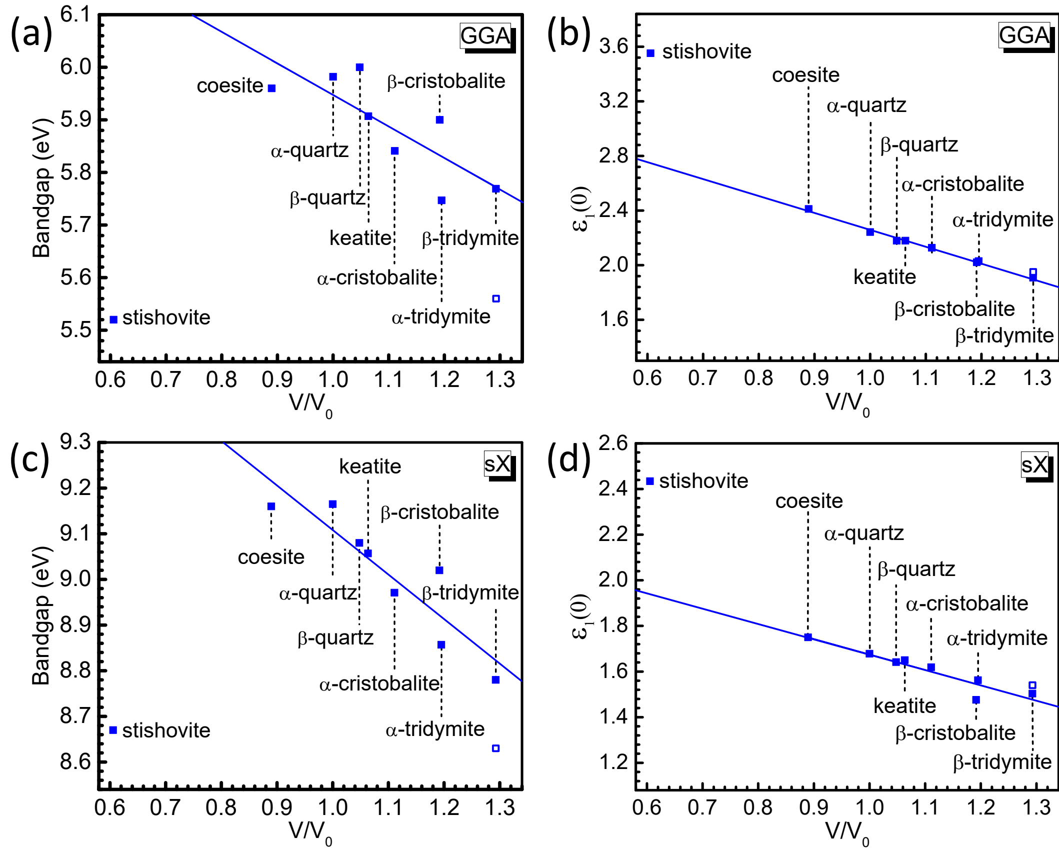


Figure 3.6 Band gap [(a) and (c)] and static dielectric constant $\varepsilon_1(0)$ [(b) and (d)] calculated using different functionals vs the relative unit cell volume V/V_0, where V_0 is the crystal volume of α-quartz. The less stable ideal cubic β-cristobalite is plotted as hollow squares for comparison. The special case of stishovite is excluded from the fitting.

To further examine the role of density in varying the optical properties of SiO_2 polymorphs, the refractive index is calculated, which is one of the most accessible optical parameters in experiments. Maj [135] derived a relationship between the mean refractive index and density for SiO_2 polymorphs, based on the dispersion formula from the Lorentz electron theory. Maj used empirical data and observed that SiO_2 polymorphs obey the linear Gladstone–Dale law. The specific refraction, i.e., the slope of the refractive index vs density plot, is calculated to be 0.2 cm^3/g. The refractive index calculated in this manner aligns remarkably well with experimental measurements, showing a negligible discrepancy of less than 0.01. In Fig. 3.7, the refractive index for the 4:2-coordinated silica is plotted against the corresponding density, along with the available experimental results for comparison. As discussed in Sec. 3.3.3, GGA yields a more accurate analysis of the optical properties of silica; therefore, the focus is on results obtained with this functional. The main observation from Fig. 3.7 is a strong linear relationship between the refractive index and the density, with a slope of 0.18 cm^3/g. This behaviour agrees with the prediction of the Gladstone–Dale law.

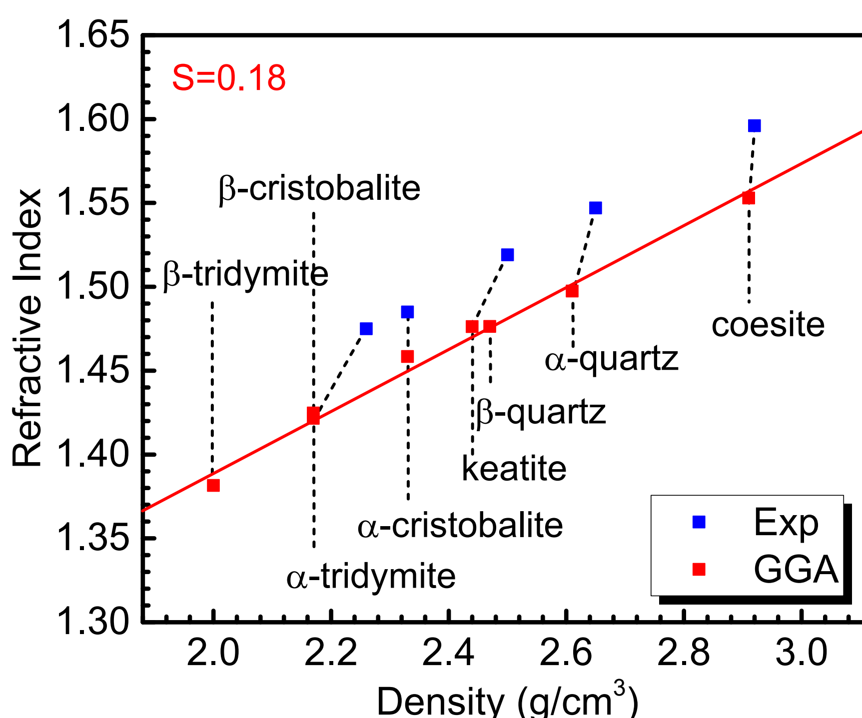


Figure 3.7 Refractive index vs density for 4:2-coordinated SiO_2 polymorphs calculated using the GGA functional. Experimental values [135] are plotted for comparison.

3.5 Conclusions

This study investigates the properties of nine different silica polymorphs within the framework of DFT, analysing the underlying correlation among their crystal structures, electronic properties, and optical properties. Additionally, the study assesses the validity of various XC functionals in predicting the properties of different silica phases. It is found that the band gap values and the static dielectric constants follow an inverse linear relationship with the unit cell volume, except for the high-density stishovite with a different coordination system. A linear correlation between the refractive index and density for the 4:2-coordinated SiO_2 is also established. Calculations using different functionals reveal significant deviations in the simulated properties. While the traditional GGA scheme suffers from band gap underestimation, it provides valuable guidance on the optical properties of silica polymorphs. In contrast, while the sX functional accurately predicts the band gap values, it tends to underestimate the dielectric constant. This study aims to provide essential complements to the systematic theoretical studies of silica polymorphs with different calculation schemes.

CHAPTER 4: SCHOTTKY BARRIER HEIGHTS OF DEFECT-FREE METAL–ZNO, CDO, MGO AND SRO INTERFACES

The SBHs of defect-free interfaces of ZnO, CdO, MgO, and SrO with various metals and different terminations have been investigated using density functional supercell calculations. A U-type treatment is applied to the metal d and O p states of the oxides where necessary to correct the density functional band gap error. The p-type SBHs are found to decrease linearly with increasing metal work function. For each oxide, the pinning factors for non-polar and polar interfaces are similar. S values are 0.26, 0.56, 0.74, and 0.96 for CdO, ZnO, MgO, and SrO, respectively. The calculated pinning factors are generally consistent with the MIGS model in terms of variation in ionicity and dielectric constant. For ZnO, MgO, and SrO, a significant shift of SBHs from the non-polar to the polar interfaces is observed, with values of 0.4 eV, 1 eV, and 0.5 eV, respectively, which can be explained by the formation of the interfacial dipole. The results also provide useful insights for describing CoFe–MgO interfaces in magnetic tunnel junctions (MTJs).

4.1 Introduction

Metal–oxide contacts have attracted substantial attention in numerous studies due to their importance in various applications, including light emission [136,137], catalysis [138], field-effect transistors [139], and MTJs [140–142]. One primary concern when using oxides in modern microelectronic devices is the contact resistance, which can be tuned by manipulating the interface SBHs [143]. Consequently, understanding the formation mechanism of the SBH is crucial.

The magnitude of the SBH has traditionally been determined by examining its dependence on the metal work function. Weakly interacting surfaces follow the Schottky–Mott rule, where the SBH is the difference between the EA of semiconductor and the work function of metal. For stronger interaction, charge transfer between the metal Fermi level and the states induced in the semiconductor band gap results in a FLP

effect. The SBH can be expressed using Eq. (1.4), with the degree of pinning defined by the pinning factor S, as represented in Eq. (1.5). The S factor varies between 0 (for a strongly pinned interface, i.e., the Bardeen limit) and 1 (for a non-pinning interface, i.e., the Schottky limit).

Numerous models have been proposed to explain the pinning mechanism and to identify the parameters that influence the value of the SBH. Currently, the magnitude of SBH is believed to depend on both intrinsic factors like the MIGS [36–40] and extrinsic factors like interfacial point defects [40]. The MIGS is believed to provide a good intrinsic model to describe the pinning effect at various metal–oxide contacts. In this case, Mönch [38,39] found that S followed the empirical formula as expressed in Eq (1.6). It is worth noting that for semiconductors with narrow or moderate band gaps, the contribution of MIGS is more pronounced, as MIGS are experientially shown to have longer decay lengths in these materials [39,144]. In wide band gap materials, MIGS still contribute to the pinning factor, but other factors, such as interfacial dipoles and extrinsic defects, may become more prominent in determining the pinning effect at interfaces.

In the conventional MIGS model, gap states originate exclusively from the VB and CB states of the semiconductor, making them intrinsic in nature. Consequently, the pinning factor should depend solely on the properties of the semiconductor, rather than the characteristics of the contact metals or the contacting orientations. However, the MIGS model does not fully explain the behaviours of compound metals. A study by Nishimura *et al.* [145] observed that metallic germanides can mitigate Fermi level pinning (FLP), while also demonstrating a distinct substrate orientation-dependent SBH on Germanium. This contrasts with the elemental metals where FLP is nearly perfect with marginal orientation dependence.

Similar findings have also been reported for silicide contacts on silicon [146]. The authors attribute Fermi level depinning to defect gap states. Notably, the interfaces between metal silicides and Si inherently exhibit a covalent lattice structure. This leads to the creation of mis-coordinated defect sites at the interface that generate additional localised interface states. These states add to the density of MIGS and drive the Fermi level across the band gap, leading to an increased slope parameter S. These defect states have been proposed as a necessary extension to the MIGS model, establishing a more discernible dependency of the SBH on both the selected metal and the interface orientation.

The SBH can also be influenced by the type of termination in the case of the polar surface of a compound semiconductor, which can be either anion-terminated or cation-terminated. Observations have indicated that the SBH depends on surface termination at interfaces such as Al–X junctions (X = Ge, GaAs, AlAs, and ZnSe) [147], and metal–HfO_2 interfaces [148]. The difference in SBHs can be attributed to the surface charges on the semiconductor side and the induced image charges on the metal side, which establish a potential difference that varies the SBH. Understanding the termination effects is essential for comprehending the interfacial electronic structure and for engineering interfaces with specific properties, such as reduced contact resistance or enhanced rectification ratios.

To advance the knowledge of interfaces and enable the development of more efficient and reliable electronic devices, it is crucial to uncover the underlying microscopic mechanisms responsible for the chemical trends in SBHs. The investigation focuses on studying various metals in contact with selected metal oxides with different terminations, to thoroughly examine the applicability of the MIGS model and the influence of terminations on the electronic properties of metal–oxide contacts.

It is essential to carefully select representative materials that consider these factors while maintaining industrial relevance. The optical dielectric constant is a critical parameter in the MIGS model, affecting the pinning factor, while the ionicity of a material significantly impacts charge separation at the interface and the formation of distinct dipole layers. With these considerations in mind, four oxides (ZnO, CdO, MgO, and SrO) have been chosen, covering a wide range of optical dielectric constants and ionicities. This selection allows for an in-depth evaluation of the MIGS model and its effectiveness in various contexts, as well as the investigation of the impact of different terminations and ionicities on the formation of interface dipoles.

Commentary on specific applications related to the chosen oxides is provided. ZnO demonstrates a wide array of fascinating physical characteristics, such as high optical transparency and adjustable electronic properties, making it suitable for a multitude of applications, particularly in optoelectronic devices. The scope of potential applications broadens more when ZnO is combined with other Group II oxides like MgO and CdO to form alloys and heterostructures [137,149]. MgO is of particular importance as an insulator in MTJs [140–142]. A critical aspect of MTJs used in spin-transfer torque magnetic random-access memory (STT-MRAM) is that certain properties, including the perpendicular magnetic anisotropy (PMA), might be affected

by the atomic nature of the magnetic metal–oxide interface [150,151]. As a result, further exploration of these metal–oxide interfaces is greatly important.

Although previous investigations have explored some of these interfaces [152–156], a comprehensive theoretical analysis of the SBH trends, taking into account the aforementioned factors, has not been conducted to the best of current knowledge. The present study aims to fill this gap by considering a broader selection of metals and oxides, as well as assessing various factors that influence the SBH trends.

4.2 Calculation Methods

The calculations in this study employ the CASTEP code [110], utilising the plane wave pseudopotential method and the PBE version of the GGA for the XC functional. Norm-conserving pseudopotentials with a plane wave cutoff energy of 680 eV are implemented. Oxygen pseudopotentials are generated using the OPIUM method [111].

The SBHs are calculated using slab supercells of five layers of metal, seven layers of oxide, and 15 Å of vacuum. This configuration ensures that the properties of the central region of the slabs mimic those of the corresponding infinite bulk materials, effectively representing the metal–oxide interface. The thickness of the vacuum layer is sufficient to prevent interactions between the slab and its periodic images. The chosen thicknesses for the metals, the oxides, and the vacuum are a result of a careful balance between the computational efficiency and accuracy of the results. Additionally, the oxide–vacuum surfaces are hydrogen-terminated to passivate any dangling bonds and reduce spurious surface effects, with the H–O bonds allowed to relax laterally.

For ZnO, the zincblende structure is employed to avoid the internal fields of the wurtzite phase and to facilitate lattice matching with cubic metals. The equilibrium rock salt structure is studied for CdO, MgO, and SrO. Metals are adopted in their face-centred cubic (FCC) structure, except for Fe, which is used in its body-centred cubic (BCC) structure, to achieve a smaller lattice mismatch with a reduced model size. Since metal work function is not highly sensitive to lattice constant, metals are strained to match the surface slabs of oxides when necessary. The lattice mismatch between metal and oxide for all interface structures remains within 5%. Lattice matching of the metal–oxide supercells is listed in Table 4.1. Multiple interface models with various initial metal and oxide positions are designed to search for the global minimum structure.

Geometries are relaxed to yield energy differences below 5×10^{-6} eV/atom and forces below 0.01 eV/Å.

SBH values are extracted using the core-level method [157,158] to enhance the precision in determining the VBM, using Eq. (4.1), where E_{core}^{int} represents the core-level state in the interfacial model and ΔV is the energy difference between the core-level state and VBM of the bulk crystalline oxide. The semi-core 2s state of the O atom away from the interface region is selected as the reference core-level state,

$$\phi_p = E_{core}^{int} + \Delta V - E_F. \quad (4.1)$$

Table 4.1 Lattice matching of various metal–oxide interfaces. The notation a × a = b × b implies that an a × a sized supercell of the metal surface is matched with a b × b sized supercell of the oxide surface.

Oxide	Metal	Matching	
		Non-polar	Polar
CdO	Ru, Os, Re, Rh, Ir, Pt	$\sqrt{8} \times \sqrt{8} = \sqrt{5} \times \sqrt{5}$	…
	MoO_3	$2 \times 2 = \sqrt{5} \times \sqrt{5}$	…
ZnO	Ti, Ru, Os, Re, Rh, Pd, Pt	$2 \times 2 = \sqrt{3} \times \sqrt{3}$	$2 \times 2 = 1 \times 1$
	Ni	$\sqrt{7} \times \sqrt{7} = 2 \times 2$	$\sqrt{5} \times \sqrt{5} = 1 \times 1$
	MoO_3	$\sqrt{3} \times \sqrt{3} = 2 \times 2$	$\sqrt{2} \times \sqrt{2} = 1 \times 1$
MgO	Sc, Hf, Zr, Ti, W, Ru, Re, Rh, Pd, Ir, Pt, Co, Fe	$1 \times 1 = 1 \times 1$	$1 \times 1 = 1 \times 1$
	Cr	$\sqrt{5} \times \sqrt{5} = 2 \times 2$	$2 \times 2 = \sqrt{3} \times \sqrt{3}$
	MoO_3	$\sqrt{5} \times \sqrt{5} = \sqrt{8} \times \sqrt{8}$	$2 \times 2 = \sqrt{7} \times \sqrt{7}$
SrO	Sc, Hf, Zr, Ti	$\sqrt{5} \times \sqrt{5} = 2 \times 2$	$2 \times 2 = \sqrt{3} \times \sqrt{3}$
	Ru, Rh, Pd, Ir, Pt	$\sqrt{2} \times \sqrt{2} = 1 \times 1$	$\sqrt{7} \times \sqrt{7} = 2 \times 2$
	MoO_3	$1 \times 1 = 1 \times 1$	$1 \times 1 = 1 \times 1$

The GGA functional is commonly employed for calculating electronic structures. However, it tends to underestimate the band gap of semiconductors and insulators. This issue is particularly pronounced in ZnO and CdO, with a calculated band gap of only ~0.8 eV for ZnO and a negative indirect band gap for CdO [83,159–164].

More accurate band structure calculations can be achieved with the GW and similar methods [106,160,161], but these approaches are computationally intensive. Alternative options include hybrid functional methods, such as HSE [76,162] and sX [83,163,164] functionals, which are more costly than GGA but not as computationally demanding as GW. However, HF calculations can exhibit a divergence in the derivative of the orbital energies for metallic systems, leading to inaccurate descriptions of these systems [76,165]. Furthermore, while hybrid functionals are more efficient compared to the GW method, they can still be computationally expensive for large supercells, which are typically required for constructing metal–oxide interfaces with desired lattice matching.

The GGA + U scheme is proposed as an efficient approach to correct the band gap error and handle large SBH supercells. The U parameter was initially introduced for open-shell transition metals to describe the effect of electron–electron repulsion within the metal d states. This method was extended to closed-shell transition metal compounds with filled cation d states, where a U potential on the d states would increase the band gap [159,161]. However, applying the correction only to the metal d orbitals requires an unphysically large U value to align the calculated band gap with experimental results. In this work, a moderate U combination is applied to both anion p states and cation d states of ZnO and CdO to correct the overestimated p–d coupling and overcome the band gap underestimation. For MgO and SrO, it is demonstrated that a U term can be applied to the O 2p states to widen their band gaps, even without the involvement of any d states. This more general application of U terms falls within the Agapito–Curtarolo–Buongiorno Nardelli (ACBN0) pseudo-hybrid Hubbard density functional [166], providing a cost-effective and accurate approach to study these systems.

4.3 Results and Discussion

4.3.1 Band Structures of Oxides

To accurately describe interfacial SBHs, it is crucial to precisely define the band structures of oxides. In particular, the band gap underestimation by traditional DFT methods must be corrected to reflect experimental band gap values. Band structures of oxides have been calculated using GGA, GGA + U, and sX methods, with results depicted in Figs. 4.1 and 4.2. Additionally, EAs are determined using non-polar surfaces of supercells with vacuum spacing to obtain absolute band edge energies relative to the vacuum level. Detailed band energies are summarised in Table 4.2 and compared to available experimental results. In the following discussion, the emphasis is placed on examining the GGA + U method. The calibre of electronic properties computed using this method is contrasted against the benchmark set by sX calculations. Reference to relevant theoretical work and experimental studies is incorporated for a comprehensive comparison.

The band structure of ZnO has been the subject of numerous studies. Previous results using the sX method [83,163,164] align closely with those using the quasiparticle calculations [160,161] and match the experimental band gap of 3.4 eV, whereas the HSE functional requires an increased HF fraction α from 25% to 37% to reproduce the result [162]. The details of the band structure obtained by the sX calculations in this work are consistent with prior works using the same functional [83,164], as well as GW calculations [161]. As for the GGA + U scheme, U = 4 eV is applied to the Zn 3d states and U = 9 eV to the O 2p states. This combination of U values effectively increases the band gap to 3.40 eV. Despite the general similarity between the band structures obtained by sX and GGA + U in the current work, it is worth noting that the GGA + U scheme yields a narrower separation between the upper and lower sections of the VB as compared to the sX results, as shown in the left panel of Figs. 4.1(b) and 4.1(c). This difference can be attributed to the on-site Coulomb interaction. The inclusion of the U term, which represents a repulsive interaction, leads to a greater delocalisation of electrons. As a result, the upper VB (associated with O 2p) and lower VB (associated with Zn 3d) become more dispersive, leading to a reduced gap between the upper and lower VBs.

Previous research on CdO band structure has yielded diverse results [83,159,161,163,164,167–169]. The present work, employing sX calculation, shows an indirect L–Γ gap of 0.55 eV, and a direct gap of 2.40 eV at Γ. The band structure details align closely with the previous sX calculation [83]. On the other hand, to address the unusual negative band gap predicted by the GGA functional, U = 2 eV is applied to the Cd 4d states and U = 4 eV to the O 2p states. This results in an indirect L–Γ gap of 1.17 eV, and a direct gap of 2.27 eV at Γ. The calculated band gap value is consistent with experimental values [167,168]. Additionally, the band structure details derived from the GGA + U calculation show similarity to those derived from the HSE03 + GW approach [161], particularly in terms of the gap between the upper and lower VB, which is less pronounced in the sX result [Figs. 4.1(b) and 4.1(c)]. Notably, the lower VB which corresponds to the Cd 4d orbital does not exhibit a significant downward shift as in the case of ZnO. This observation could potentially be attributed to the lesser degree of interaction among electrons in 4d orbitals compared to those in 3d orbitals. Therefore, the electron–electron interaction introduced by U generally exerts a reduced influence on the electronic structure of CdO in contrast to ZnO.

MgO demonstrates a straightforward band structure due to the absence of d orbitals participating in bonding. Both the sX functional and the GGA + U scheme with U = 8.5 eV applied to O 2p states, predict a similar direct band gap at Γ of approximately 7.8 eV. This is in accordance with previous theoretical studies utilising hybrid functional [83,161,164] and full GW approximation [170], as well as experimental measurements [171]. The top of the VB consists of the triple degenerate level dominated by the O 2p states, while the CBM demonstrates an s-type character. Significantly, the GGA + U calculation reveals a wider VB compared to the sX scheme, which is an anticipated outcome considering the inclusion of U in the O 2p orbital.

SrO is less studied compared to the previously mentioned metal oxide. The sX calculation predicts an indirect band gap of 5.37 eV with the VBM at Γ and the CBM at X, while the direct band gap at Γ is measured at 5.83 eV. By including a value of U = 8 eV on O 2p states, the GGA + U method yields an indirect Γ–X gap of 5.36 eV and a direct Γ–Γ gap of 6.06 eV. These computed band gap values align well with the previous theoretical work using the LDA functional with quasiparticle corrections [172]. To date, no experimental data regarding the indirect band gap is available. However, the experimental band gap values at Γ are reported to be 5.89 eV [173] and 6.08 eV [174], which match the present calculations. The band structure details obtained from

both sX and GGA + U methods are remarkably consistent. SrO features a narrow O 2p valence bandwidth, which is attributed to its high ionicity. This characteristic has been recognised through other methodologies, such as the LDA + GW [172] and the tight-binding method with scaling proposed by Pantelides [175]. It is essential to note that in the cases of MgO and SrO, the addition of the U term essentially functions as a scissors operator, expanding the GGA gap while largely preserving the band shapes.

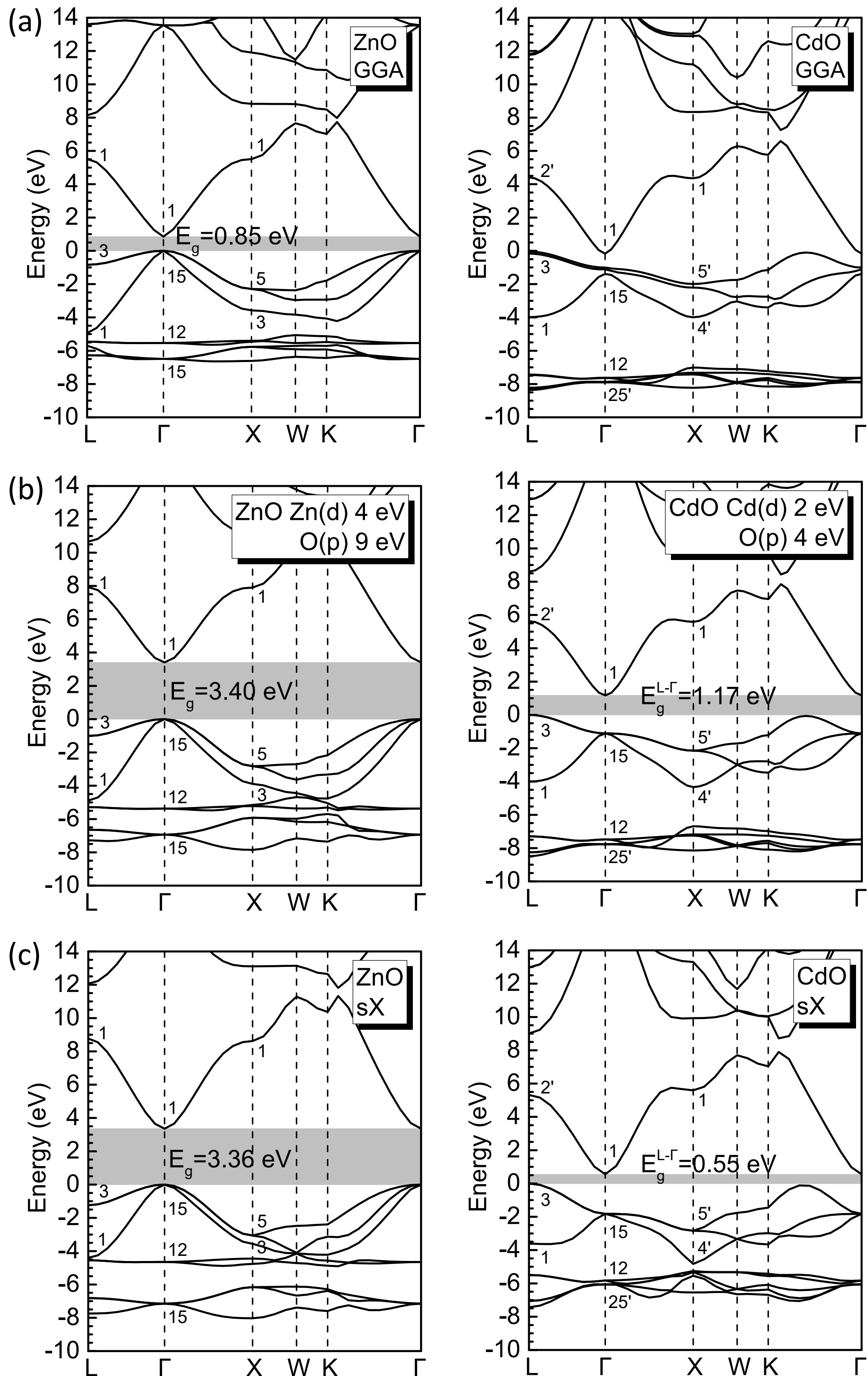


Figure 4.1 Calculated electronic band structures of zincblende ZnO (left panel) and rock salt CdO (right panel), using (a) GGA, (b) GGA + U, and (c) sX, respectively. The minimum band gap is labelled and shaded in grey. The reference point of energy zero is set at the VBM.

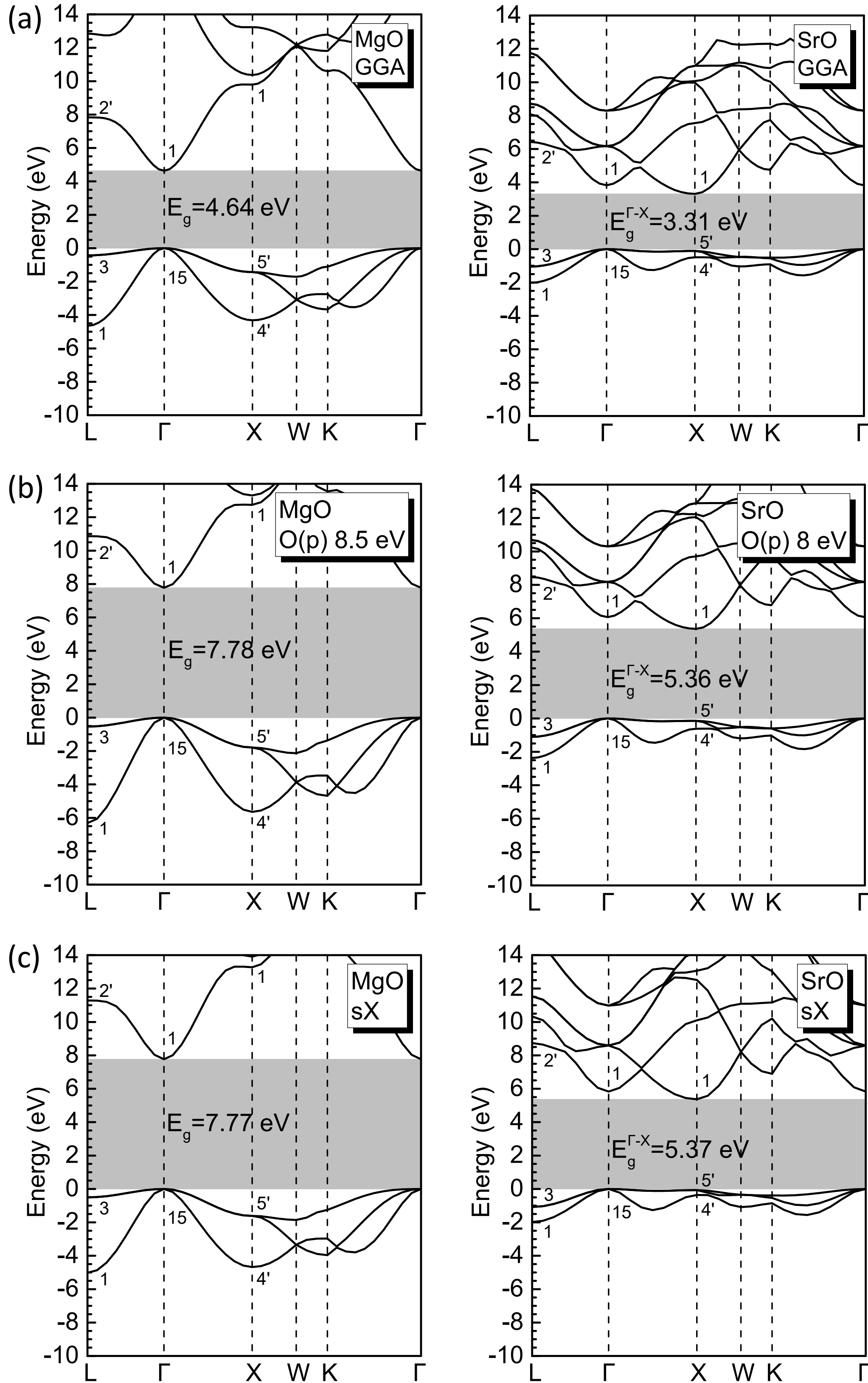


Figure 4.2 Calculated electronic band structures of rock salt MgO (left panel) and SrO (right panel), using (a) GGA, (b) GGA + U, and (c) sX, respectively. The minimum band gap is labelled and shaded in grey. The reference point of energy zero is set at the VBM.

Table 4.2 Band energies for cubic CdO, ZnO, MgO, and SrO, calculated using GGA, GGA + U, sX, and HSE methods, compared to experimental data. The positions of the lowest CB (c) and the highest VB (v) at high symmetry points are studied. For ZnO, MgO, and SrO, the energy zero is set at the uppermost VB at Γ. For CdO, the energy zero is set at L_{3v}. The calculated EA values are listed as a reference for the absolute energy scale.

		EA	Γ_{1c}	Γ_{15v}	$L_{2'c}$	L_{3v}	X_{1c}	$X_{5'v}$	Minimum gap
CdO	GGA	4.15	−0.15	−1.40	4.47	0.00	4.37	−2.07	…
	GGA + U	4.62	1.17	−1.10	5.62	0.00	5.58	−2.12	1.17 (L–Γ)
	sX	5.25	0.55	−1.85	5.32	0.00	5.65	−2.82	0.55 (L–Γ)
	HSE (α = 0.25)	5.03	0.92	−1.18	5.95	0.00	5.56	−2.12	0.92 (L–Γ)
	Experiment	5.00 [31]	2.28 [167], 2.32 [168] (Γ–Γ)		…		…		1.09 [167], 1.98 [168] (L–Γ)
MgO	GGA	0.44	4.64	0.00	7.80	−0.47	9.79	−1.45	4.64 (Γ–Γ)
	GGA + U	0.95	7.78	0.00	10.86	−0.51	12.73	−1.79	7.78 (Γ–Γ)
	sX	1.10	7.77	0.00	11.29	−0.45	13.29	−1.60	7.77 (Γ–Γ)
	HSE (α = 0.45)	0.97	7.76	0.00	11.02	−0.71	12.32	−1.61	7.76 (Γ–Γ)
	Experiment	0.85 [31]	7.77 [171] (Γ–Γ)		…		…		7.77 [171] (Γ–Γ)
SrO	GGA	0.11	3.81	0.00	6.44	−1.02	3.31	−0.16	3.31 (Γ–X)
	GGA + U	0.43	6.06	0.00	8.45	−1.10	5.36	−0.16	5.36 (Γ–X)
	sX	0.94	5.83	0.00	8.75	−1.05	5.37	−0.08	5.37 (Γ–X)
	HSE (α = 0.29)	0.95	5.97	0.00	8.77	−1.21	5.31	−0.12	5.31 (Γ–X)
	Experiment	0.64 [176]	5.89 [173], 6.08 [174] (Γ–Γ)		5.97 [173] (L–L)		6.28 [173], 5.79 [174] (X–X)		…
		EA	Γ_{1c}	Γ_{15v}	L_{1c}	L_{3v}	X_{1c}	X_{5v}	Minimum gap
ZnO	GGA	4.18	0.85	0.00	5.48	−0.83	5.52	−2.32	0.85 (Γ–Γ)
	GGA + U	4.58	3.40	0.00	7.90	−1.02	7.94	−2.81	3.40 (Γ–Γ)
	sX	4.57	3.36	0.00	8.71	−1.28	8.63	−3.10	3.36 (Γ–Γ)
	HSE (α = 0.37)	4.42	3.43	0.00	8.64	−0.75	8.03	−2.07	3.43 (Γ–Γ)
	Experiment	4.60 [31]	3.44 [136] (Γ–Γ)		…		…		3.44 [136] (Γ–Γ)

4.3.2 Electronic Properties at the Metal–Oxide Interfaces

For ZnO, both the non-polar (110) and polar (111) interfaces are investigated. In the case of MgO and SrO, the calculations are centred on the non-polar (001) and polar (111) interfaces. For CdO, the variability of SBHs is substantial relative to its small band gap, which could adversely affect fitting accuracy. Consequently, the focus is solely on its non-polar interface. this study specifically addresses O-terminated interfaces in the polar case, as they exhibit greater energetic favourability for metal contacts.

Figure 4.3 reveals a general linear dependence of SBHs on metal work function for all four oxides. It is important to note that this plot is the result of a systematic study; the data points do not necessarily align to form a perfectly straight line, and other factors beyond the metal work function can significantly influence the SBHs. Any scattering of data points represents inherent variations of the interfaces studied, rather than measurement uncertainty.

For CdO, ZnO, and MgO, the data points demonstrate a relatively tight clustering, suggesting a clear linear correlation between the metal work function and the SBH. In the case of SrO, the plot displays significantly more variation. While there still appears to be an overall trend of a larger work function correlating with a smaller barrier height, the relationship is less defined due to the pronounced scatter. This increased scatter for SrO suggests that the electronic behaviours at the interface are greatly influenced by the unique characteristics of the specific metal–SrO contact. This extends beyond the basic linear model, illustrating a more profound dependence on the nature of interfacial bonding and the specific properties of the contact metals.

The slopes of different terminations are similar for each oxide. A downward shift of the p-type SBHs is evident when comparing O-terminated polar faces to their non-polar counterparts. The pinning factor S can be derived from the fitted slope in Fig. 4.3.

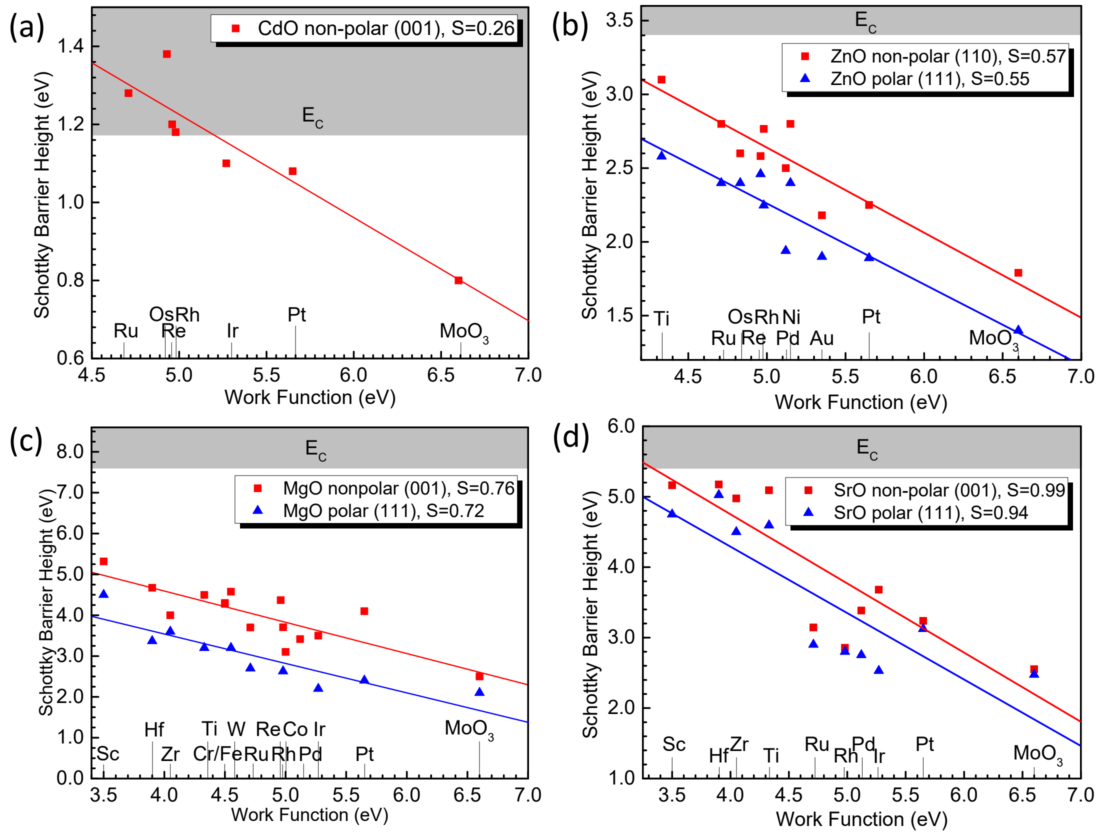


Figure 4.3 Calculated p-type SBHs for various metals on (a) CdO, (b) ZnO, (c) MgO, and (d) SrO. Work function values are sourced from Michaelson [177]. Similar pinning factors are observed across different terminations of ZnO, MgO, and SrO. A significant shift of SBHs can be observed when comparing the non-polar and polar interfaces.

The calculated S values and the SBH differences between various terminations are summarised in Table 4.3. The pinning factor demonstrates a correlation with both ionicity [178,179] and dielectric constant [31], although the latter has a more significant impact on S. CdO exhibits strongly pinned interfaces with an S value of 0.26, while ZnO and MgO display moderate S parameters of approximately 0.56 and 0.74, respectively. In contrast, SrO is weakly pinned, with an S value close to unity.

To assess the applicability of the MIGS model, the calculated S factors are compared with the empirical S factors derived from Eq. (1.6) using the experimental optical dielectric constants ε_∞ [31], as depicted in Fig. 4.4. The power law relationship between S and ε_∞ predicted by the MIGS model is validated for CdO, ZnO, and MgO by comparison. However, the FLP for the metal–SrO contacts is weaker than the empirical trend indicates. In this case, the lower S value is mainly attributed to the high ionicity and large band gap of SrO. The strong electrostatic interactions between the ions in SrO cause the electrons to be more tightly bonded and localised, leading to a reduction in the orbital overlap with the contacting metals. As a result, the density of MIGS is lowered at metal–SrO interfaces and the product of N and δ in Eq. (1.5) is decreased. This results in deviations from the empirical MIGS formula [Eq. (1.6)], highlighting the limitations of the MIGS model in accurately predicting the behaviours of materials with high ionicity, such as SrO.

Table 4.3 Comparison of pinning factors between the calculated values from Fig. 4.3 and empirical values derived from Eq. (1.6), along with the difference in SBHs for non-polar and polar interfaces. The oxide ionicities are sourced from Levine [178,179], and the dielectric constants provided are experimental values [179].

		Calculated S	SBH shift (eV)	Ionicity	ε_∞	Empirical S
CdO	…	0.26	…	0.78	6.2	0.27
ZnO	Non-polar	0.57	0.4	0.65	4.0	0.53
	Polar	0.55				
MgO	Non-polar	0.76	1.0	0.84	3.0	0.71
	Polar	0.72				
SrO	Non-polar	0.99	0.5	0.93	3.3	0.65
	Polar	0.94				

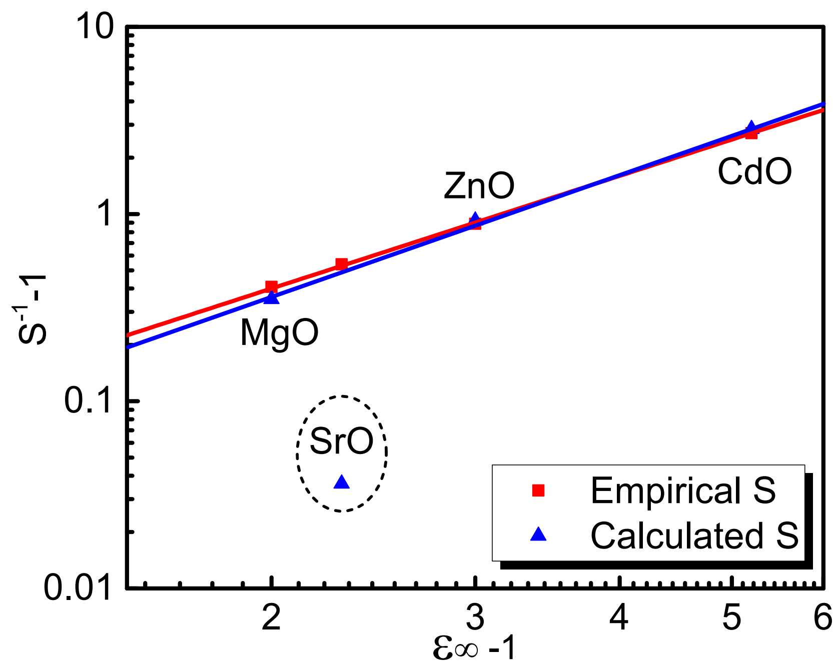


Figure 4.4 Calculated pinning factor S as a function of the experimental optical dielectric constant ε_∞ [31], along with the reference empirical value for S derived from Eq. (1.6). The special case of SrO (circled) is excluded from the fitting.

The interfacial electronic characteristics of various metal–oxide interfaces are further examined to unravel the underlying mechanisms influencing the variation in pinning factor values and the observed shifts in SBHs. This investigation delves into different aspects of the interfaces, including the partial density of states (PDOS), the wavefunction at the metal–oxide interfaces, and the charge density difference, to illuminate the correlation between the interfacial structures, the formation of MIGS, and their implications on SBH values. Pd is chosen as the contact metal for comparison due to its chemical stability, which reduces the chances of unwanted reactions that could complicate the interpretations of the results.

The PDOS and the wavefunction at metal–oxide interfaces are illustrated in Fig. 4.5, with the focus placed specifically on the interactions of Pd with ZnO and SrO. These oxides are selected for their substantial difference in pinning strength, which is expected to highlight the most prominent differences in their interfacial electronic structures. By focusing on the non-polar oxide interfaces, the impact of additional interface dipoles can be minimised, and the influence of MIGS can be more easily visualised.

The left panel of Fig. 4.5(a) displays the PDOS of individual Pd and ZnO layers in the interface region. The interfacial states are pronounced and span across the original band gap of ZnO. These gap states share comparable features with Pd and decay as they extend deeper into the ZnO bulk, which is a distinctive characteristic of MIGS as evanescent states of the metal wave function within the forbidden energy gap of the semiconductor. The right panel of Fig. 4.5(a) provides a more intuitive view of the interface-induced states. The wavefunction on the Pd side shows significant d orbital character, and the attenuation of states can be observed on the ZnO side, which is consistent with the PDOS results. Electrons are localised around the O atoms, which can be attributed to the fact that the upper VB of ZnO is primarily composed of O 2p states.

In contrast, the PDOS of the Pd–SrO interface presents minimal perturbation on the SrO side, as shown in the left panel of Fig. 4.5(b). The induced states within the band gap of SrO are negligible and confined only to the first layer at the interface, indicating that there are fewer available states to accommodate charge transfer and leading to an unpinned interface. The wave function plot in the right panel of Fig. 4.5(b) reveals that the charge is concentrated at the interfacial bond formed between Pd and O

atoms, suggesting that the induced gap states primarily result from the formation of weak covalent bonds rather than MIGS.

While MIGS are evident at these interfaces, especially in the case of ZnO, their contribution to the pinning effect is generally minimal. In practical scenarios, most of the pinning observed experimentally tends to be attributed to defects [137,180]. In the case of ZnO, oxygen vacancies have been shown to contribute to pinning [180], with an energy level of 0.7 eV below the CB edge. Taking these factors into account, the MIGS theory still provides a successful prediction for the interface pinning effect.

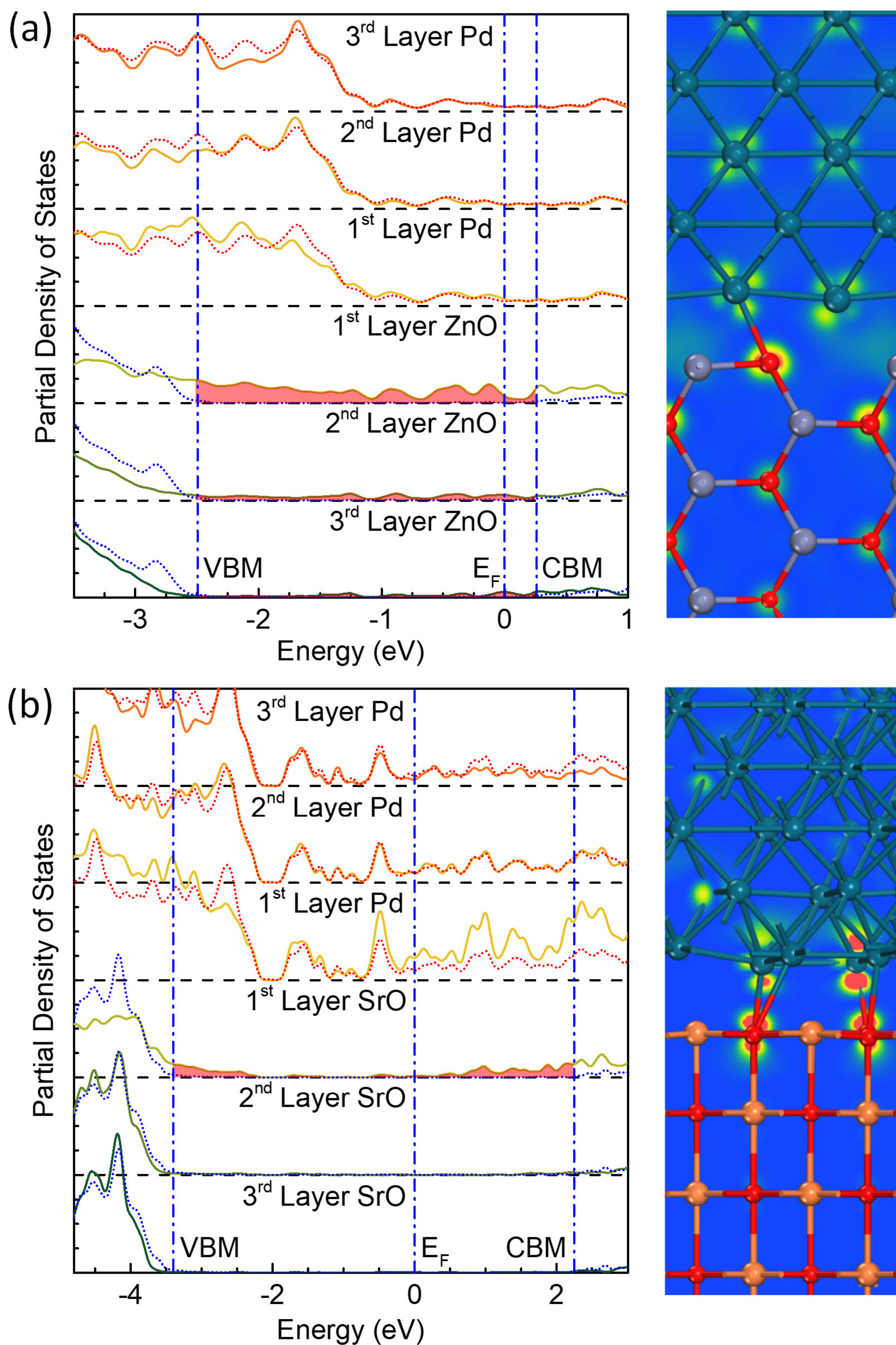


Figure 4.5 PDOS for (a) Pd–ZnO and (b) Pd–SrO interfaces, with the VBM, CBM, and Fermi level labelled. Interface-induced gap states are shaded in red. The dotted red lines represent the PDOS of Pd, while the dotted blue lines represent the PDOS of ZnO or SrO. The right panel of each image displays the wavefunction with energy levels around the metal Fermi energy.

An intriguing observation from the results is that, while different terminations of a given semiconductor share the same pinning strength S at the interface, there is a considerable difference between the non-polar and polar terminated interfaces. The SBHs of ZnO are approximately 0.4 eV higher when the orientation is changed from polar (111) to non-polar (110), while for MgO and SrO, the (001) SBH values lie roughly 0.5 eV and 1 eV above the O-terminated (111) SBHs, respectively. Notably, the shift in SBH changes inversely with the optical dielectric constant of the oxides, which is consistent with the findings of previous studies [147,148]. These findings can be explained by the additional dipole created at the polar interface, which includes the charge transfer effect and the image effect. The charge transfer between the metal and the induced interface states can introduce a net charge at the interface, the magnitude of which depends on the Fermi energy compared to the CNL energy. The charge density difference Δρ is calculated as follows:

$$\Delta\rho = \rho_{metal-oxide} - \rho_{metal} - \rho_{oxide}, \quad (4.2)$$

where $\rho_{metal-oxide}$ represents the total charge density of the metal–oxide contact, while ρ_{metal} and ρ_{oxide} denote the charge densities of isolated metal and oxide slabs with geometries identical to those in the combined system.

The charge density difference calculated with Eq. (4.2) is depicted in Fig. 4.6 using 2D energy slices oriented along the z-direction. The Pd–ZnO and Pd–MgO contacts are chosen here, as they represent the two extreme cases in terms of SBH shifts. The charge accumulation around O and Pd exhibits characteristics of s and d orbitals, respectively. Meanwhile, a depopulation of the O p orbital and Pd d_{z2} orbital can be observed.

For non-polar configurations, both Pd–ZnO and Pd–MgO interfaces display a significant enhancement of charge density between Pd and the interface O atoms, coupled with a corresponding decrease at the Pd atoms. This effect appears more pronounced for the MgO system, suggesting an augmented electron transfer at the Pd–MgO junction.

Contrarily, for polar configurations, electrons exhibit greater localisation around the interface O atoms, indicating a mild charge redistribution. This can be rationalised by the existence of a preformed dipole at the polar surface—the negatively charged oxygen atoms present increased resistance to charge redistribution at the interface upon metal contact, which consequently reduces the p-type SBH.

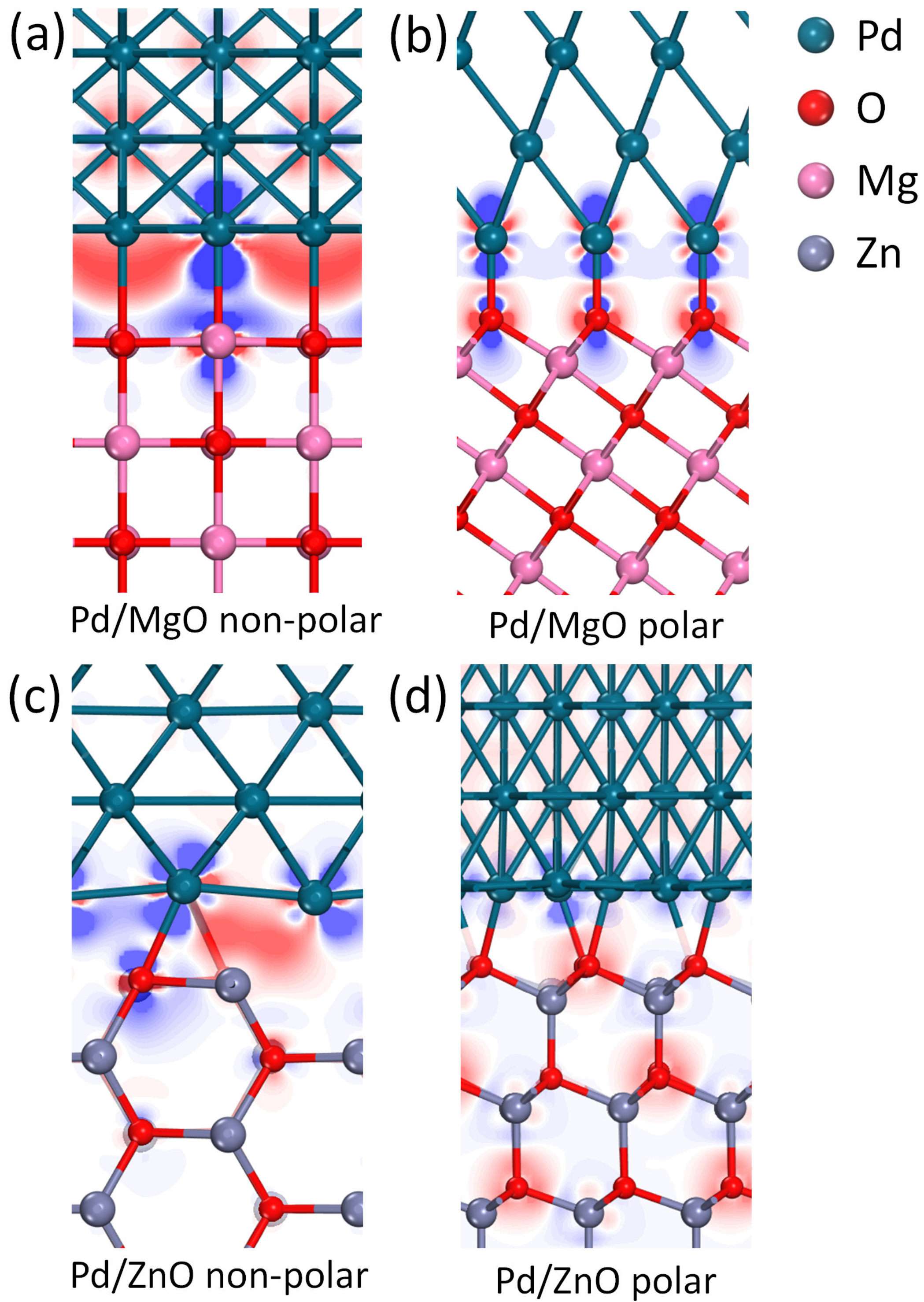


Figure 4.6 Charge density difference for (a) Pd–MgO non-polar, (b) Pd–MgO polar, (c) Pd–ZnO non-polar, and (d) Pd–ZnO polar interfaces. Red regions indicate charge accumulation, while blue regions represent charge depletion.

Comparing models for metal–MgO interfaces in MTJs is particularly intriguing due to their recent technological relevance [140–142]. MgO serves as the tunnel barrier between two electrodes made of BCC CoFe alloy, with boron initially added as an amorphising agent. Annealing subsequently causes boron to be released into the underlying Ta layer, resulting in the crystallisation of CoFe templated on the cubic MgO. The final interface is notably flat and well-separated [141,142]. The tunnelling conductance and magnetoresistance, being crucial aspects of MTJ functionality, exhibit high sensitivity to the symmetry of MIGS at the Fermi energy. Additionally, the formation of the sub-stochiometric layer of FeO could potentially degrade the performance of MTJs [181,182]. The stability of spin polarisation promotes electrode spin directions perpendicular to the MgO plane [141].

First-principles investigations have been reported at the Fe–MgO(100) structures to explain the interfacial magnetic anisotropy. Nakamura *et al.* [150,183] proposed that the weak hybridisation of O and Fe orbitals results in a significant out-of-plane magnetocrystalline anisotropy (MCA), and the presence of an interfacial FeO layer enables MCA switching by an electric field. Yang *et al.* [151] stated that the interfacial PMA is attributed to the interactions between O p_z and Fe d_{z2} orbitals hybridised with $d_{xz(yz)}$ orbitals. They also demonstrated weakened PMA value with over- or under-oxidised interfacial geometries. Recent high-resolution transmission electron microscopy (TEM) imaging supports the sharpness of the CoFe–MgO interface, indicating that any non-stoichiometric interfacial layer is minimal [142], which highlights the significance of metal–insulator interfaces and SBHs for MTJs.

4.4 Conclusions

In conclusion, SBHs of different metal–oxide interfaces are systematically investigated. Different terminations of the same oxide are found to exhibit similar pinning strength, which increases with the oxide ionicity. For more covalent oxides like CdO, MIGS are the primary contributors to FLP. For more ionic oxides like ZnO and MgO, although MIGS are experimentally less important than defects for the pinning effect, the pinning factor S predicted by first-principles calculations still aligns with the empirical MIGS model. SrO forms unpinned interfaces, and the simple model is not applicable due to its high ionicity. These findings suggest that MIGS could serve as a useful simplified model, with its predictions being more accurate for less ionic oxides.

There is a notable upward shift in SBH values when the oxide termination changes from polar to non-polar. The difference in SBHs can be explained by dipole formation due to charge transfer at the interface. The magnitude of the shift is found to vary inversely with the oxide dielectric constant.

CHAPTER 5: ELECTRONIC PROPERTIES OF CaF_2 BULK AND INTERFACES

The electronic band structures of the ultra-wide band gap CaF_2 are investigated using both the sX and the GGA + U scheme. The hybrid functional approach demonstrates excellent agreement with experimental results, and the inclusion of an on-site Coulomb interaction for F 2p orbitals also accurately reproduces the experimental band gap. The GGA + U method is further applied to study Si–CaF_2 and metal–CaF_2 interfaces. The insulating Si–CaF_2 interfaces with a clear band gap are constructed based on the electron counting rule. The supercell calculations of Si–CaF_2 interfaces reveal a type-I band alignment, with the VBOs following a descending trend from (001) to (111) and finally to (110). The calculation convergence of GGA + U is further tested with metallic contacts. Metal–CaF_2 interfaces exhibit a weak pinning effect, and different CaF_2 terminations share a similar pinning factor S of approximately 0.9, attributable to the highly ionic nature of CaF_2. The GGA + U approach is validated as a valuable tool for studying fluoride and related interfaces.

5.1 Introduction

Calcium fluoride (CaF_2) is a well-known FCC crystal with an ultra-wide band gap (~11.8 eV) and a relatively high dielectric constant (~8.43). Its interesting luminescent properties contribute to its potential application in bioimaging [184]. The significant progress recently achieved in molecular beam epitaxy (MBE) growth of thin CaF_2 layers also makes it a possible gate insulator in 2D material-based field-effect transistors (FETs) [185]. Moreover, CaF_2 was introduced as an insulation layer to facilitate the high-temperature applications of InGaN-based Schottky photodiodes (SPDs) through the implementation of a metal–insulator–semiconductor (MIS) structure [186]. Consequently, a thorough analysis of the interfaces between CaF_2 and various semiconductors or metals is imperative for understanding the experimentally observed properties at a microscopic scale, to provide valuable insights that can further advance device performance.

The accurate theoretical representation of CaF_2-based systems requires the precise prediction of band gaps for the materials involved. However, the calculation of

these quantities using DFT with LDA or GGA functionals often results in severe band gap underestimation. This deficiency arises from the improper treatment of electron self-interaction, leading to errors when describing electron and hole states. The case of CaF_2 presents a unique challenge due to the complexity of its electronic states, which is particularly noticeable in the lower CBs, where delocalised Ca 4s and F 3s orbitals contribute to a highly dispersive band, while the strongly localised Ca 3d orbitals lead to several flat bands [187].

To address this deficiency, advanced methods have been developed. One such method is the GW approximation [106,188–190], which calculates quasiparticle energies utilising the Green's function (G) and the screened Coulomb interaction (W). This approach can rectify band gap errors; however, it is computationally demanding and not well-suited for large supercells. An accurate description of Ca 3d states in the lower CBs requires specific treatment of Ca 3s and 3p semi-core states as valence states [189,190], leading to a considerable increase in computational requirements, even for bulk calculations.

An alternative method involves hybrid functional calculations, such as HSE [76,190] and sX [83,191] functionals. These calculations are less computationally intensive and have proven to be successful in treating most semiconductors and related heterostructures. However, the applicability of these hybrid functionals to wide gap materials remains uncertain, and there is a scarcity of studies focusing on CaF_2 [190,191]. Furthermore, the addition of HF exchange can cause a divergence at the Fermi energy with metallic systems, such as Schottky barriers or metallic grain boundaries [76,165].

A third method, GGA + U, was initially introduced to address electron–electron repulsion in transition metal and rare earth compounds [66] by adding the Hubbard parameter U to the open-shell metal d states. This approach was subsequently extended to closed-shell transition metal compounds, such as ZnO, CdO, and Cu_2O [159,192]. However, the correction of band gap values often requires unphysically large U values. It was later observed that combinations of moderate U values on both metal d states and oxygen p states can effectively increase the band gap [193,194]. Recent studies have demonstrated that adding a U term exclusively to the O 2p states, without involving any d states, can reproduce the band structures of oxides like MgO and SrO [194]. In general, GGA + U is considered a viable alternative method for electronic structure correction, offering the efficiency of traditional DFT functionals alongside the accuracy of hybrid

functionals. It is worth noting that incorporating the U term in HfO_2 resulted in an overestimation of EA [193], though this issue was not observed in MgO [194]. In this study, the GGA + U scheme is applied to the similarly ionic CaF_2 system.

As band edge levels from periodic bulk calculations cannot be directly compared with the experimental references, it is crucial to assess the accuracy of theoretical schemes in predicting ionisation potential (IP) and EA at the surface, as well as band offset at the interface [190]. Validating theoretical values against experimental measurements requires employing realistic interface supercell slab models that consider detailed surface reconstructions and specific interface bonding configurations.

The Si–CaF_2 interface serves as an excellent test case for this purpose. The closely matched lattice constants of CaF_2 (5.46 Å) and Si (5.43 Å) facilitate the growth of high-quality CaF_2 layers on silicon substrates using MBE [195]. There are extensive reports on the structural [196–198] and electronic [196,199–203] properties of the Si–CaF_2(111), which provide valuable experimental references for evaluating the accuracy of theoretical calculations. In this study, the investigation goes beyond the extensively researched Si–CaF_2(111) interface, to also include the modelling of interfaces in the (001) and (110) orientations. By examining these additional configurations, the aim is to provide a more complete picture of the underlying mechanisms and potential device applications involving the Si–CaF_2 interface.

In addition to the Si–CaF_2 interfaces, this study examines the technologically relevant metal–CaF_2 interfaces. As a notable class of superionic crystals [204], CaF_2 is anticipated to form weakly interacting interfaces with metals, following the Schottky–Mott rule. The degree of FLP should approach the Schottky limit, which corresponds to a near-unity pinning factor S ($S = \partial\varphi/\partial\Phi_M$, where φ represents the SBH and Φ_M denotes the metal work function) [38,40]. Despite the potential applications of these interfaces, there are currently limited reports available on metal–CaF_2 interfaces.

This work aims to test the accuracy of various DFT schemes (GGA, GGA + U, and sX) in describing different properties of the representative wide band gap fluoride, CaF_2. By examining both Si–CaF_2 and metal–CaF_2 interfaces, this study intends to fill the gap in the existing theoretical data on these appealing interfaces. A comprehensive understanding of interfacial electronic properties obtained through this research will be valuable for potential applications involving CaF_2-based interfaces.

5.2 Calculation Methods

The calculations employ the CASTEP code [110], utilising the plane wave pseudopotential method and the PBE version of GGA for the XC functional. Ultra-soft pseudopotentials with a plane wave cutoff energy of 400 eV are adopted. Energies are converged to below 10^{-6} eV/atom and forces to below 10^{-3} eV/Å. A dense 11 × 11 × 11 k-point mesh is used for the Si primitive cell to include the conduction band minimum (CBM), which is not located at the high symmetry k-points. Meanwhile, a 7 × 7 × 7 k-point mesh is used for the CaF_2 primitive cell.

For the Si–CaF_2 interfaces, supercells are constructed with approximately 20 Å of CaF_2 and 20 Å of Si with no vacuum layer separation. Details of the models with different Si–CaF_2 orientations will be discussed in Sec. 5.3.2. For the metal–CaF_2 interfaces, the F-terminated polar (111) and (001) orientations, as well as the non-polar (110) orientation, are studied with various metals. The models are constructed using six layers of metals in their FCC structure, approximately 20 Å of CaF_2, and 15 Å of vacuum.

The band offsets are extracted using the core-level method [157,158] to enhance the precision in determining the VBM, using Eq. (5.1), where E_{core}^{int} is the core level state in the interfacial model, ΔV is the energy difference between the core level state and VBM of the bulk materials, and E_F is the metal work function. The SBHs are evaluated using the same method and Eq. (4.1) as defined in Chapter 4. The semi-core 2s state of F atom away from the interface region is chosen as the reference core level state here,

$$\Delta E_V = (E_{core-A}^{int} - E_{core-B}^{int}) + (\Delta V_A - \Delta V_B). \tag{5.1}$$

5.3 Results and Discussion

5.3.1 Band Structures of CaF_2

Detailed band structures of CaF_2 calculated using different functionals are presented in Fig. 5.1, with the energy levels at high symmetry points (L, Γ, and X) provided in Table 5.1. CaF_2 exhibits an indirect band gap, with the highest VB state located at the X point and the lowest CB state at the Γ point, as depicted in Fig. 5.1.

The top of the VB primarily originates from F 2p orbitals, with only minor contributions from Ca, as presented in the right panel of Fig. 5.1(a). The presence of two F atoms in a unit cell results in two distinct VB levels at the Γ point ($\Gamma_{25'}$ and Γ_{15}). The VBM lies at the X point rather than the Γ point can be attributed to the phase reversal that occurs between two translation-related F atoms in the x-direction, whose x coordinates differ by $\frac{1}{2}$a (a being the lattice constant of the cubic crystal). At the X point, the bonding conditions between the p_x orbitals of these two F atoms transform into completely antibonding or bonding states due to this phase reversal. As a result, the VB states split, causing the $X_{2'}$ state to lie above the Γ_{15} state, which makes the $X_{2'}$ state the VBM for CaF_2.

The bottom of the CB predominantly consists of Ca 3d orbitals, with some contributions from Ca 4s and F 3s orbitals. The Γ_1 state at the bottom of the CB is formed by the Ca 4s and F 3s orbitals. The delocalised nature of s orbitals leads to a higher degree of orbital overlap between neighbouring atoms, resulting in stronger interactions and the formation of a highly dispersive band, as depicted in the left panel of Fig. 5.1(a). The total density of states (TDOS) at the bottom of the CB exhibits an s-type character, which is a smooth tail extending toward the band gap. This can be seen in the right panel of Fig. 5.1(a) and further supports the contribution of delocalised Ca 4s and F 3s orbitals at the CBM. Above the Γ_1 state, there are several flat bands primarily associated with the Ca 3d orbitals. The Γ_{12} and $\Gamma_{25'}$ states are two typical examples, characterised by limited dispersion and weak interactions between the orbitals due to the localised nature of d orbitals. The $\Gamma_{25'}$ state also exhibits a small mixture of F p orbitals, as shown by the red line in the right panel of Fig. 5.1(a), which indicates the interactions between the Ca 3d and F p states.

The inclusion of the sX functional raises the CB, resulting in a substantially wider band gap compared to the GGA method. The direct gap (Γ–Γ) and the indirect gap (X–Γ) are 12.04 eV and 11.78 eV, respectively. These values are close to the experimental observations of 12.1 eV for the direct gap and 11.8 eV for the indirect gap [205]. The top VB states maintain the same structure, and the valence bandwidth slightly increases from 2.64 eV in GGA to 2.75 eV in sX, addressing the compressed VB reported in previous work [191]. The dispersion of the CB undergoes significant modification, with the CBM at Γ being 1.35 eV lower than the X_1 state, which is comparable to earlier GW results [189].

Various U value combinations are tested on Ca d and F p orbitals. The U parameters are selected by minimising the error between the calculated GGA + U and sX band energies for the VB and CB states [193] at the high symmetry points L, Γ, and X. The calculations reveal that adding U to Ca d states negatively impacts the band gap value, which is different from the case of ZnO [194]. The distinct electronic structures and bonding types of these materials contribute to the differing effects of the Hubbard U correction on their band gaps. The covalent nature of ZnO results in a significant interaction between the Zn 3d and O 2p orbitals in the VB. Applying a Hubbard U correction to the d orbitals of Zn can better represent the localised nature of the d electrons, reducing the p–d hybridisation and adjusting the repulsion between the Zn d orbitals and the CB levels. This process effectively lowers the energy levels associated with the d orbitals, leading to an increased band gap. On the other hand, the CB of CaF_2 includes Ca 3d orbitals. When the U correction is applied to these orbitals, they become more localised, potentially weakening the interactions between the Ca 3d and 4s orbitals. This results in the reduction of the 4s orbital energy that constitutes the CBM, leading to a narrower band gap.

Alternatively, incorporating U into F p states in CaF_2 effectively increases the band gap, similar to the case of MgO [194]. By adding U = 9.5 eV to the F 2p states, the band gap of CaF_2 can be corrected to align closely with the sX results, yielding a value of 11.78 eV, as shown by the red lines in Fig. 5.1(b). With this correction, both the top VB and the bottom CB maintain the GGA band structure. The valence and conduction bandwidths are expanded to 3.60 eV and 3.46 eV, respectively.

It is important to note that different calculation methods can position band edges differently, even when they yield the same band gap values. This phenomenon can directly influence the prediction of heterostructure band offsets. Therefore, the IP and EA of bulk cubic CaF_2 are also considered, using a non-polar (110) CaF_2 slab with a thick vacuum layer to validate the GGA + U method in an absolute energy scale. The results are summarised in Table 5.1. Overall, the results calculated by the GGA + U method are in good agreement with both the sX calculations and experimental values. The accurate prediction of IP and EA values by the GGA + U method is comparable to the case of MgO [194] and surpasses the case of HfO_2 [193].

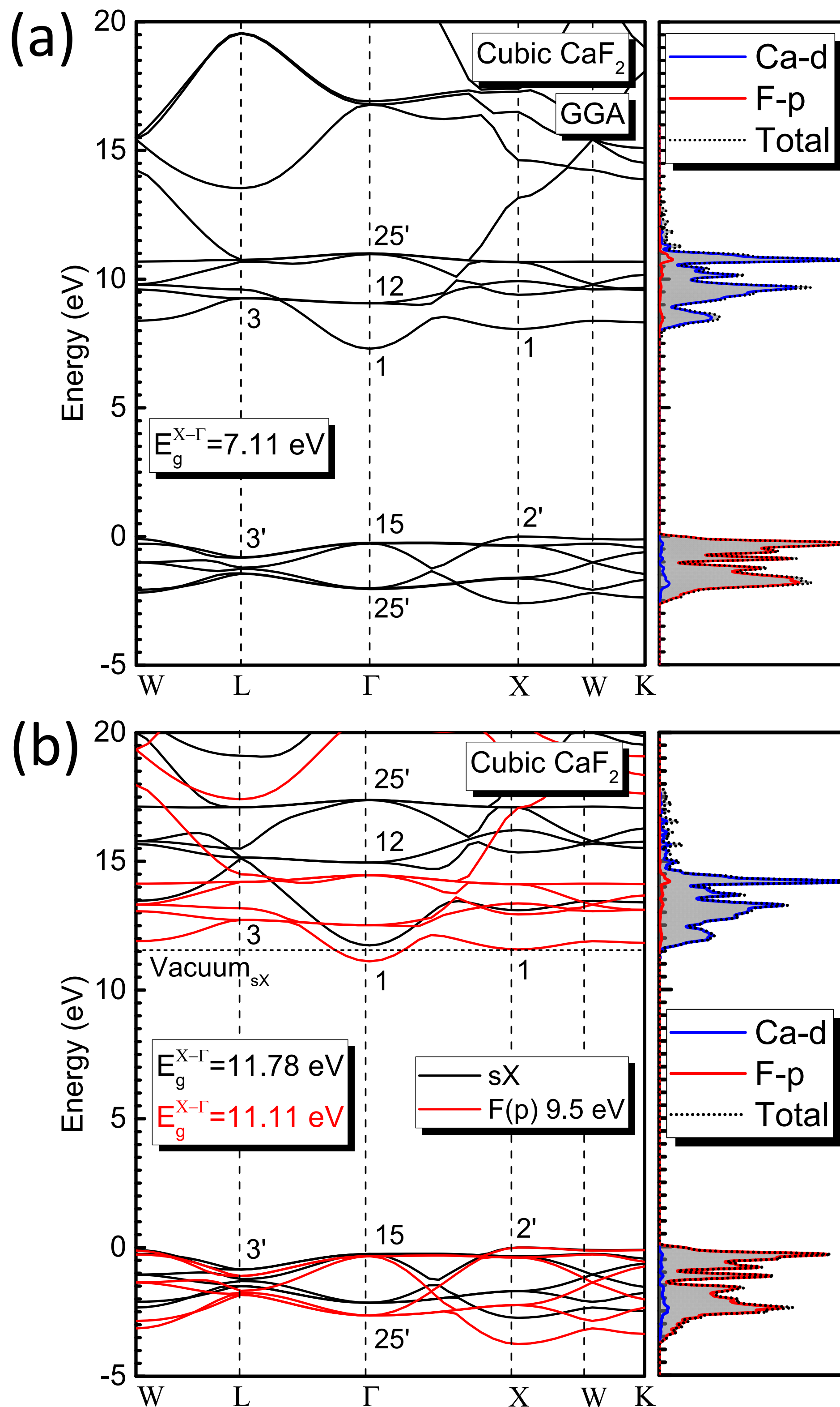


Figure 5.1 Comparison of calculated band structures and PDOS for CaF_2 using (a) the GGA functional, (b) the sX functional (represented by black lines), and the GGA + U scheme (represented by red lines). The TDOS of bulk CaF_2 is shaded in the DOS plots. The VBM at X is set to an energy level of zero.

Table 5.1 Band energies, EA, and IP values for CaF_2, calculated using GGA, GGA + U, and sX methods, compared to experimental data. The positions of the lowest CB (c) and the highest VB (v) at high symmetry points are studied. The uppermost VB at X is set at energy zero.

<table>
<tr><th></th><th></th><th>L_{3c}</th><th>$L_{3'v}$</th><th>Γ_{1c}</th><th>Γ_{15v}</th><th>X_{1c}</th><th>$X_{2'v}$</th><th>Minimum gap (X–Γ)</th><th>EA</th><th>IP</th></tr>
<tr><td>GGA</td><td></td><td>9.23</td><td>−0.77</td><td>7.11</td><td>−0.28</td><td>8.05</td><td>0.00</td><td>7.11</td><td>−0.30</td><td>6.81</td></tr>
<tr><td>GGA + U</td><td></td><td>12.70</td><td>−1.07</td><td>11.11</td><td>−0.34</td><td>11.57</td><td>0.00</td><td>11.11</td><td>−0.39</td><td>10.71</td></tr>
<tr><td>sX</td><td></td><td>15.05</td><td>−0.86</td><td>11.78</td><td>−0.24</td><td>13.13</td><td>0.00</td><td>11.78</td><td>−0.22</td><td>11.55</td></tr>
<tr><td rowspan="2">Experiment</td><td>[205]</td><td colspan="6">…</td><td>11.80</td><td>…</td><td>…</td></tr>
<tr><td>[206]</td><td colspan="6">…</td><td>12.11</td><td>−0.15</td><td>11.96</td></tr>
</table>

5.3.2 Electronic Properties at the Si–CaF_2 Interfaces

Having thoroughly examined the electronic structures of CaF_2, the investigation proceeds to examine the Si–CaF_2 interfaces. The Si–CaF_2 interfaces in (111), (001), and (110) orientations are examined, with the modelling details outlined below.

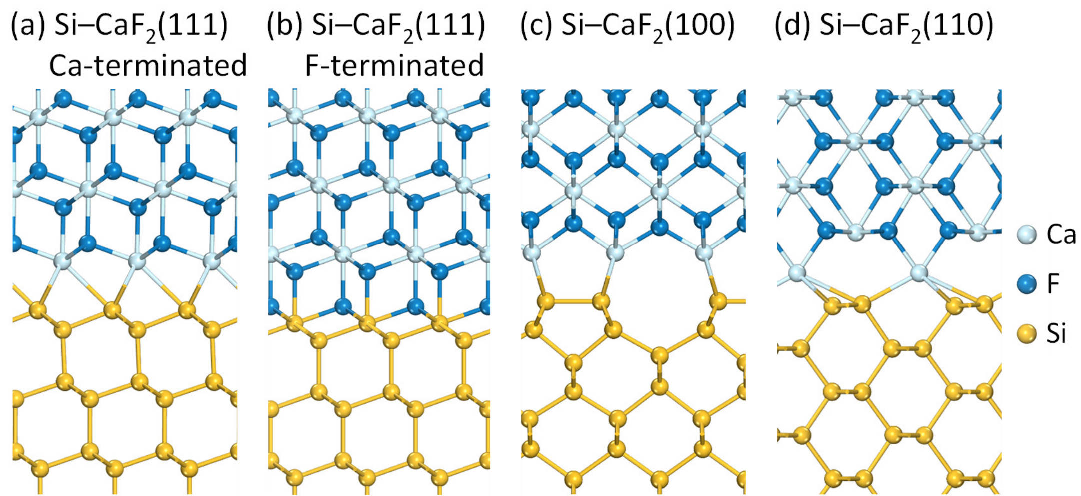
For the Si–CaF_2(111) interface, two types of terminations can be found on the CaF_2(111) surface: Ca-terminated and F-terminated. For the Ca-terminated case, the T_4 configuration with B-type orientation along the z-direction is adopted, as depicted in Fig. 5.2(a). The interface is charge neutral, with Ca atoms positioned at the high symmetry T_4 sites and F atoms occupying H_3 sites on the Si(111) (1 × 1) surface. The validity of this interface morphology has been extensively justified experimentally [197,207], and theoretical works based on this model have demonstrated accurate reproduction of experimentally measured band alignment [190]. Regarding the F-terminated case, there are experimental observations supporting the formation of Si–F bonds at the interface [196]; however, detailed information on interface bonding remains limited. Consequently, various F-terminated interface models are examined to identify the most energetically favourable configuration. The most structurally stable configuration features two interfacial fluorine layers, with F atoms positioned at the hollow sites of the first-layer Si atoms and adopting an A-type orientation, as illustrated in Fig. 5.2(b).

For the Si–CaF_2(001) interface, previous research has found that when CaF_2 is deposited onto Si(001) at temperatures below 600 °C, the lattice orientation of CaF_2 mimics the orientation of the substrate [208]. Atomic force microscopy (AFM) images and reflection high-energy electron diffraction (RHEED) results indicate that the Si(001) surface is reconstructed in the form of dimer rows, leading to a 2 × 1 periodicity. Additionally, the formation of interfacial Si–Ca bonds is supported by photoemission [208]. Based on the available experimental data, a 2 × 1 reconstructed interface is proposed, which consists of fivefold Si sites with lateral Si–Si bonds and Ca atoms positioned on top of the second layer of Si atoms [Fig. 5.2(c)]. The proposed Si–CaF_2(001) interface model is inspired by the well-established sevenfold model for Si–silicide(001) interfaces [209,210].

The present reports are not conclusive about the detailed atomic geometry at the Si–CaF_2(110) interface. The interface model is constructed with Si(110)‖CaF_2(110) and Si[110]‖CaF_2[110]. However, the direct contact between the ideal (110) faces of

Si and CaF_2 does not yield an insulating interface, as states now appear within the band gap: CaF_2 exhibits ionic bonding without fixed coordination, similar to ZrO_2, and the (110) face of CaF_2 is neutral [211]; meanwhile, the Si(110) surface has one half-filled dangling bond per surface Si atom. To accommodate the extra electrons donated by Si and create an insulating interface, all F atoms at the first CaF_2 layer are removed to satisfy the electron counting rule [211]. This results in a Ca-terminated interface with fivefold coordinated Ca as shown in Fig. 5.2(d).

The GGA + U method, with U = 9.5 eV added to the F p orbitals, is employed for interface structural relaxation and electronic property calculations. The band gap of Si calculated using the GGA functional yields an indirect band gap of 0.68 eV. Although this value is smaller than the experimental band gap of 1.17 eV, it is important to emphasise that the band gap of Si is not the primary focus of this investigation. The difference is considered acceptable in light of the large gap value of CaF_2. The main focus of this study is to explore the band offset values at the Si–CaF_2 interface. The local DOSs of Si and CaF_2 presented in Fig. 5.2(d) demonstrate that all the interfacial models are insulating, without any gap states in the band gap. Consequently, these models are proven to perfectly satisfy the electron counting requirement, providing a solid foundation for accurate analysis of the band alignment.



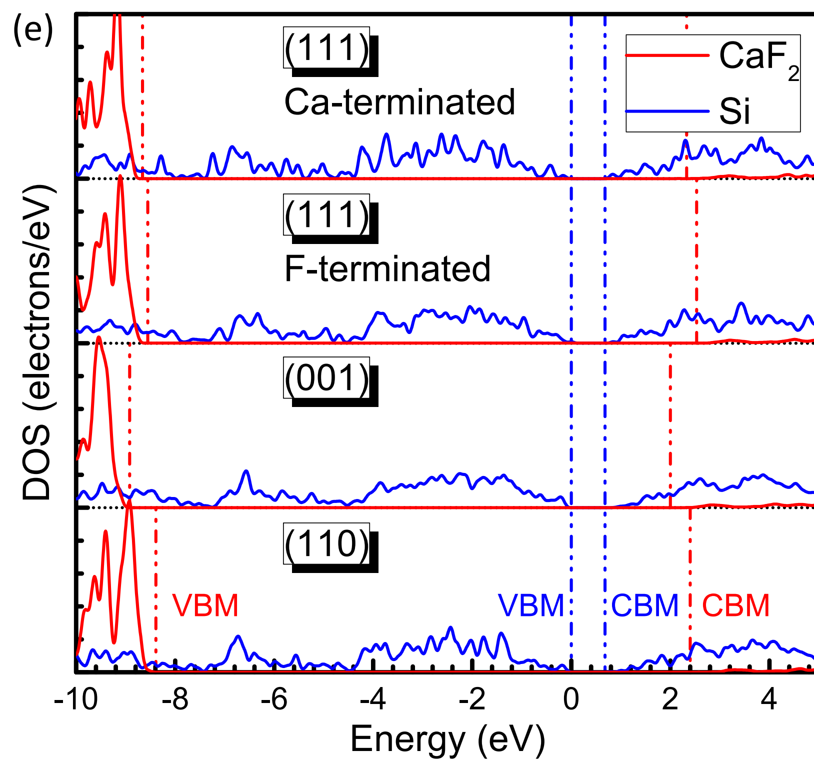


Figure 5.2 Atomic structures of the Si–CaF_2 interfaces for (a) Ca-terminated (111), (b) F-terminated (111), (c) (001), and (d) (110) orientations, respectively. The PDOS plots for these four interface models are presented in (e). The PDOS are derived from the bulk layers distant from the interfacial region. The VBM of the Si bulk is set to an energy level of zero.

Figure 5.3 presents the band offset values extracted using the core-level scheme. All three Ca-terminated interfaces exhibit a type-I band alignment. The calculated VBOs (ΔE_V) for the (111), (001), and (110) orientations are 8.66 eV, 8.92 eV, and 8.39 eV, respectively. The substrate orientation influences the band alignment and displays a relationship of $\Delta E_V(001) > \Delta E_V(111) > \Delta E_V(110)$, with a corresponding inverse trend for CBOs.

Comparing these results with available experimental and theoretical data is informative. Experimental VBOs for the Si–CaF_2(111) range from 8.5 eV to 8.9 eV [199–201], while hybrid functionals like PBE0, HSE, and GW methods used in previous theoretical work deduced VBOs from 8.28 eV to 9.01 eV [190]. The GGA + U results yield a reasonable agreement with existing results, with errors of at most 0.4 eV.

An asymmetry of band offsets is observed across all three orientations, with the VBOs consistently larger than the CBOs. These CBO values are substantial enough to prevent electron injection, validating CaF_2 as an effective insulating layer. Notably, the largest CBO value, found in the (110) orientation, presents an advantageous characteristic for potential applications such as resonant tunnelling diodes [212], although this orientation is not the most experimentally favoured.

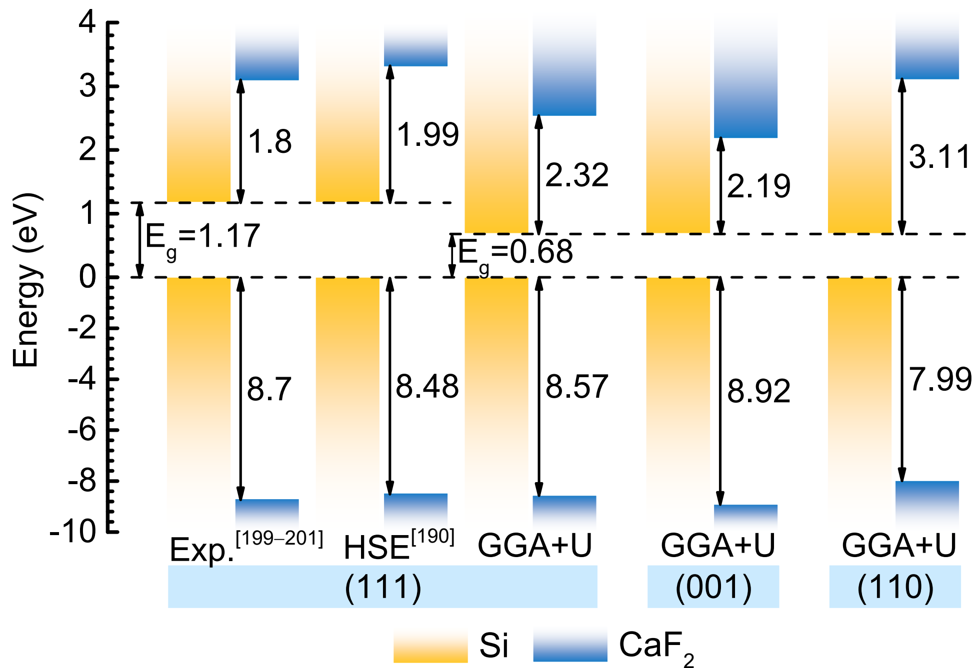


Figure 5.3 Energy band alignment diagram for Ca-terminated (111), (001), and (110) oriented interfaces. Experimental data [199–201] and theoretical data calculated using HSE [190] are provided for comparison. The CBOs of the reference data are based on the VBOs and the experimental band gap of 11.8 eV in Ref. [190].

To quantitatively characterise the electronic charge distribution at different interfaces and explain the VBO variation, Mulliken atomic population analysis is performed. Fig. 5.4 summarises the gross atomic charges by Mulliken analysis for individual layers within the interface model. It should be noted that, for the Si–CaF_2(110) case, F atoms are counted starting from the second layer due to their removal from the first layer at the interface.

The analysis reveals that the charge transfer between Si and CaF_2 is confined to the interface region, involving only Si and Ca atoms. The interfacial effect rapidly diminishes and reverts to the bulk charge population distribution starting from the third layer of Si and CaF_2. Charge transfer from Ca to Si leads to dipole formation at the interface, causing an upward shift of the CaF_2 VBM in relation to the Si VBM, and consequently reducing the VBO of the (110) interface. At the Si–CaF_2(110) interface, the Si:Ca ratio is 2:1, and each Ca atom transfers 0.93e of charge to the first Si layer. For the other two conditions, the Si:Ca ratios are equal to 1, and the charge transfer for the (111) and (001) orientations are 0.72e and 0.53e, respectively. The amount of charge transfer indicates the strength of the dipole. Therefore, the (110) oriented interface, exhibiting the most pronounced charge rearrangement, has the smallest VBO.

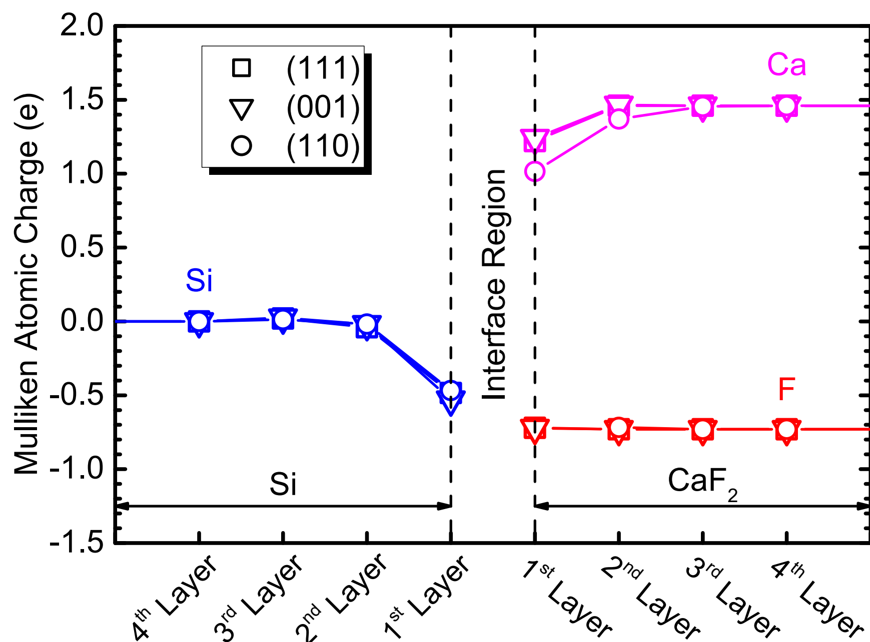
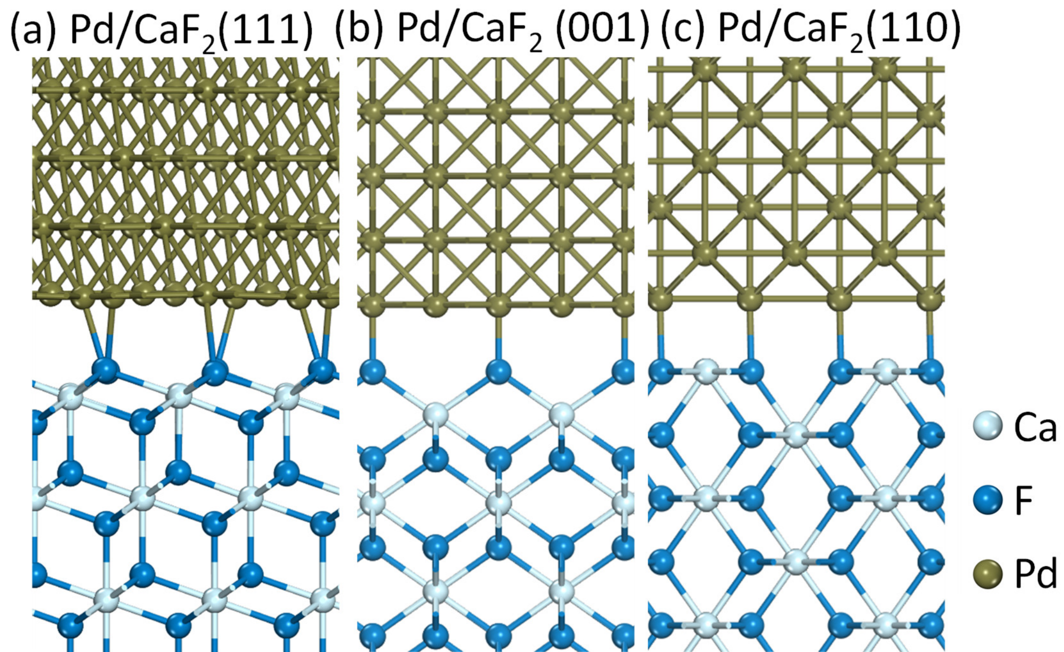


Figure 5.4 Mulliken atomic charges for Si, Ca, and F atoms in individual layers. Positive charges represent electron depletion, while negative charges indicate electron accumulation. The interface region is labelled to differentiate between the two sides.

5.3.3 Electronic Properties at the Metal–CaF_2 Interfaces

The metal–CaF_2 interfaces are investigated using the GGA + U scheme to study the SBHs and assess the calculation convergence of the system. The F-terminated polar (111) and (001) surfaces, as well as the non-polar (110) surface, are employed to form interfaces with various metals spanning a wide work function range from 3.5 eV (Sc) to 5.6 eV (Pt) [177]. Although terminated with F^- ions, the (111) surface of CaF_2 can be considered as $F^-Ca^{2+}F^-$ trilayers stacked along the [111] direction [211]. The non-polar $F^-Ca^{2+}F^-$ unit yields a closed-shell configuration, making the (111) termination stable [Fig. 5.5(a)]. Applying the same concept to the F-terminated (001) surface, half of the F^- ions in one layer are equally distributed to the Ca^{2+} ions on both sides, leaving extra F^- ions at the surface. Therefore, half of the F atoms are removed from the first layer to achieve a stable termination, as illustrated in Fig. 5.5(b). The non-polar (110) surface is neutral and remains unchanged [Fig. 5.5(c)].

The SBHs are extracted from the interface calculations and plotted against the metal work function in Fig. 5.5(d). A linear dependence of SBH on the metal work function is observed. The pinning factor S is calculated by the fitted slope and is found to be approximately 0.9 for each termination, indicating weakly pinned interfaces approaching the Schottky limit [40]. The calculated S value is consistent with the highly ionic nature of CaF_2 [204] and matches the empirical value of 0.89 predicted by the MIGS model [Eq. (1.6)]. Fig. 5.5(d) also reveals a noticeable downward shift in SBHs of approximately 0.4 eV when transitioning from the (110) to the (111) orientation. And a further decrease of roughly 0.1 eV is observed when moving from the (111) to the (001) orientation.



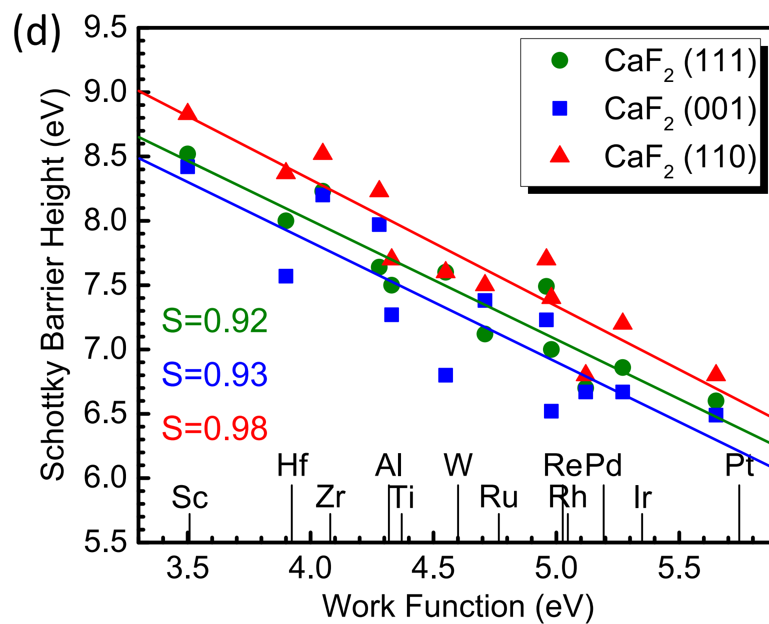


Figure 5.5 Interface models for Pd–CaF_2 with CaF_2 in (a) (111), (b) (001), and (c) (110) terminations. The calculated p-type SBHs of various metals on different terminations of CaF_2 are shown in (d). Work function values are sourced from Michaelson [177].

5.4 Conclusions

In summary, this study extensively investigates the band structure of CaF_2 using different calculation methods. Both sX and GGA + U approaches can effectively address the band gap issue of the GGA functional, providing improved accuracy in calculations. Furthermore, the GGA + U scheme is utilised to systematically study the interfaces between different CaF_2 terminations in contact with Si and various metals. The calculated band offsets at Si–CaF_2 follow a relation of $\Delta E_V(001) > \Delta E_V(111) > \Delta E_V(110)$, which is attributed to varied dipole formation corresponding to different interface orientations. The metal–CaF_2 interfaces exhibit a weak pinning effect, consistent with the MIGS prediction, and demonstrate shifts in SBHs between different CaF_2 terminations. The precise description of the electronic properties of CaF_2 verifies the effectiveness of the sX approach in treating wide band gap materials. Additionally, the GGA + U scheme, with the U term applied only to the anion p states, is validated as an efficient and reliable calculation method for fluorides such as CaF_2.

CHAPTER 6: METAL CONTACTS WITH MOIRÉ INTERFACES ON WSe_2 FOR AMBIPOLAR APPLICATIONS

The rational design of metal contacts on TMDs has the potential to substantially enhance the performance of 2D devices. In this study, the contacts between p-type and ambipolar WSe_2 and various metals employing different interfacial geometries have been thoroughly investigated. The results indicate that most metals stabilise in their on-top sites, while noble metals preferentially form Moiré interfaces. The construction of Moiré interfaces enhances the stability of physisorptive sites, consequently weakening the FLP and enabling the modulation of SBHs at these interfaces, similar to the MoS_2 case. The unusual structural and electronic properties of Ir, Pd, Au, In, and Al have been assessed in detail. Pd with its stable Moiré site, exhibits a distinctively low p-type SBH, whereas Au presents a metastable low p-type SBH. Both In and Al possess notably long interfacial bonding and low n-type SBHs. The research suggests that WSe_2 can be employed in high-performance ambipolar devices through the rational design and appropriate selection of contacting metals.

6.1 Introduction

The TMDs have recently gained attention as potential successors to graphene for valuable 2D systems [213–216]. Unlike graphene, 2D TMDs are semiconductors with band gaps, making them ideal channel materials for FETs. Their ultra-thin body, exceptional mechanical flexibility, and relatively high carrier mobility are suitable for end-of-roadmap logic devices such as TFETs [217,218].

As device downscaling continues, achieving low-resistance metal contacts poses a significant challenge, obscuring the intrinsic remarkable electronic properties of 2D semiconductors. Currently, there is no stable and dependable method for inducing heavy doping in 2D materials [219]. Given the absence of a controllable and enduring substitutional doping approach, the work function of metal becomes crucial in injecting the desired types of carriers into 2D materials. Consequently, the contact resistance in 2D material-based devices predominantly depends on the formation of Schottky

barriers at the interfaces. The SBHs restrict the efficiency of carrier tunnelling or excitation into the respective bands of 2D TMDs.

The majority of TMDs, such as MoS_2 [220–223], are classified as n-type semiconductors, while only a minority, including WSe_2 [224–226], exhibit bipolar or p-type behaviours due to their significantly lower band edge energies with respect to the vacuum level. Recently, several studies have focused on the contacts between metals and n-type MoS_2 [220–223,227–233]. Liu *et al.* [223] reported the creation of an ideal weakly bonded metal–MoS_2 interface by physically laminating metal electrodes on top of mechanically exfoliated MoS_2 flakes using a polydimethylsiloxane (PDMS) stamp, avoiding any energetic deposition or disordering process. MoS_2 contacts formed with noble or precious metals, such as Ag, Cu, Au, Pd, and Pt, approach the Schottky–Mott limit with an interfacial pinning factor S of 0.96. Although they successfully demonstrated the fundamental limit of ideal metal–MoS_2 interfaces, fabricating realistic electronic devices remains challenging. Wang *et al.* [232] achieved ultra-clean vdW contacts between In/Au electrodes and MoS_2 using a standard laboratory electron-beam evaporator under normal vacuum, resulting in excellent device performance. Wang *et al.* [233] observed that Ag has the potential to form an Ohmic contact on MoS_2 through X-ray photoelectron spectroscopy. However, these works are not systematic and lack the rigour necessary for accurate analysis.

Compared to studies on metal–MoS_2 interfaces, research on metal–WSe_2 interfaces is relatively limited [224–227,234–238]. Fang *et al.* [224] investigated various metal contacts (Pd, Ag, Ni, Au, Ti, and Gd) for back-gated WSe_2 FETs, discovering that the high work function Pd forms a contact that exhibits clear p-type conducting behaviours without ambipolar transport. Liu *et al.* [235] studied a series of low work function metals (In, Ag, Al, and Ti), suggesting that In and Ag could form Ohmic contacts with low resistances and facilitate high-performance n-type WSe_2 devices. Kong *et al.* [238] recently reported a doping-free strategy for achieving complementary WSe_2 circuit functions using the same contact metal Au, but different integration methods. They performed a low-energy vdW integration process, resulting in an atomically sharp and clean interface and a depinned Fermi level. The device with vdW integrated Au electrode demonstrates explicit p-type characteristics, in contrast to the conventional evaporated Au electrode exhibiting pronounced n-type behaviours.

Recently, there have also been studies regarding metal contacts to monolayer TMDs with hexagonal boron nitride (h-BN) as an insertion layer to enhance device

performance [239,240]. However, it is important to note that the dry vdW transfer techniques applied in these studies do not provide full surface coverage. Additionally, the high chemical vapour deposition (CVD) temperatures required for h-BN deposition (780–1100 °C) [241] are not applicable for both MoS_2 and WSe_2. In summary, while controlling the SBHs at monolayer TMDs by inserting h-BN as a tunnel barrier is scientifically valuable for understanding contact issues, it is not technologically feasible. A more practical approach would be to adjust the SBHs by exploring interfaces of metals on TMDs, aiming to provide useful guidelines for viable applications.

The SBH at the interface is influenced by the FLP effect, which is mainly attributed to the MIGS. The pinning factor S is defined as the change in the Schottky barrier with respect to the metal work function and can be calculated using Eq. (1.5). The S factor ranges between 0 (for a strongly pinned interface, i.e., Bardeen limit) and 1 (for an unpinned interface, i.e., Schottky limit). A larger S value allows the SBH to be more easily tuned by adjusting the metal work function, which is electrically desirable. Unfortunately, strong FLP is observed in numerous 2D devices, where the Fermi level is pinned within the band gap, making the control of SBH quite challenging. To increase the S factor, constructing weakly bonded vdW metal–TMD junctions is required, which results in interfaces that are free from chemical disorder and exhibit reduced MIGS interaction [223].

The metal–semiconductor interfaces can be classified as either chemisorptive or physisorptive, depending on the degree of interaction at the contacts. Generally, the chemisorptive interfaces are characterised by the formation of strong chemical bonds, resulting in robust and stable interfaces, typically accompanied by strong FLP. Conversely, the physisorptive interfaces involve the formation of vdW bonding, and the weakly bonded interfaces exhibit a reduced pinning effect. The interfacial bonding is typically an inherent characteristic of 3D semiconductors, which can be difficult to modify. However, for layered 2D semiconductors, a Moiré lattice can be leveraged to alter the interlayer interactions by applying a twisted angle between the contacting layers. The formation of Moiré lattice enables the achievement of weakly bonded physisorptive interfaces without directly lengthening the bond [242–247], thereby avoiding the introduction of additional contact resistance due to an enhanced tunnelling barrier.

h-BN exhibit a good lattice match with the (111) surface of Fe and Ni, as well as the (0001) surface of Co [248–251]. These metal–h-BN interfaces provide a clear

separation between chemisorptive and physisorptive sites, which contribute to a better understanding of the contact nature. Taking the interfaces between h-BN and the lattice-matched Ni(111) as an example [249,250]: the N on top of the Ni site involves hybridisation between the N and Ni atoms, forming shorter and stronger chemisorptive bonds; on the other hand, the B on top of Ni site forms longer B–Ni bonds and a flat interface, indicating the formation of a weakly interacted physisorptive interface.

Considering the distinct interfacial behaviours observed in h-BN, it has been proposed as a model system to study the effects of Moiré configuration. Lu *et al.* [250] discovered that the pseudo-incommensurate structures introduce weak vdW bonding, leading to physisorptive interfaces with larger interlayer spacing. Furthermore, the FLP was found to be dramatically different for different h-BN bonding sites: the chemisorptive interfaces exhibit strong pinning with an S value of approximately 0.25, while the physisorptive interfaces essentially allow for Fermi level depinning, with an S value close to 1.

The concept of constructing a weakly bonded physisorptive interface using a Moiré lattice can indeed be applied to metal–TMD interfaces. In the previous study on n-type MoS_2, the SBHs are found to follow clear trends, which not only depend on the work functions of metals, but also to a lesser extent on the contact type (chemisorptive or physisorptive) and the interfacial bond length [252]. The Moiré configurations are observed to stabilise physisorptive sites, resulting in a larger pinning factor S of ~0.37, compared to chemisorptive interfaces with S of ~0.24. The SBHs of physisorptive interfaces are found to be above those of chemisorptive interfaces and closer to the CBM. The low work function metals, such as In and Ag, are of particular interest because they are stabilised in their Moiré sites and exhibit relatively small n-type SBHs. These findings provide an explanation for the experimentally observed low contact resistance at the In and Ag contacts on MoS_2 [232,233].

Given these clearer behaviours of n-type MoS_2, whose Fermi level lies closer to the CB edge, it would be beneficial to conduct a similar analysis for p-type and ambipolar WSe_2, with its Fermi level lying near mid-gap or towards its VB edge [234,253]. This study provides a comprehensive analysis of the more complex system of monolayer WSe_2 in contact with various metals, adopting different bonding methodologies, to explore suitable interfacial configurations and contacting metals that enable weakly pinned interfaces and tuneable SBHs.

6.2 Calculation Methods

The calculations employ the plane wave VASP code [254] with the GGA functional. The PAW pseudopotentials with a 500 eV cutoff energy are adopted. Energies are converged to 10^{-5} eV/atom, and forces to 0.02 eV/Å. No additional hybrid functional calculation of the band gap is performed, as the GGA band gap of monolayer WSe_2 of 1.64 eV is close enough to the experimental optical gap of approximately 1.65 eV. Moreover, using hybrid functionals could result in singularities in metallic systems. Spin–orbit coupling (SOC) is not incorporated in this study. Although including the strong SOC in WSe_2 would raise the VBM by 0.23 eV [234], this is small compared to the overall band gap. Meanwhile, SOC calculations are computationally expensive, especially considering the variety of interface supercells in this work. Besides, this study focuses on the chemical trend of SBHs, which is marginally affected by band splitting. The effects of dispersive interactions, such as vdW forces, are included using the optB88-vdW functional [255], as done previously [242,252]. This functional is chosen because it is optimised for heavier elements, whereas other treatments of vdW bonding mainly considered molecules of second-row elements [256].

The interface supercells comprise of a WSe_2 monolayer and four layers of metal, with approximately 20 Å of vacuum spacing maintained between the slabs. The in-plane lattice constants are fixed to the experimental lattice constant of WSe_2 (a = 3.29 Å), while vertical metal distances can relax, as metal work functions largely depend on atomic volumes and not on the lattice constants. The metals are presented in the FCC structures, with metal work functions ranging from 3.50 eV (Sc) to 5.65 eV (Pt) [177]. The details of the interface models are discussed in Sec. 6.3.1.

6.3 Results and Discussion

6.3.1 Interface Modelling

The metals in contact with the monolayer WSe_2 are classified into four groups based on the similarity of their lattice constants, in descending order: Sc, Hf, Zr, and In, with their primitive cells perfectly matched to that of WSe_2; Ag, Al, Ti, and Au with smaller radii; followed by noble metals Ru, Pd, Ir, and Pt; and finally Cu, Co, and Ni with the smallest lattice constants. Supercell models of the metal–WSe_2 interfaces describe metals in the same group sharing the same heterostructures, which generally

fall into three types: on-top, hollow, and Moiré (the lattice matching details are summarised in Table 6.1). In Fig. 6.1, the interfaces between noble metals and WSe_2 are presented as an example to explain the differences between the three contacting sites.

The first two cases consist of WSe_2 axes directly aligned along the metal(111) axes. For the on-top configuration, the W atoms are positioned above the centres of the triangles formed by the FCC, hexagonal closest packed (HCP), and top sites of the metal layers, and the Se atoms are placed right above the metal FCC, HCP, and top sites, as shown in Fig. 6.1(a). In this case, most Se atoms lie on top of the metal atoms from the first metal layer. The hollow configuration is similar to the on-top case but with a relative shift of the contacting layers in the xy plane. All W and Se atoms sit above the centres of the triangles formed by the metal FCC, HCP, and top sites, as depicted in Fig. 6.1(b). The on-top and hollow configurations have relatively small unit cells and interlayer spacing, which aim to achieve stronger chemisorptive interfaces.

The third case is the Moiré interface, where the $WSe_2(3 \times 3)$ lattice is rotated to a 13.9° orientation of the metal($\sqrt{13} \times \sqrt{13}$) lattice, as shown in Fig. 6.1(c). This configuration positions most selenide atoms away from any on-top sites, resulting in longer interfacial distances, to construct a physisorptive interface. There is a wide range of possible Moiré lattices, some with long lattice vectors, to enable close lattice matches if desired. Other metals, such as Ni, Co, and Cu with smaller lattice constants, or metals like Zr, Hf, and Sc with larger radii, could also form interfaces with similar on-top, hollow, or Moiré features. It should be noted that the lattice-matched interface between Ni and h-BN offers the clearest separation of chemisorptive and physisorptive sites, while this separation is only partial for less well-matched interfaces like metal–MoS_2 or metal–WSe_2.

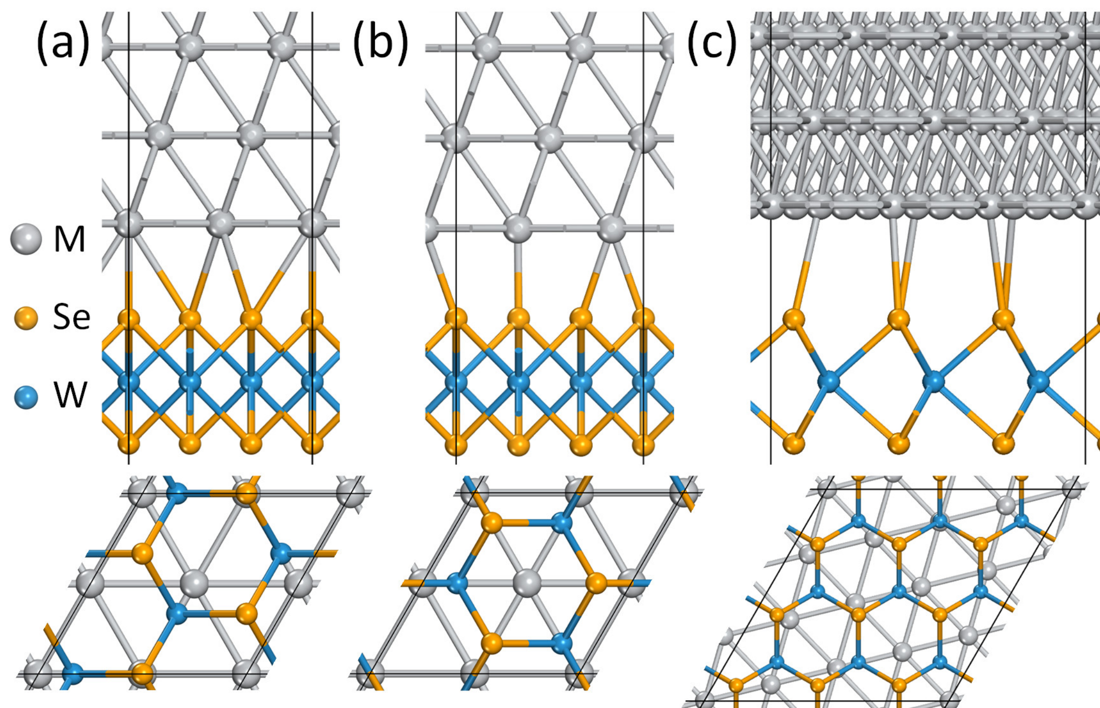


Figure 6.1 The atomic structures of interfaces between noble metals and WSe_2 in (a) on-top, (b) hollow, and (c) Moiré configurations. The on-top and hollow models employ the metal(2×2)–WSe_2($\sqrt{3} \times \sqrt{3}$) matching, while the Moire model uses metal($\sqrt{13} \times \sqrt{13}$)–$WSe_2$($3 \times 3$) matching.

Table 6.1 Lattice matching of metal–WSe_2 supercells in on-top, hollow, and Moiré configurations. Positive and negative strain values indicate that the metal has a larger or smaller lattice than the WSe2 slab, respectively, meaning that the metal is compressed or stretched. The corresponding minimum bond length, binding energy, and SBH are also listed.

Top/Hollow: metal(2 × 2)–WSe_2(√3 × √3); Moiré: metal(3 × 3)–WSe_2(√7 × √7) 19.1°					
	Method	Mismatch (%)	Min. bond length (Å)	$E_{binding}$ (eV/#WSe_2)	SBH (eV)
Ag	Top	3.57	2.96	−0.06	0.98
	Hollow	3.57	3.14	−0.05	0.80
	Moire	1.70	2.97	−0.24	1.07
Al	Top	0.54	3.09	−0.45	1.21
	Hollow	0.54	3.36	−0.45	1.16
	Moire	−1.27	3.17	−0.32	1.34
Ti	Top	2.30	2.60	−0.55	0.94
	Hollow	2.30	2.66	−0.52	0.97
	Moire	0.46	2.63	−0.53	1.01
Au	Top	3.85	3.18	−0.31	0.32
	Hollow	3.85	3.31	−0.30	0.32
	Moire	1.98	3.06	−0.31	0.65
Top/Hollow: metal(2 × 2)–WSe_2(√3 × √3); Moiré: metal(√13 × √13)–WSe_2(3 × 3) 13.9°					
	Method	Mismatch (%)	Min. bond length (Å)	$E_{binding}$ (eV/#WSe_2)	SBH (eV)
Ru	Top	−4.77	2.51	−0.03	0.91
	Hollow	−4.77	2.63	−0.02	0.90
	Moire	−0.88	2.55	−0.15	0.89
Pd	Top	−1.50	2.54	−0.31	0.50
	Hollow	−1.50	2.64	−0.32	0.53
	Moire	2.52	2.58	−0.62	0.54
Ir	Top	−3.51	2.58	−0.27	0.61
	Hollow	−3.51	2.67	−0.29	0.73
	Moire	0.44	2.61	−1.03	0.68
Pt	Top	−1.01	2.62	−1.02	0.52
	Hollow	−1.01	2.77	−1.02	0.56
	Moire	3.04	2.62	−0.82	0.52

Top/Hollow: metal($\sqrt{7} \times \sqrt{7}$)–WSe_2(2 × 2); Moiré: metal($\sqrt{12} \times \sqrt{12}$)–$WSe_2$($\sqrt{7} \times \sqrt{7}$) 10.9°					
	Method	Mismatch (%)	Min. bond length (Å)	$E_{binding}$ (eV/#WSe_2)	SBH (eV)
Cu	Top	3.29	2.67	−0.13	0.95
	Hollow	3.29	2.77	−0.04	0.99
	Moire	2.23	2.61	−0.21	0.96
Co	Top	0.43	2.31	−1.04	0.77
	Hollow	0.43	2.38	−1.04	0.77
	Moire	−0.60	2.32	−0.90	0.81
Ni	Top	−0.02	2.35	−1.13	0.71
	Hollow	−0.02	2.41	−1.13	0.70
	Moire	−1.04	2.33	−1.03	0.71
Top/Hollow: metal(3 × 3)–WSe_2(3 × 3)					
	Method	Mismatch (%)	Min. bond length (Å)	$E_{binding}$ (eV/#WSe_2)	SBH (eV)
Sc	Top	−0.43	2.81	−0.65	1.09
	Hollow	−0.43	3.06	−0.62	1.10
Hf	Top	−3.38	2.84	−0.31	0.93
	Hollow	−3.38	3.10	−0.17	0.91
Zr	Top	−2.20	2.85	−0.42	0.96
	Hollow	−2.20	3.09	−0.32	0.96
In	Top	3.38	3.37	−0.21	1.36
	Hollow	3.38	3.69	−0.21	1.22

6.3.2 Interface Stability and Bonding

To compare the stability of different interface configurations, the corresponding interface binding energy is calculated using the following formula:

$$E_{binding} = (E_{metal\text{-}WSe_2} - n \times E_{metal} - m \times E_{WSe_2})/\#WSe_2, \qquad (6.1)$$

where $E_{metal\text{-}WSe_2}$, E_{metal}, and E_{WSe_2} are the total energy of the contact supercell, the metal(1 × 1) surface slab, and the WSe_2(1 × 1) surface slab, respectively. $\#WSe_2$ represents the number of WSe_2 formula units. n and m are the number of isolated slabs in these supercells. A lower binding energy indicates a more stable interface. Table 6.1 provides a summary of the binding energy associated with various metal–WSe_2 contacts, alongside the corresponding interfacial bond length, which is a crucial factor for interface geometry.

The chemical trends of interfacial energy and interfacial bond length for each interfacial geometry are shown in Fig. 6.2 against the metal work function, with the most stable configurations labelled by solid symbols. Fig. 6.2(a) reveals that most of the metals are stabilised in their on-top geometries, as indicated by the solid black squares. The hollow geometries are also possible and exhibit similar or slightly less stability compared to the on-top sites for all metal cases. Generally, Moiré sites are favoured for noble metals and tend to have low binding energies of approximately −0.25 eV per WSe_2 unit.

Considering only the most stable configuration for each metal, it is observed that the (negative) binding energies tend to peak for more electropositive metals like Co, Ni, and Pt on the right. Metals whose Fermi energies lie around mid-gap, such as Ru, appear to have the lowest binding energy. The binding energies then increase negatively for reactive metals like Sc, Hf, and Zr, as was the case in the MoS_2 system, with Sc showing the highest binding energy among these metals.

Figure 6.2(b) illustrates the variation of the interfacial bond length with the metal work function. The physical separation between metal and WSe_2 is larger compared to the case of MoS_2, which can be explained by the larger atomic size of WSe_2. Metals that prefer their on-top sites have metal–Se bond lengths that vary from approximately 2.3 Å for the more stable interfaces of Ni and Co, to roughly 2.8 Å for the reactive metals (Sc, Hf, and Zr). In and Al exhibit unusually large bond lengths over 3 Å. The bond lengths of hollow sites are larger compared to the on-top sites, which is

reasonable considering there are three interfacial bonds per selenide atom. The interfaces stabilised in Moiré sites generally have an interlayer bonding of approximately 2.6 Å.

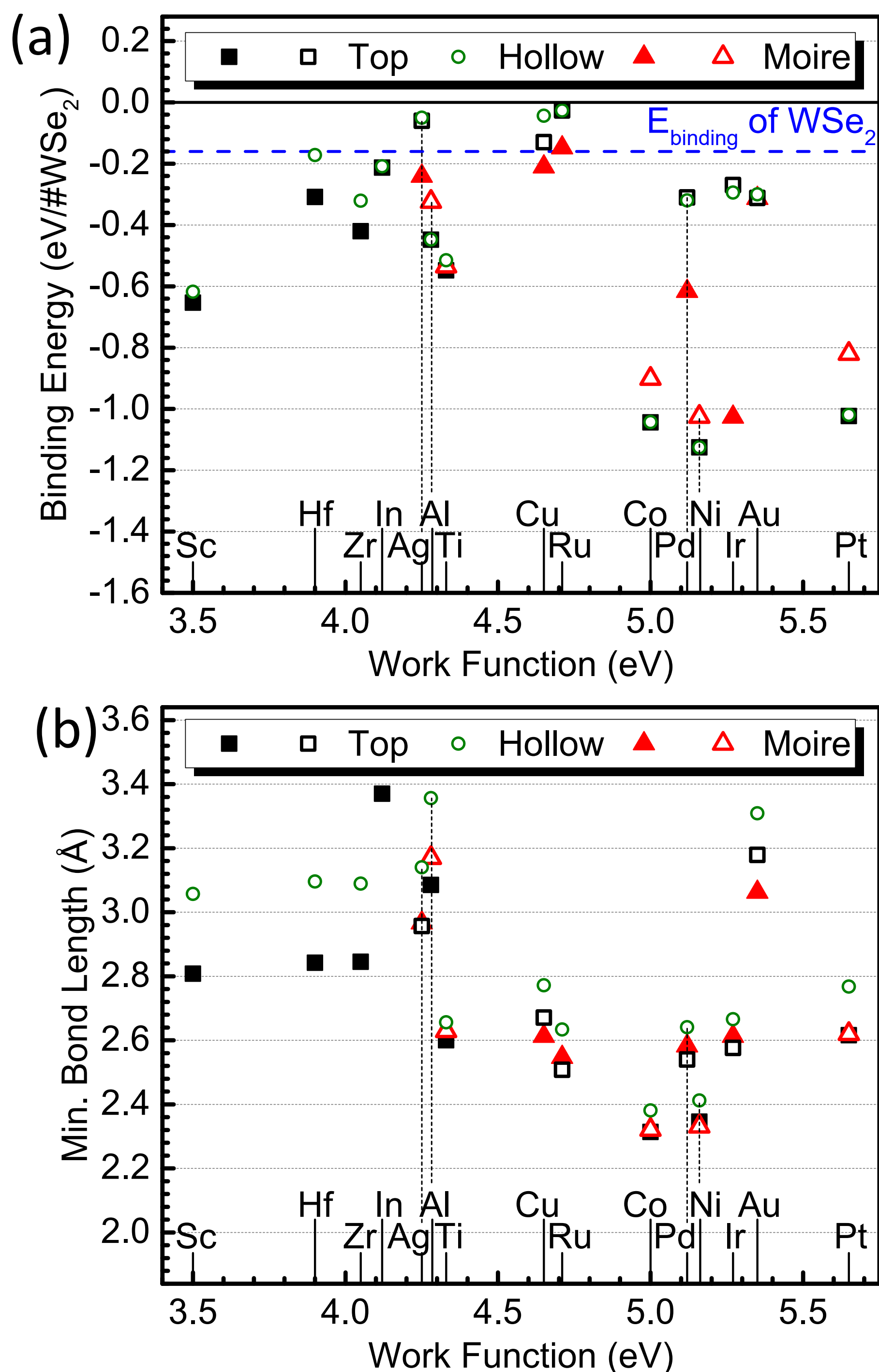


Figure 6.2 (a) Binding energy and (b) minimum bond length for different metal–WSe_2 interfaces with respect to the metal work function [177]. Filled and open symbols represent the more stable and less stable configurations, respectively. The binding energy of the monolayer WSe_2 is labelled with the blue dashed line for reference.

For a clearer understanding of the correlation between interfacial parameters, the data from Fig. 6.2 are combined into a single plot of binding energy vs interfacial bond length, as shown in Fig. 6.3. This enables a more straightforward categorisation of chemisorptive and physisorptive interfaces. While there is not an absolute standard for such classification, it is generally accepted that interfaces with larger bond lengths (comparable to the typical vdW gap of 3 Å) or smaller binding energies (approaching the binding energy between WSe_2 monolayers of 0.16 eV) can be considered as physisorptive. On the other hand, interfaces with smaller bond lengths and more negative binding energies are typically viewed as chemisorptive. It is worth noting that these classifications are empirical and should be interpreted within the context of the specific system under investigation.

Figure 6.3 demonstrates the correlation between interfacial binding energy and bond length, with shorter distances generally yielding more negative binding energies. Co and Ni exhibit the smallest bond lengths and the most negative binding energies in their on-top geometries, indicating the formation of chemisorptive interfaces. A range of metals, mostly noble metals, follow with larger bond lengths (~2.6 Å) but varying binding energies: Ir and Pt have the highest binding energies, comparable to those of Co and Ni, and are classified as chemisorptive; Ti and Pd have binding energies of approximately 0.6 eV, and are considered chemisorptive; Cu and Ru have low binding energies that are close to the binding energy between WSe_2 monolayers, which correspond to physisorptive interfaces. The reactive metals Sc, Hf, and Zr exhibit stable on-top structures with bond lengths of approximately 2.8 Å. These metals are chemically distinct from other metals and will react with WSe_2. Therefore, they are not classified into the two groups and are represented as blue squares. Finally, the noble metals Ag and Au, together with the s-electron metals In and Al, have the longest interfacial bonds. These contacts, except for Al, exhibit relatively low binding energies and interlayer bonds exceeding 3 Å, which demonstrate clear vdW behaviour. Consequently, these contacts are characterised as physisorptive. Al with its on-top structure has moderate binding energy and is considered a borderline case. It is important to note that all these metals are adopted in their physisorptive Moiré geometries for the MoS_2 interfaces [252]. However, for the WSe_2 interface, Al and In retain their large bond length but adopt the on-top geometries. This is reasonable, as the lattice constant of WSe_2 is 3% larger than that of MoS_2, leading to different lattice mismatches when aligning the lattices.

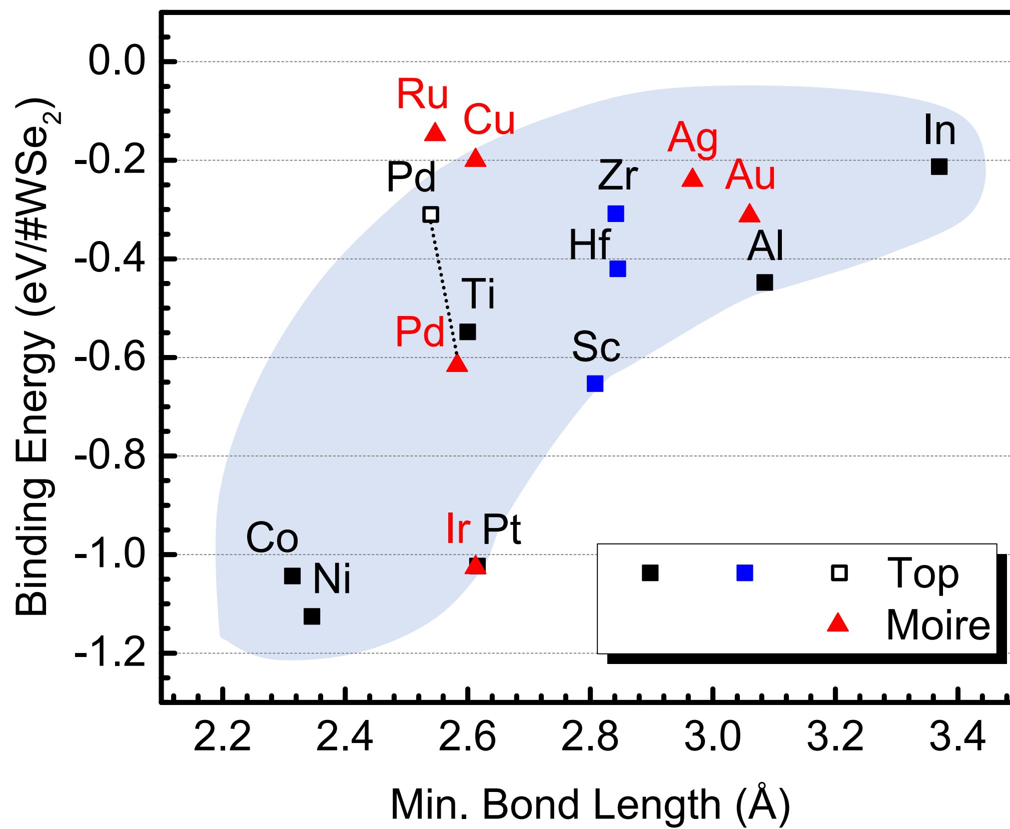


Figure 6.3 Binding energy vs minimum interfacial bond length plot for the most stable configurations. The less stable on-top site of Pd is also plotted in the open square for comparison. Note that the low work function metals Sc, Zr, and Hf (filled blue squares) form reactive sites.

The categorised interfaces are summarised in Fig. 6.4. It can be deduced that noble metals are more easily stabilised in their Moiré pattern, and constructing the Moiré lattice increases the likelihood of obtaining physisorptive bonding states.

The cases of In and Al with the on-top structure stand out among the physisorptive interfaces, especially In, with an unusually large bond length of 3.37 Å, and a small binding energy of −0.21 eV, demonstrating remarkable physisorptive features without being a noble metal. These findings are consistent with previous experimental and theoretical works: Wang *et al.* [232] observed an ultra-clean interface between In and WSe_2, with a large spacing of 2.94 Å; Liu *et al.* [235] simulated the Al–WSe_2 system and the relaxed contact region has an interlayer gap of 3.22 Å. Similar behaviours of In and Al in their Moiré configurations are also observed in a previous study of MoS_2 [252], consequently, it can be deduced that the bond length of In and Al is a distinctive feature irrespective of the interfacial geometry.

Additional remarks are necessary regarding the cases of Pd and Ir, as they remain stabilised in their Moiré configurations but display chemisorptive properties. The unusual behaviours of the abovementioned metals will become clearer once the interface electronic properties are analysed in Sec. 6.3.3.

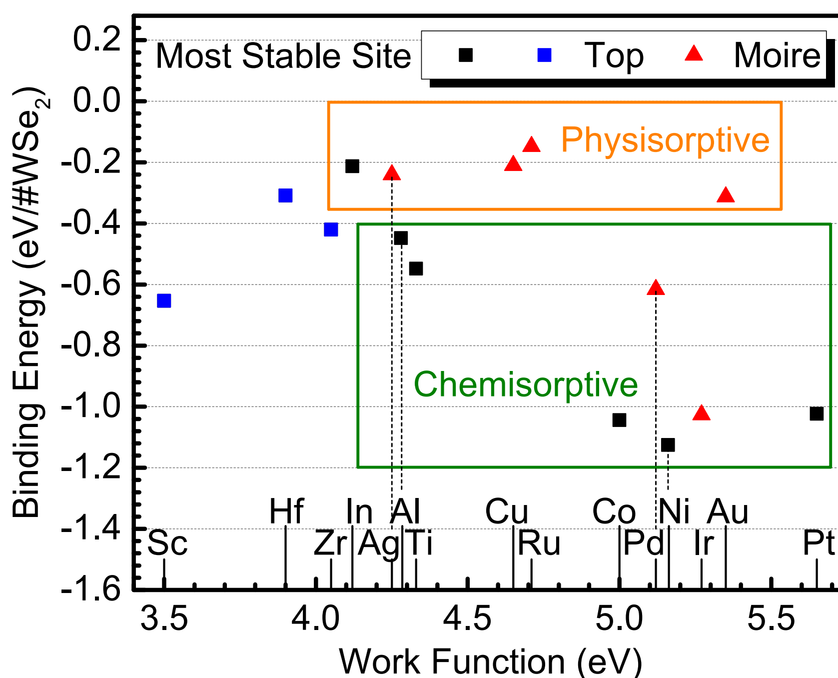


Figure 6.4 Binding energies of the most stable sites, classified into the physisorptive and chemisorptive groups. The low work function metals Zr, Hf, and Sc (blue squares) are reactive and do not fit into either group.

6.3.3 Schottky Barrier Heights and Interfacial States

To investigate whether the superior structural features of physisorptive Moiré contacts can be utilised to improve device performance, the SBHs of metal–WSe_2 interfaces are calculated, as shown in Fig. 6.5. The p-type SBH is extracted from the energy separation between the Fermi level of the contact metal and the VBM of WSe_2 using the core-level method [157,158]. Fig. 6.5(a) displays the SBH values for various configurations, including both stable and metastable sites. In Fig. 6.5(b), the SBH values for the most stable interface configurations are shown, which are separated into two categories according to their different bonding geometries.

The metal–WSe_2 contacts adopted in their on-top and Moiré configurations primarily fall on two separate lines. The SBHs of the chemisorptive on-top interfaces (excluding In and Al) lie on a line with slope $S = 0.24$, while those of the physisorptive Moiré interfaces lie on a line with slope $S = 0.45$. The Fermi level depinning effect induced by the Moiré sites is largely similar to that found for metals on MoS_2 [252], except that the Moiré line shifts down towards the VB. Although the SBHs follow a clear linear chemical trend, the correlation of binding energy and bond length with the metal work function is not as straightforward. This is because interface energy and structural trends are more closely connected to properties like electron configuration, rather than the metal work function.

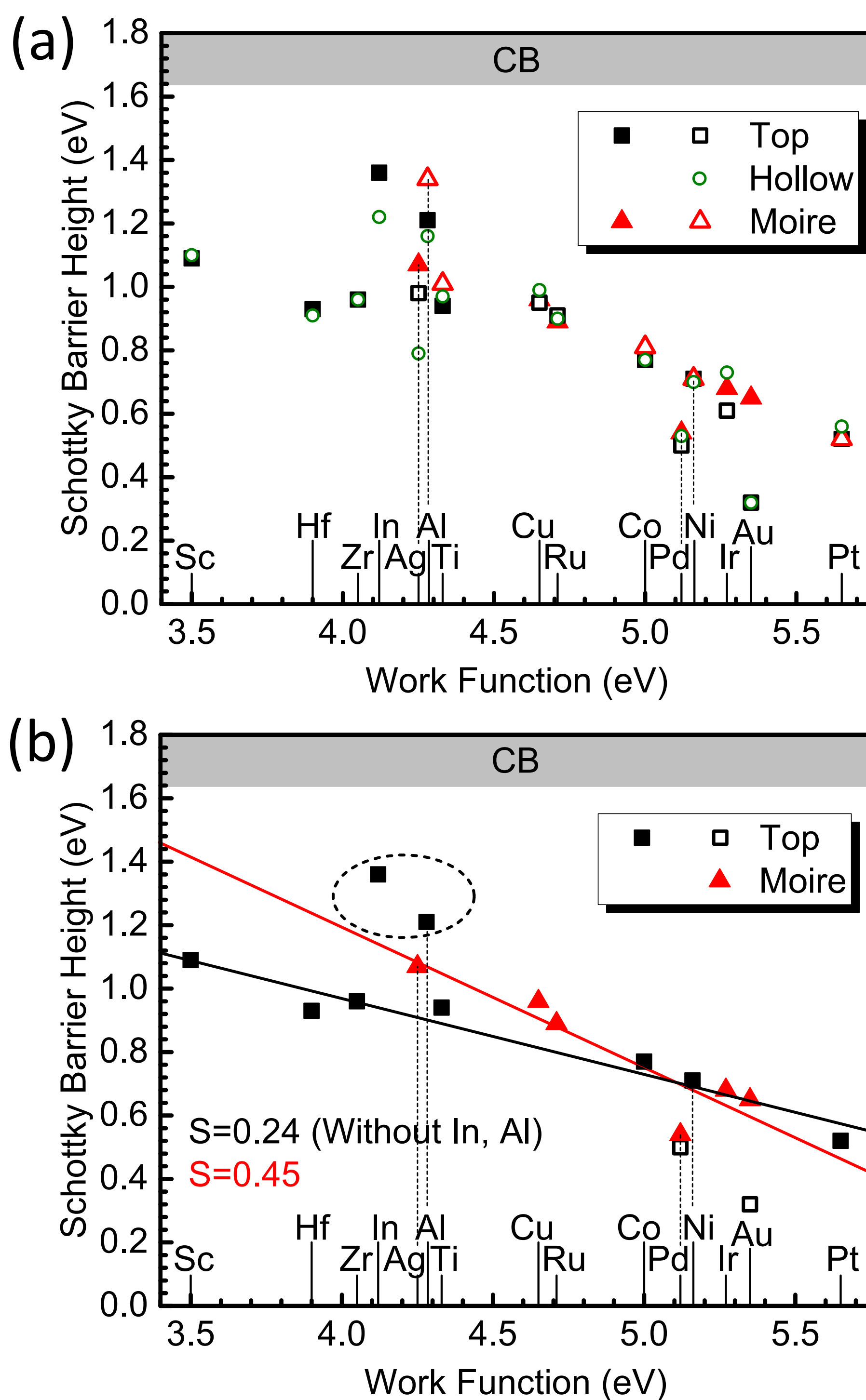


Figure 6.5 (a) Calculated p-type SBHs for different metal–WSe_2 interfaces vs metal work function. The filled and open symbols represent the more and less stable configurations, respectively. (b) The most stable configuration for each metal. SBHs of the less stable on-top sites for the Pd–WSe_2 and Au–WSe_2 interfaces are also plotted for comparison. Fermi level depinning is observed from the on-top sites to the Moiré sites, with the pinning factor increasing from 0.24 to 0.45. The special cases of In and Al (circled) are excluded from the fitting.

The cases of Pd and Ir require special remarks. Pd shows a weakly bonded on-top site (−0.31 eV/#WSe_2) and a more strongly bonded Moiré site (−0.62 eV/#WSe_2), as seen in Fig. 6.2(a). It is important to emphasise that different interface structures exhibit similar small SBH values of ~0.5 eV, which fall below the usual trend and are highly suitable for p-type devices. Ir represents a more extreme case, with an even more stable Moiré geometry and comparable SBHs for different configurations.

To further evaluate the Moiré contacts of these two metals and explain their unique structural and electronic behaviours, the PDOS is plotted in Figs. 6.6(b) and 6.6(c). In Fig. 6.6(a), the PDOS of monolayer WSe_2 without contacts is displayed as a reference, exhibiting its intrinsic electronic characteristics. The Fermi level of the undoped WSe_2 lies at the middle of the band gap, and there are no states near E_F. After contacting with Ir and Pd, the Fermi level shifts towards the VB, as shown in Figs. 6.6(b) and 6.6(c), indicating that WSe_2 is p-doped by Ir and Pd. For the Ir contact, the PDOS spreads all over the WSe_2 band gap, demonstrating that the VBs and CBs of the original monolayer WSe_2 are strongly disrupted, representing the metallisation of WSe_2 in the contact region. Prominent orbital overlaps are also observed between Pd and WSe_2, confirming the formation of strong interfacial bonds.

Additionally, a distinct hybridisation between the d orbitals of the metals and W atoms is observed, despite being separated by a Se layer, as evident from Figs. 6.6(b) and 6.6(c). Taking the Pd–WSe_2 contact as an example, the PDOS of Pd d orbitals and W d orbitals share some similar features, one of which is the presence of double peaks at approximately −1.2 eV. Previous theoretical works also identified that the hybridisation between d orbitals from the contact metals and W atoms is responsible for the interfacial interaction, resulting in better electron injection and lower contact resistance [227,234,235]. Notably, Ir has an open d-shell and is capable of directly hybridising with WSe_2. On the other hand, although Pd has full 4d orbitals, most of its 4d states are located at high energies and within the lower half of the original WSe_2 band gap. Consequently, Pd can effectively hybridise with the WSe_2 VB states.

To further investigate the electronic properties of different interface configurations, the wavefunction diagrams of Ir and Pd at the on-top and Moiré sites are provided in Figs. 6.6(d) and 6.6(e). From the wavefunction plots, the orbital filling condition of interfacial atoms at the Fermi level can be deduced. For both Ir and Pd, the relaxed interface models indicate that constructing the Moiré interfaces can form more

interfacial bonds without significantly increasing the interlayer distance. The slightly fluctuated nature of the Moiré interfaces, as compared to the flatter on-top interfaces, confirms the formation of stronger interaction. Additionally, the wavefunction diagrams of the Moiré configurations reveal a substantial charge density at the interface region. This observation further substantiates the stronger orbital hybridisation and consequently indicates the formation of covalent bonds at these interfaces. In contrast, the on-top sites exhibit less pronounced charge transfer, which can be attributed to the relative position of interfacial atoms causing limited overlap between orbitals.

Fang *et al.* [224] have achieved high-performance monolayer WSe_2 p-type FETs with Pd serving as the contact metal for the source and drain, benefitting from the lowered contact resistance for hole injection through the use of a low SBH at Pd contacts, as found in this work. It is presumed that the Moiré site dominates experimentally, as this configuration has much lower binding energy.

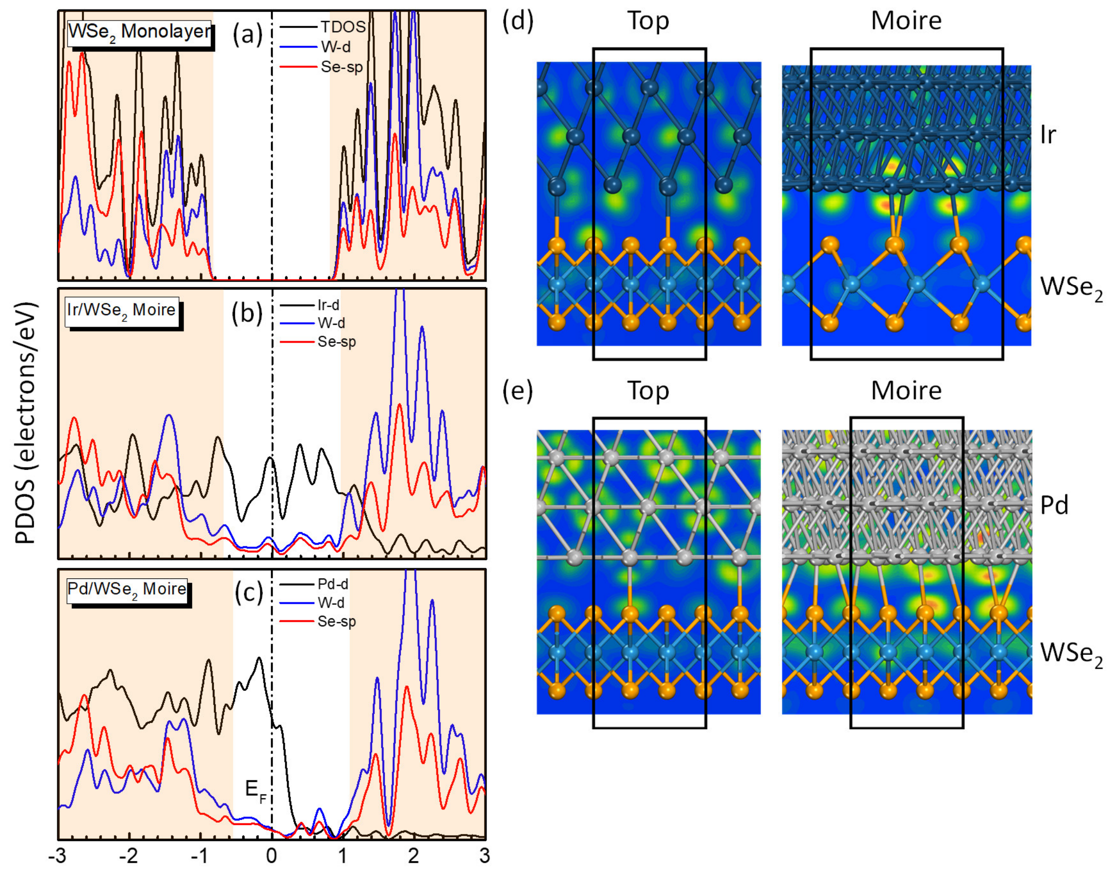


Figure 6.6 PDOS of (a) monolayer WSe_2, as well as the Moiré configurations of (b) Ir–WSe_2 and (c) Pd–WSe_2. The CB and VB of WSe_2 are shaded in orange. The Fermi level E_F of the monolayer WSe_2, as well as the E_F of Ir and Pd at the contacts, are labelled by the dash-dot lines. The wavefunction diagrams with energy around the metal Fermi energy are compared for different bonding sites of (d) Ir–WSe_2 and (e) Al–WSe_2.

Au exhibits very similar low binding energies in the physisorptive range for all three configurations [Fig. 6.2(a)], with the energy difference within 0.01 eV per WSe_2. It is important to stress that there is a recognised drop of p-type SBH from the Moiré site (0.65 eV) to the on-top site (0.32 eV).

In Fig. 6.7, the PDOS plots of Au–WSe_2 in different bonding sites are presented. There are much fewer overlap interface states distributed in the original WSe_2 band gap compared to those in the Ir and Pd Moiré contacts. The VBs and CBs are well-preserved, with most of the WSe_2 states in the Au–WSe_2 system remaining identifiable as in pure WSe_2 [Fig. 6.6(a)]. The PDOS analysis of the Au–WSe_2 contacts demonstrates the lack of orbital overlap and weakly interacted interfaces, consistent with the calculated binding energies of Au–WSe_2, which fall in the physisorptive region.

In both on-top and Moiré contacts, their Fermi levels shift downward in energy and form p-type contacts, while the E_F of the on-top contact is closer to the original VB edge, introducing a much lower p-type SBH. The remarkable reduction in the p-type SBH can be explained by the specific orbital overlap condition at different interfaces. As shown in Fig. 6.7(a), the Se sp orbitals, which play a more prominent role in the VB, coincide and hybridise with Au d orbitals. On the other hand, Fig. 6.7(b) reveals that the W d orbitals, which retain a higher weight at the bottom of CB, share more comparable characteristics with the Au d orbitals. Therefore, it can be deduced that there is a better alignment of Au d and Se sp orbitals for the on-top site, contributing more to the VB and resulting in more efficient p-type doping. In contrast, the overlap between Au d and W d orbitals is favoured at the Moiré site, which makes a certain degree of contribution to the CB.

Although a low p-type SBH has been observed for the on-top configuration of the Au–WSe_2 interface, it is important to note that this configuration is only metastable. The similar binding energies of the on-top and Moiré configurations suggest that either site could potentially occur. Overall, Pd remains the best choice of metal for p-type FETs to minimise contact resistance.

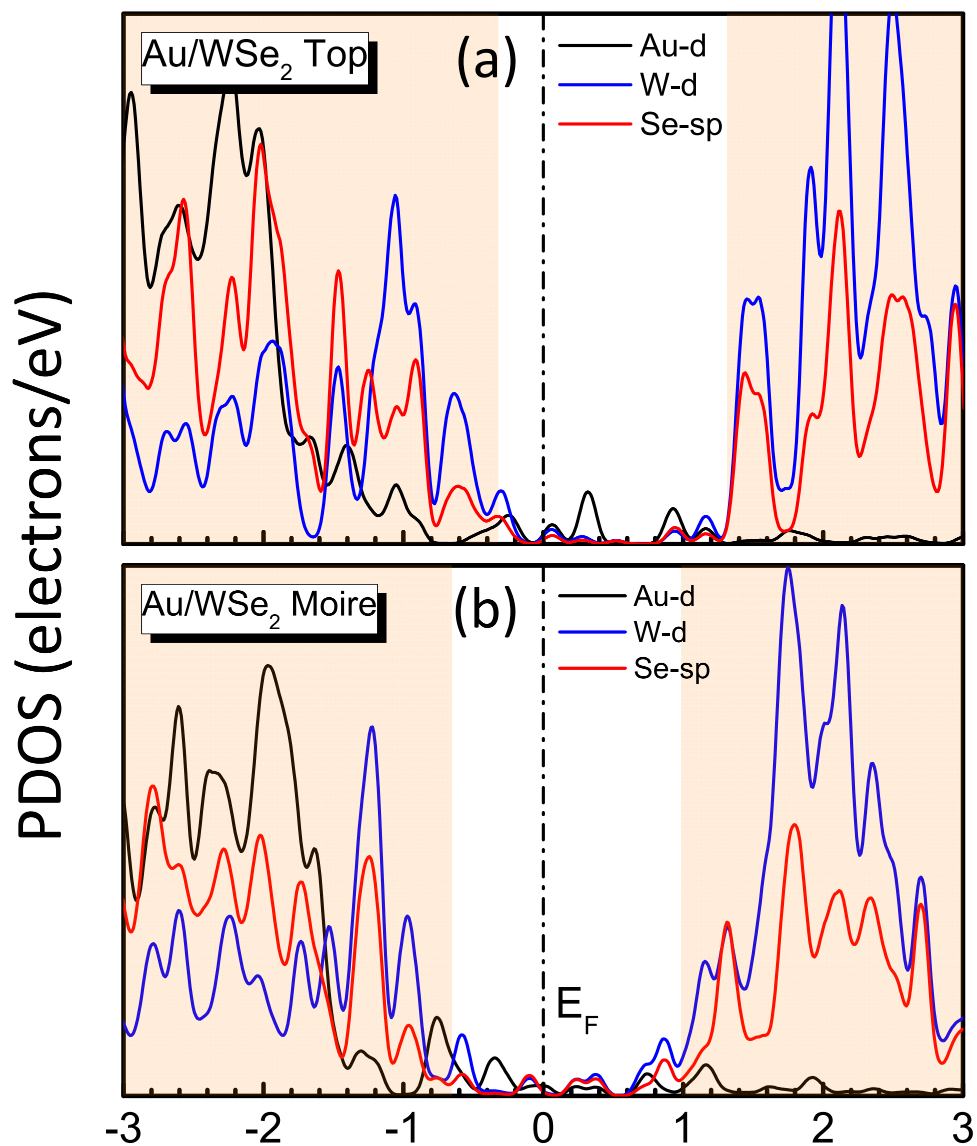


Figure 6.7 PDOS of Au–WSe_2 in (a) on-top and (b) Moire configurations. The CB and VB of WSe_2 are shaded in orange, and the E_F of Au is labelled by the dash-dot lines.

As a complementary component to p-type FETs, an ideal n-type FET should also have low contact resistance by exhibiting a low n-type SBH. Interestingly, s-electron metals In and Al deviate from the trend in Fig. 6.5(b), with In displaying the lowest n-type SBH of only 0.24 eV. From the PDOS plots [Figs. 6.8(a) and 6.8(b)], it can be observed that the perturbation of In and Al contacts leads to the broadening of the WSe_2 CB, while the VB largely retains its original features. However, apart from the slightly disrupted CB, only a few states are introduced in the original WSe_2 band gap, indicating minimal orbital overlap.

It is worth noting that Al lacks d orbitals, and In has fully occupied d orbitals which locate deep in energy. As a result, their s and p orbitals are responsible for interacting with WSe_2. The absence of effective metal d orbitals leads to subtle orbital overlap compared to previous cases, resulting in a less perturbed electronic structure of WSe_2, as evidenced by the PDOS plots. The weakly interacted interfaces revealed by the PDOS plots are consistent with the anomalous low binding energies of Al and In.

The wavefunction analysis of In–WSe_2 and Al–WSe_2 in their on-top sites are depicted in Figs. 6.8(c) and 6.8(d). Both In and Al exhibit negligible structural distortion at their relaxed interfaces. The vertical interlayer distance of over 3 Å between the Se layer and the metal slab indicates a typical vdW separation. There is no observable orbital hybridisation at the interfaces as revealed by the wavefunction plots. Besides, the charge transfer from the metal side to WSe_2 is negligible. It is believed that the charges distributed around the WSe_2 atoms originate from the decay of metal wavefunction. These observations suggest that the on-top configurations of In and Al are less interrupted and weakly bonded.

The low DOS at E_F, as shown by the PDOS plots [Figs. 6.8(a) and 6.8(b)], along with the rapid attenuation of MIGS into the WSe_2 layer [Figs. 6.8(c) and 6.8(d)], imply a considerably depressed or even vanishing FLP. The weakly pinned interfaces of In–WSe_2 and Al–WSe_2 with on-top configurations account for the unusual behaviours of n-type SBHs observed in Fig. 6.5: the SBH values lie above the standard trend of the other on-top contacts and closer to the trend of Moiré contacts.

In contrast to the calculated low n-type SBH for the Al–WSe_2 system in this work, Liu *et al.* [235] reported that Al forms high-resistance Schottky contacts with WSe_2. The contradiction can be explained by the weak doping capability of Al due to the absence of d orbitals, as well as the wide tunnel barrier that restricts electron injection efficiency. Conversely, In is superior in terms of its n-type SBH, and the low

contact resistance observed in experiments suggests that the large interlayer distance does not compromise the contact performance. Instead, it facilitates the formation of a clean interface with minimal defect creation [232].

It is worth noting that Kong *et al.* [238] used Au contact with two different process approaches resulting in different p-type SBHs for the doping-free complementary WSe_2 circuits. Nevertheless, based on the results discussed earlier, a more advantageous strategy is to use In for n-type WSe_2 FETs and Pd for p-type WSe_2 FETs. This approach would likely yield the best overall performance for ambipolar devices.

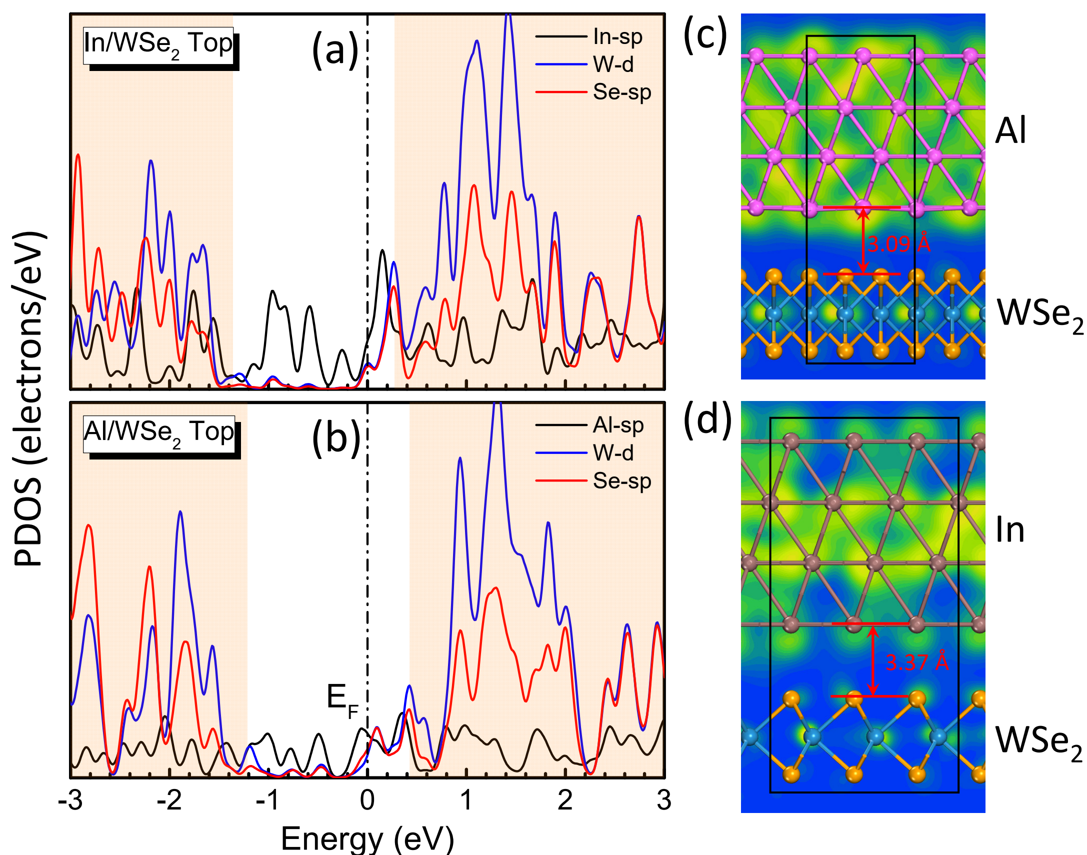


Figure 6.8 PDOS of the on-top configurations of (a) In–WSe_2 and (b) Al–WSe_2. The CB and VB of WSe_2 are shaded in orange, and the E_F of In and Al are labelled by the dash-dot lines. The wavefunction diagrams with energy around the metal Fermi energy are compared for the on-top configurations of (c) In–WSe_2 and (d) Al–WSe_2, with interlayer distances labelled.

6.4 Conclusions

In conclusion, p-type WSe_2, similar to n-type MoS_2, demonstrates the separation of chemisorptive and physisorptive interfaces concerning the contact metals and their corresponding configurations. The majority of the stable contacts for both compounds are chemisorptive, which exhibit a smaller pinning factor S. The physisorptive interfaces are favoured for noble metals and Moiré configurations, which possess a weaker pinning effect. For WSe_2, In and Al change their stable geometries from Moiré to on-top while maintaining their significantly large bond lengths and relatively low n-type SBHs. Pd presents unusual behaviours when in contact with WSe_2, forming a relatively strong bond with a physisorptive geometry and exhibiting the lowest p-type SBH. Au shows an even lower p-type SBH, but this occurs at a metastable site. This work provides valuable theoretical insights for constructing low-resistance metal–WSe_2 contacts, which could be instrumental in the design of novel 2D semiconductor devices.

CHAPTER 7: CONCLUSIONS AND OUTLOOK

This thesis presents a comprehensive investigation into various materials and contacts used in the development of MOSFETs, with a focus on understanding the underlying physics of metal–semiconductor junctions and heterojunctions. Density functional calculations are employed to provide valuable insights into MOSFET design and functionality.

Chapter 3 studies the structural, electronic, and optical properties of nine SiO_2 polymorphs, utilising both traditional GGA and state-of-the-art sX functionals, to assess the validity of various XC functionals in predicting the properties of different silica phases. While GGA underestimates band gap values, it offers valuable guidance on the optical properties of silica polymorphs. Conversely, sX provides accurate band gap predictions but underestimates the dielectric constant.

Chapter 4 systematically studies the SBHs at defect-free metal–oxide interfaces (ZnO, CdO, MgO, and SrO) with various terminations using the GGA + U method. Different oxide terminations exhibit similar pinning strength, and a significant upward shift in SBH values occurs when the oxide termination changes from polar to non-polar. The study concludes that MIGS is a reliable model for describing pinning effects and demonstrates the applicability of adding U values to O p orbitals in addition to their traditional applications on the metal d orbitals.

Chapter 5 investigates the electronic properties of CaF_2 bulk and interfaces. The band structures of CaF_2 are investigated with both the hybrid functional and the GGA + U scheme. The GGA + U method is further applied to study Si–CaF_2 and metal–CaF_2 interfaces with different CaF_2 terminations. Type-I band alignment is found at the Si–CaF_2 contacts, while the metal–CaF_2 contacts have weak pinning factors, consistent with the MIGS predictions. Different terminations of CaF_2 share similar pinning factors, which is similar to previous findings in the metal–oxide contacts. Following the calculations of SiO_2 properties from the previous chapter, the calculations on CaF_2 band structures further demonstrate the validity of the sX approach in material with a wide band gap. Meanwhile, the study extends the application of the

GGA + U scheme to fluoride and proves that the U-type treatment can be a useful tool in studying such fluoride interfaces and contacts.

Chapter 6 studies the metal contacts on the 2D TMDs, extending calculations from n-type MoS_2 to p-type and ambipolar WSe_2. The construction of Moiré configurations enables the achievement of physisorptive interfaces, inducing the Fermi level depinning effect. Several noteworthy metal contacts with unusual interface properties are identified, exhibiting low n-type SBHs (In, Al) or p-type SBHs (Pd, Au). This study provides valuable theoretical insights for constructing low-resistance metal–TMD contacts and guidance for the design of novel 2D semiconductor devices.

As 2D materials continue to attract attention, prioritising the exploration of novel 2D materials and heterostructures is essential for future research. Gaining a deeper understanding of their unique properties and behaviours could lead to the development of innovative applications and devices with unparalleled performance characteristics. As mentioned in Chapter 6, contact resistance is a significant challenge in the development of 2D TMD-based devices, which can be addressed by reducing the SBHs at the interface. MIGS is considered the primary cause of pinning; therefore, it is logical to seek contacts that suppress MIGS at the interface to achieve weak pinning. This work suggests constructing such interfaces by achieving rotational Moiré contacts, while another strategy involves using semimetal–TMD contacts. Since semimetals exhibit near-zero DOS at the Fermi level, the orbital overlap will be minimal, and fewer MIGS will be induced. As, Sb, and Bi are well-known Group V semimetals and could be potential candidates. Recent experimental work demonstrated excellent Ohmic contacts between Bi and various monolayer semiconductors, including MoS_2, WS_2, and WSe_2. Consequently, it would be worthwhile to perform theoretical calculations on these Group V semimetals in contact with WSe_2, complementing this study on WSe_2-based systems.

Another potential area for future research is to broaden the calculations on metal–WSe_2 contacts to HfS_2-based systems. HfS_2 is a relatively unexplored TMD compared to MoS_2 and WSe_2. However, it has a larger EA, making it an appealing candidate for achieving low contact resistance in n-type carrier transport. Investigating metal–HfS_2 interfaces with Moiré configurations could complement the existing knowledge on metal–TMD contacts.

REFERENCES

[1] J. Bardeen and W. H. Brattain, Phys. Rev. 74, 230 (1948).
[2] G. E. Moore, Electronics 38, (1965).
[3] K. K. Ng, Complete Guide to Semiconductor Devices, 2nd ed. (Wiley/IEEE Press, 2002).
[4] M. A. Khan, Q. Chen, C. J. Sun, J. W. Yang, M. Blasingame, M. S. Shur, and H. Park, Appl. Phys. Lett. 68, 514 (1996).
[5] D. Kahng and M. M. Atalla, in IRE Solid State Device Research Conference (1960).
[6] F.-C. Hsu, R. S. Muller, C. Hu, and P.-K. Ko, IEEE Trans. Electron Devices 30, 1354 (1983).
[7] S. M. Sze and M. K. Lee, Semiconductor Devices: Physics and Technology, 3rd ed. (John Wiley & Sons, 2012).
[8] V. K. Khanna, in Integrated Nanoelectronics: Nanoscale CMOS, Post-CMOS and Allied Nanotechnologies (Springer India, 2016), pp. 73–93.
[9] L. D. Yau, Solid-State Electron. 17, 1059 (1974).
[10] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, IEEE J. Solid-State Circuits 9, 256 (1974).
[11] S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, IEEE Electron Device Lett. 18, 209 (1997).
[12] J. Robertson, Rep. Prog. Phys. 69, 327 (2005).
[13] C.-H. Lee, S.-H. Hur, Y.-C. Shin, J.-H. Choi, D.-G. Park, and K. Kim, Appl. Phys. Lett. 86, 152908 (2005).
[14] H. J. Cho, Y. D. Kim, D. S. Park, E. Lee, C. H. Park, J. S. Jang, K. B. Lee, H. W. Kim, Y. J. Ki, I. K. Han, and Y. W. Song, Solid-State Electron. 51, 1529 (2007).
[15] K. Yim, Y. Yong, J. Lee, K. Lee, H.-H. Nahm, J. Yoo, C. Lee, C. Seong Hwang, and S. Han, NPG Asia Mater. 7, e190 (2015).
[16] P. M. Zeitzoff and H. R. Huff, AIP Conf. Proc. 788, 203 (2005).
[17] Y. Kamata, Mater. Today 11, 30 (2008).
[18] S. Oktyabrsky and P. Ye, editors, Fundamentals of III-V Semiconductor MOSFETs (Springer US, 2010).
[19] D. Jena, Proc. IEEE 101, 1585 (2013).
[20] F. Schwierz, J. Pezoldt, and R. Granzner, Nanoscale 7, 8261 (2015).
[21] Y. Liu, N. O. Weiss, X. Duan, H.-C. Cheng, Y. Huang, and X. Duan, Nat. Rev. Mater. 1, 1 (2016).
[22] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, in 2008 IEEE International Electron Devices Meeting (2008), pp. 1–3.
[23] H. Mertens, R. Ritzenthaler, V. Pena, G. Santoro, K. Kenis, A. Schulze, E. D. Litta, S. A. Chew, K. Devriendt, R. Chiarella, S. Demuynck, D. Yakimets, D. Jang, A. Spessot, G. Eneman, A. Dangol, P. Lagrain, H. Bender, S. Sun, M. Korolik, D. Kioussis, M. Kim, K.-. H. Bu, S. C. Chen, M. Cogorno, J. Devrajan, J. Machillot, N. Yoshida, N. Kim, K. Barla, D. Mocuta, and N. Horiguchi, in IEEE International Electron Devices Meeting (IEDM) (2017), pp. 37.4.1-37.4.4.
[24] S. Zhu, J. Chen, M.-F. Li, S. J. Lee, J. Singh, C. X. Zhu, A. Du, C. H. Tung, A. Chin, and D. L. Kwong, IEEE Electron Device Lett. 25, 565 (2004).
[25] W. Schottky, Z. Physik 113, 367 (1939).
[26] N. F. Mott, Proc. R. Soc. Lond. A 171, 27 (1939).
[27] W. E. Meyerhof, Phys. Rev. 71, 727 (1947).

[28] W. Mönch, Rep. Prog. Phys. 53, 221 (1990).
[29] L. Brillson, Surf. Sci. Rep. 2, 123 (1982).
[30] J. Robertson and B. Falabretti, J. Appl. Phys. 100, 014111 (2006).
[31] J. Robertson, J. Vac. Sci. Technol. A 31, 050821 (2013).
[32] A. M. Cowley and S. M. Sze, J. Appl. Phys. 36, 3212 (1965).
[33] V. Heine, Phys. Rev. 138, A1689 (1965).
[34] S. G. Louie and M. L. Cohen, Phys. Rev. B 13, 2461 (1976).
[35] S. G. Louie, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B 15, 2154 (1977).
[36] J. Tersoff, Phys. Rev. Lett. 52, 465 (1984).
[37] J. Tersoff, Phys. Rev. B 30, 4874 (1984).
[38] W. Mönch, Phys. Rev. Lett. 58, 1260 (1987).
[39] W. Mönch, Appl. Surf. Sci. 92, 367 (1996).
[40] J. Robertson, J. Vac. Sci. Technol. B 18, 1785 (2000).
[41] C. B. Duke, J. Vac. Sci. Technol. B 3, 1170 (1985).
[42] K. Stiles and A. Kahn, Phys. Rev. Lett. 60, 440 (1988).
[43] S. Aftab and M. Z. Iqbal, J. Mater. Chem. C 10, 17414 (2022).
[44] K. Choudhary, K. F. Garrity, S. T. Hartman, G. Pilania, and F. Tavazza, Phys. Rev. Mater. 7, 014009 (2023).
[45] M. M. Furchi, A. Pospischil, F. Libisch, J. Burgdörfer, and T. Mueller, Nano Lett. 14, 4785 (2014).
[46] M. Shanmugam, R. Jacobs-Gedrim, E. Sang Song, and B. Yu, Nanoscale 6, 12682 (2014).
[47] O. M. Nayfeh, C. N. Chleirigh, J. Hennessy, L. Gomez, J. L. Hoyt, and D. A. Antoniadis, IEEE Electron Device Lett. 29, 1074 (2008).
[48] L. Wang, E. Yu, Y. Taur, and P. Asbeck, IEEE Electron Device Lett. 31, 431 (2010).
[49] X. Yan, C. Liu, C. Li, W. Bao, S. Ding, D. W. Zhang, and P. Zhou, Small 13, 1701478 (2017).
[50] N. Oliva, J. Backman, L. Capua, M. Cavalieri, M. Luisier, and A. M. Ionescu, npj 2D Mater. Appl. 4, 1 (2020).
[51] R. L. Anderson, Solid-State Electron. 5, 341 (1962).
[52] C. G. Van de Walle and R. M. Martin, J. Vac. Sci. Technol. B 4, 1055 (1986).
[53] C. G. Van de Walle and R. M. Martin, Phys. Rev. B 35, 8154 (1987).
[54] A. Baldereschi, S. Baroni, and R. Resta, Phys. Rev. Lett. 61, 734 (1988).
[55] E. T. Yu, J. O. McCaldin, and T. C. McGill, Solid State Phys. 46, 1 (1992).
[56] F. Capasso and G. Margaritondo, editors, Heterojunction Band Discontinuities: Physics and Device Applications (North-Holland, 1987).
[57] J. Robertson, Appl. Phys. Lett. 94, 152104 (2009).
[58] J. Robertson, Phys. Stat. Soldi. A 207, 261 (2010).
[59] E. Fermi, Z. Physik 48, 73 (1928).
[60] L. H. Thomas, Proc. Cambridge Phil. Roy. Soc. 23, (1927).
[61] P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
[62] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[63] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. 45, 566 (1980).
[64] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B 46, 6671 (1992).
[65] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
[66] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B 44, 943 (1991).
[67] V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk, and G. A. Sawatzky, Phys. Rev. B 48, 16929 (1993).

[68] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys. Condens. Matter 9, 767 (1997).
[69] J. Hubbard and B. H. Flowers, Proc. Roy. Soc. London, Series A 285, 542 (1997).
[70] D. Baeriswyl, D. K. Campbell, J. M. P. Carmelo, F. Guinea, and E. Louis, The Hubbard Model: Its Physics and Its Mathematical Physics (Springer Science & Business Media, 1995).
[71] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B 73, 195107 (2006).
[72] A. D. Becke, J. Chem. Phys. 98, 1372 (1993).
[73] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. 98, 11623 (1994).
[74] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. 105, 9982 (1996).
[75] R. Martin, in Electronic Structure: Basic Theory and Practical Methods (Cambridge University Press, 2020), pp. 188–214.
[76] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. 118, 8207 (2003).
[77] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. 124, 219906 (2006).
[78] D. M. Bylander and L. Kleinman, Phys. Rev. B 41, 7868 (1990).
[79] C. B. Geller, W. Wolf, S. Picozzi, A. Continenza, R. Asahi, W. Mannstadt, A. J. Freeman, and E. Wimmer, Appl. Phys. Lett. 79, 368 (2001).
[80] R. Asahi, A. Wang, J. R. Babcock, N. L. Edleman, A. W. Metz, M. A. Lane, V. P. Dravid, C. R. Kannewurf, A. J. Freeman, and T. J. Marks, Thin Solid Films 411, 101 (2002).
[81] A. J. Freeman, J. Comput. Appl. Math. 149, 27 (2002).
[82] M. C. Gibson, S. Brand, and S. J. Clark, Phys. Rev. B 73, 125120 (2006).
[83] S. J. Clark and J. Robertson, Phys. Rev. B 82, 085208 (2010).
[84] T. Hori, Gate Dielectrics and MOS ULSIs: Principles, Technologies and Applications (Springer Science & Business Media, 2012).
[85] F. J. Himpsel, F. R. McFeely, A. Taleb-Ibrahimi, J. A. Yarmoff, and G. Hollinger, Phys. Rev. B 38, 6084 (1988).
[86] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, IEEE Trans. Electron Devices 47, 2320 (2000).
[87] H. Li, Y. Guo, J. Robertson, and Y. Okuno, J. Appl. Phys. 119, 054103 (2016).
[88] A. Ourmazd, D. W. Taylor, J. A. Rentschler, and J. Bevk, Phys. Rev. Lett. 59, 213 (1987).
[89] A. Pasquarello, M. S. Hybertsen, and R. Car, Appl. Phys. Lett. 68, 625 (1996).
[90] Y. Tu and J. Tersoff, Phys. Rev. Lett. 84, 4393 (2000).
[91] S. Tsuneyuki, M. Tsukada, H. Aoki, and Y. Matsui, Phys. Rev. Lett. 61, 869 (1988).
[92] Y. Tsuchida and T. Yagi, Nature 347, 267 (1990).
[93] S. M. Stishov and S. V. Popova, Geokhimiya 10, 837 (1961).
[94] K. T. Park, K. Terakura, and Y. Matsui, Nature 336, 670 (1988).
[95] X. Zhang and C. K. Ong, Phys. Rev. B 48, 6865 (1993).
[96] E. P. O'Reilly and J. Robertson, Phys. Rev. B 27, 3780 (1983).
[97] R. Nada, C. R. A. Catlow, R. Dovesi, and C. Pisani, Phys. Chem. Minerals 17, 353 (1990).
[98] E. Gnani, S. Reggiani, R. Colle, and M. Rudan, IEEE Trans. Electron Devices 47, 1795 (2000).
[99] Y. P. Li and W. Y. Ching, Phys. Rev. B 31, 2172 (1985).
[100] Y. Xu and W. Y. Ching, Phys. Rev. B 44, 11048 (1991).
[101] J. R. Chelikowsky, H. E. King, N. Troullier, J. Lus Martins, and J. Glinnemann, Phys. Rev. Lett. 65, 3309 (1990).

[102] N. R. Keskar and J. R. Chelikowsky, Phys. Rev. B 46, 1 (1992).
[103] D. R. Hamann, Phys. Rev. Lett. 76, 660 (1996).
[104] A. Zupan, P. Blaha, K. Schwarz, and J. P. Perdew, Phys. Rev. B 58, 11266 (1998).
[105] T. Demuth, Y. Jeanvoine, J. Hafner, and J. G. Ángyán, J. Phys. Condens. Matter 11, 3833 (1999).
[106] X. Zhu and S. G. Louie, Phys. Rev. B 43, 14142 (1991).
[107] W. Chen and A. Pasquarello, Phys. Rev. B 86, 035134 (2012).
[108] W. Chen, G. Miceli, G.-M. Rignanese, and A. Pasquarello, Phys. Rev. Mater. 2, 073803 (2018).
[109] E. K. Chang, M. Rohlfing, and S. G. Louie, Phys. Rev. Lett. 85, 2613 (2000).
[110] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. Cryst. Mater. 220, 567 (2005).
[111] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B 41, 1227 (1990).
[112] K. Kihara, Z. Kristallogr. Cryst. Mater. 148, 237 (1978).
[113] K. Kihara, Z. Kristallogr. Cryst. Mater. 152, 95 (1980).
[114] K. Kihara, Z. Kristallogr. Cryst. Mater. 157, 93 (1981).
[115] R. W. G. Wyckoff, Am. J. Sci. 9, 448 (1925).
[116] A. F. Wright and A. J. Leadbetter, Phil. Mag. 31, 1391 (1975).
[117] J. R. Smyth, J. V. Smith, Gilberto. Artioli, and Ake. Kvick, J. Phys. Chem. 91, 988 (1987).
[118] A. Kirfe, G. Will, and J. Arndt, Z. Kristallogr. Cryst. Mater. 149, 315 (1979).
[119] F. L. Galeener, Solid State Commun. 44, 1037 (1982).
[120] L. Levien, C. T. Prewitt, and D. J. Weidner, Am. Mineral. 65, 920 (1980).
[121] A. F. Wright and M. S. Lehmann, J. Solid State Chem. 36, 371 (1981).
[122] W. A. Dollase, Acta Cryst. 23, 617 (1967).
[123] P. Bettermann and F. Liebau, Contr. Mineral. and Petrol. 53, 25 (1975).
[124] T. R. Hill, H. Konishi, and H. Xu, Am. Mineral. 98, 187 (2013).
[125] N. L. Ross, J. Shu, and R. M. Hazen, Am. Mineral. 75, 739 (1990).
[126] J. R. Chelikowsky and M. Schlüter, Phys. Rev. B 15, 4020 (1977).
[127] L. A. J. Garvie, P. Rez, J. R. Alvarez, P. R. Buseck, A. J. Craven, and R. Brydson, Am. Mineral. 85, 732 (2000).
[128] P. V. Rysselberghe, J. Phys. Chem. 36, 1152 (1932).
[129] R. Warmbier, F. Mohammed, and A. Quandt, Opt. Eng. 53, 071808 (2014).
[130] F. Zheng, J. Tao, and A. M. Rappe, Phys. Rev. B 95, 035203 (2017).
[131] M. Harb, P. Labéguerie, I. Baraille, and M. Rérat, Phys. Rev. B 80, 235131 (2009).
[132] L. He, F. Liu, G. Hautier, M. J. T. Oliveira, M. A. L. Marques, F. D. Vila, J. J. Rehr, G.-M. Rignanese, and A. Zhou, Phys. Rev. B 89, 064305 (2014).
[133] L. E. Ramos, J. Furthmüller, and F. Bechstedt, Phys. Rev. B 69, 085102 (2004).
[134] Y. Kang, G. Kang, H.-H. Nahm, S.-H. Cho, Y. S. Park, and S. Han, Phys. Rev. B 89, 165130 (2014).
[135] S. Maj, Phys. Chem. Minerals. 10, 133 (1984).
[136] D. C. Look, B. Claflin, Y. I. Alivov, and S. J. Park, Phys. Stat. Sol. A 201, 2203 (2004).
[137] L. J. Brillson and Y. Lu, J. Appl. Phys. 109, 121301 (2011).
[138] S. Kattel, P. J. Ramírez, J. G. Chen, J. A. Rodriguez, and P. Liu, Science 355, 1296 (2017).
[139] R. L. Hoffman, B. J. Norris, and J. F. Wager, Appl. Phys. Lett. 82, 733 (2003).
[140] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, Nat. Mater. 3, 868 (2004).

[141] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, Nat. Mater. 9, 721 (2010).
[142] Z. Wang, M. Saito, K. P. McKenna, S. Fukami, H. Sato, S. Ikeda, H. Ohno, and Y. Ikuhara, Nano Lett. 16, 1530 (2016).
[143] R. T. Tung, Appl. Phys. Rev. 1, 011304 (2014).
[144] R. T. Tung, Phys. Rev. B 64, 205310 (2001).
[145] T. Nishimura, T. Yajima, and A. Toriumi, Appl. Phys. Express 9, 081201 (2016).
[146] J. Robertson, Y. Guo, Z. Zhang, and H. Li, J. Vac. Sci. Technol. B 38, 042208 (2020).
[147] C. Berthod, N. Binggeli, and A. Baldereschi, Phys. Rev. B 68, 085323 (2003).
[148] K.-Y. Tse, D. Liu, and J. Robertson, Phys. Rev. B 81, 035325 (2010).
[149] J. Goniakowski and C. Noguera, J. Phys. Condens. Matter 26, 485010 (2014).
[150] K. Nakamura, T. Akiyama, T. Ito, M. Weinert, and A. J. Freeman, Phys. Rev. B 81, 220409 (2010).
[151] H. X. Yang, M. Chshiev, B. Dieny, J. H. Lee, A. Manchon, and K. H. Shin, Phys. Rev. B 84, 054401 (2011).
[152] J. Goniakowski and C. Noguera, Phys. Rev. B 66, 085417 (2002).
[153] J. Goniakowski, F. Finocchi, and C. Noguera, Rep. Prog. Phys. 71, 016501 (2007).
[154] Y. Dong and L. J. Brillson, J. Electron. Mater. 37, 743 (2008).
[155] N. R. D'Amico, G. Cantele, C. A. Perroni, and D. Ninno, J. Phys. Condens. Matter 27, 015006 (2014).
[156] J. Chen, Z. Zhang, Y. Guo, and J. Robertson, Microelectron. Eng. 216, 111056 (2019).
[157] E. A. Kraut, R. W. Grant, J. R. Waldrop, and S. P. Kowalczyk, Phys. Rev. Lett. 44, 1620 (1980).
[158] Z. Zhang, Y. Guo, and J. Robertson, Appl. Phys. Lett. 114, 161601 (2019).
[159] A. Janotti, D. Segev, and C. G. Van de Walle, Phys. Rev. B 74, 045202 (2006).
[160] M. Shishkin and G. Kresse, Phys. Rev. B 75, 235102 (2007).
[161] A. Schleife, C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, Phys. Rev. B 80, 035112 (2009).
[162] F. Oba, A. Togo, I. Tanaka, J. Paier, and G. Kresse, Phys. Rev. B 77, 245202 (2008).
[163] J. Robertson, K. Xiong, and S. J. Clark, Phys. Status Solidi B 243, 2054 (2006).
[164] S. J. Clark and J. Robertson, Phys. Status Solidi B 248, 537 (2011).
[165] S. J. Clark, T. W. Hollins, K. Refson, and N. I. Gidopoulos, J. Phys. Condens. Matter 29, 374002 (2017).
[166] L. A. Agapito, S. Curtarolo, and M. Buongiorno Nardelli, Phys. Rev. X 5, 011006 (2015).
[167] F. P. Koffyberg, Phys. Rev. B 13, 4470 (1976).
[168] K. Gurumurugan, D. Mangalaraj, S. K. Narayandass, K. Sekar, and C. P. G. Vallabhan, Semicond. Sci. Technol. 9, 1827 (1994).
[169] A. Wang, J. R. Babcock, N. L. Edleman, A. W. Metz, M. A. Lane, R. Asahi, V. P. Dravid, C. R. Kannewurf, A. J. Freeman, and T. J. Marks, Proc. Natl. Acad. Sci. U.S.A. 98, 7113 (2001).
[170] U. Schönberger and F. Aryasetiawan, Phys. Rev. B 52, 8788 (1995).
[171] D. M. Roessler and W. C. Walker, Phys. Rev. 159, 733 (1967).
[172] G. Cappellini, S. Bouette-Russo, B. Amadon, C. Noguera, and F. Finocchi, J. Phys. Condens. Matter 12, 3671 (2000).
[173] A. S. Rao and R. J. Kearney, Phys. Status Solidi B 95, 243 (1979).
[174] Y. Kaneko and T. Koda, J. Cryst. Growth 86, 72 (1988).

[175] S. T. Pantelides, Phys. Rev. B 11, 5082 (1975).
[176] V. Stevanović, S. Lany, D. S. Ginley, W. Tumas, and A. Zunger, Phys. Chem. Chem. Phys. 16, 3706 (2014).
[177] H. B. Michaelson, J. Appl. Phys. 48, 4729 (1977).
[178] J. C. Phillips, Rev. Mod. Phys. 42, 317 (1970).
[179] B. F. Levine, J. Chem. Phys. 59, 1463 (1973).
[180] M. W. Allen and S. M. Durbin, Appl. Phys. Lett. 92, 122110 (2008).
[181] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, Phys. Rev. B 63, 054416 (2001).
[182] X.-G. Zhang and W. H. Butler, Phys. Rev. B 70, 172407 (2004).
[183] R. Shimabukuro, K. Nakamura, T. Akiyama, and T. Ito, Physica E 42, 1014 (2010).
[184] G. Wang, Q. Peng, and Y. Li, J. Am. Chem. Soc. 131, 14200 (2009).
[185] Y. Y. Illarionov, A. G. Banshchikov, D. K. Polyushkin, S. Wachter, T. Knobloch, M. Thesberg, M. I. Vexler, M. Waltl, M. Lanza, N. S. Sokolov, T. Mueller, and T. Grasser, 2D Mater. 6, 045004 (2019).
[186] L. Sang, M. Liao, Y. Koide, and M. Sumiya, Appl. Phys. Lett. 99, 031115 (2011).
[187] R. A. Heaton and C. C. Lin, Phys. Rev. B 22, 3629 (1980).
[188] E. L. Shirley, Phys. Rev. B 58, 9579 (1998).
[189] Y. Ma and M. Rohlfing, Phys. Rev. B 75, 205114 (2007).
[190] T. Bischoff, I. Reshetnyak, and A. Pasquarello, Phys. Rev. B 101, 235302 (2020).
[191] M. Kim, Y.-J. Zhao, A. J. Freeman, and W. Mannstadt, Appl. Phys. Lett. 84, 3579 (2004).
[192] D. O. Scanlon, B. J. Morgan, and G. W. Watson, J. Chem. Phys. 131, 124703 (2009).
[193] Y. Guo, Z. Zhang, and J. Robertson, Phys. Status Solidi RRL 15, 2100295 (2021).
[194] J. Chen, Z. Zhang, Y. Guo, and J. Robertson, J. Appl. Phys. 129, 175304 (2021).
[195] Yu. Yu. Illarionov, M. I. Vexler, V. V. Fedorov, S. M. Suturin, and N. S. Sokolov, J. Appl. Phys. 115, 223706 (2014).
[196] M. A. Olmstead, R. I. G. Uhrberg, R. D. Bringans, and R. Z. Bachrach, Phys. Rev. B 35, 7526 (1987).
[197] R. M. Tromp and M. C. Reuter, Phys. Rev. Lett. 61, 1756 (1988).
[198] J. D. Denlinger, E. Rotenberg, U. Hessinger, M. Leskovar, and M. A. Olmstead, Phys. Rev. B 51, 5352 (1995).
[199] F. J. Himpsel, U. O. Karlsson, J. F. Morar, D. Rieger, and J. A. Yarmoff, Phys. Rev. Lett. 56, 1497 (1986).
[200] D. Rieger, F. J. Himpsel, U. O. Karlsson, F. R. McFeely, J. F. Morar, and J. A. Yarmoff, Phys. Rev. B 34, 7295 (1986).
[201] A. Izumi, Y. Hirai, K. Tsutsui, and N. S. Sokolov, Appl. Phys. Lett. 67, 2792 (1995).
[202] E. Rotenberg, J. D. Denlinger, M. Leskovar, U. Hessinger, and M. A. Olmstead, Phys. Rev. B 50, 11052 (1994).
[203] E. Rotenberg, J. D. Denlinger, and M. A. Olmstead, Phys. Rev. B 53, 1584 (1996).
[204] M. Dixon and M. J. Gillan, J. Phys. C: Solid State Phys. 11, L165 (1978).
[205] G. W. Rubloff, Phys. Rev. B 5, 662 (1972).
[206] R. T. Poole, D. R. Williams, J. D. Riley, J. G. Jenkin, J. Liesegang, and R. C. G. Leckey, Chem. Phys. Lett. 36, 401 (1975).
[207] C. Deiter, M. Bierkandt, A. Klust, C. Kumpf, Y. Su, O. Bunk, R. Feidenhans'l, and J. Wollschläger, Phys. Rev. B 82, 085449 (2010).
[208] L. Pasquali, S. M. Suturin, V. P. Ulin, N. S. Sokolov, G. Selvaggi, A. Giglia, N.

Mahne, M. Pedio, and S. Nannarone, Phys. Rev. B 72, 045448 (2005).
[209] B. D. Yu, Y. Miyamoto, O. Sugino, A. Sakai, T. Sasaki, and T. Ohno, J. Vac. Sci. Technol. B 19, 1180 (2001).
[210] Z. Zhang, Y. Guo, and J. Robertson, J. Phys. Chem. C 124, 19698 (2020).
[211] P. W. Peacock and J. Robertson, Phys. Rev. Lett. 92, 057601 (2004).
[212] M. W. M. Watanabe, Y. I. Y. Iketani, and M. A. M. Asada, Jpn. J. Appl. Phys. 39, L964 (2000).
[213] S. Z. Butler, S. M. Hollen, L. Cao, Y. Cui, J. A. Gupta, H. R. Gutiérrez, T. F. Heinz, S. S. Hong, J. Huang, A. F. Ismach, E. Johnston-Halperin, M. Kuno, V. V. Plashnitsa, R. D. Robinson, R. S. Ruoff, S. Salahuddin, J. Shan, L. Shi, M. G. Spencer, M. Terrones, W. Windl, and J. E. Goldberger, ACS Nano 7, 2898 (2013).
[214] S. Das, M. Kim, J. Lee, and W. Choi, Crit. Rev. Solid State Mater. Sci. 39, 231 (2014).
[215] H. Zhang, ACS Nano 9, 9451 (2015).
[216] S. Das, J. A. Robinson, M. Dubey, H. Terrones, and M. Terrones, Annu. Rev. Mater. Res. 45, 1 (2015).
[217] S. Das, A. Prakash, R. Salazar, and J. Appenzeller, ACS Nano 8, 1681 (2014).
[218] G. Fiori, F. Bonaccorso, G. Iannaccone, T. Palacios, D. Neumaier, A. Seabaugh, S. K. Banerjee, and L. Colombo, Nat. Nanotechnol. 9, 768 (2014).
[219] Y. Wang, Y. Zheng, C. Han, and W. Chen, Nano Res. 14, 1682 (2021).
[220] S. Das, H.-Y. Chen, A. V. Penumatcha, and J. Appenzeller, Nano Lett. 13, 100 (2013).
[221] Y. Guo, D. Liu, and J. Robertson, ACS Appl. Mater. Interfaces 7, 25709 (2015).
[222] C. Kim, I. Moon, D. Lee, M. S. Choi, F. Ahmed, S. Nam, Y. Cho, H.-J. Shin, S. Park, and W. J. Yoo, ACS Nano 11, 1588 (2017).
[223] Y. Liu, J. Guo, E. Zhu, L. Liao, S.-J. Lee, M. Ding, I. Shakir, V. Gambin, Y. Huang, and X. Duan, Nature 557, 696 (2018).
[224] H. Fang, S. Chuang, T. C. Chang, K. Takei, T. Takahashi, and A. Javey, Nano Lett. 12, 3788 (2012).
[225] H. Zhou, C. Wang, J. C. Shaw, R. Cheng, Y. Chen, X. Huang, Y. Liu, N. O. Weiss, Z. Lin, Y. Huang, and X. Duan, Nano Lett. 15, 709 (2015).
[226] S. Das and J. Appenzeller, Appl. Phys. Lett. 103, 103501 (2013).
[227] J. Kang, W. Liu, D. Sarkar, D. Jena, and K. Banerjee, Phys. Rev. X 4, 031005 (2014).
[228] I. Popov, G. Seifert, and D. Tománek, Phys. Rev. Lett. 108, 156802 (2012).
[229] C. Gong, L. Colombo, R. M. Wallace, and K. Cho, Nano Lett. 14, 1714 (2014).
[230] S. McDonnell, C. Smyth, C. L. Hinkle, and R. M. Wallace, ACS Appl. Mater. Interfaces 8, 8289 (2016).
[231] C.-I. Lu, C. J. Butler, J.-K. Huang, Y.-H. Chu, H.-H. Yang, C.-M. Wei, L.-J. Li, and M.-T. Lin, npj 2D Mater. Appl. 1, 1 (2017).
[232] Y. Wang, J. C. Kim, R. J. Wu, J. Martinez, X. Song, J. Yang, F. Zhao, A. Mkhoyan, H. Y. Jeong, and M. Chhowalla, Nature 568, 70 (2019).
[233] X. Wang, S. Y. Kim, and R. M. Wallace, ACS Appl. Mater. Interfaces 13, 15802 (2021).
[234] Y. Wang, R. Xi Yang, R. Quhe, H. Zhong, L. Cong, M. Ye, Z. Ni, Z. Song, J. Yang, J. Shi, J. Li, and J. Lu, Nanoscale 8, 1179 (2016).
[235] W. Liu, J. Kang, D. Sarkar, Y. Khatami, D. Jena, and K. Banerjee, Nano Lett. 13, 1983 (2013).
[236] W. Zhang, M.-H. Chiu, C.-H. Chen, W. Chen, L.-J. Li, and A. T. S. Wee, ACS Nano 8, 8653 (2014).

[237] C. M. Smyth, L. A. Walsh, P. Bolshakov, M. Catalano, R. Addou, L. Wang, J. Kim, M. J. Kim, C. D. Young, C. L. Hinkle, and R. M. Wallace, ACS Appl. Nano Mater. 2, 75 (2019).
[238] L. Kong, X. Zhang, Q. Tao, M. Zhang, W. Dang, Z. Li, L. Feng, L. Liao, X. Duan, and Y. Liu, Nat. Commun. 11, 1866 (2020).
[239] X. Cui, E.-M. Shih, L. A. Jauregui, S. H. Chae, Y. D. Kim, B. Li, D. Seo, K. Pistunova, J. Yin, J.-H. Park, H.-J. Choi, Y. H. Lee, K. Watanabe, T. Taniguchi, P. Kim, C. R. Dean, and J. C. Hone, Nano Lett. 17, 4781 (2017).
[240] G. Pande, J.-Y. Siao, W.-L. Chen, C.-J. Lee, R. Sankar, Y.-M. Chang, C.-D. Chen, W.-H. Chang, F.-C. Chou, and M.-T. Lin, ACS Appl. Mater. Interfaces 12, 18667 (2020).
[241] R. Kumar, S. Sahoo, E. Joanni, R. K. Singh, R. M. Yadav, R. K. Verma, D. P. Singh, W. K. Tan, A. Pérez del Pino, S. A. Moshkalev, and A. Matsuda, Nano Res. 12, 2655 (2019).
[242] M. Farmanbar and G. Brocks, Phys. Rev. B 93, 085304 (2016).
[243] K. Reidy, G. Varnavides, J. D. Thomsen, A. Kumar, T. Pham, A. M. Blackburn, P. Anikeeva, P. Narang, J. M. LeBeau, and F. M. Ross, Nat. Commun. 12, 1290 (2021).
[244] X. Zhao, J. Qiao, S. M. Chan, J. Li, J. Dan, S. Ning, W. Zhou, S. Y. Quek, S. J. Pennycook, and K. P. Loh, Nano Lett. 21, 3262 (2021).
[245] A. Ebnonnasir, B. Narayanan, S. Kodambaka, and C. V. Ciobanu, Appl. Phys. Lett. 105, 031603 (2014).
[246] F. Tumino, C. S. Casari, A. Li Bassi, and S. Tosoni, J. Phys. Chem. C 124, 12424 (2020).
[247] Y. Shao, S. Song, X. Wu, J. Qi, H. Lu, C. Liu, S. Zhu, Z. Liu, J. Wang, D. Shi, S. Du, Y. Wang, and H.-J. Gao, Appl. Phys. Lett. 111, 113107 (2017).
[248] M. Piquemal-Banci, R. Galceran, F. Godel, S. Caneva, M.-B. Martin, R. S. Weatherup, P. R. Kidambi, K. Bouzehouane, S. Xavier, A. Anane, F. Petroff, A. Fert, S. M.-M. Dubois, J.-C. Charlier, J. Robertson, S. Hofmann, B. Dlubak, and P. Seneor, ACS Nano 12, 4712 (2018).
[249] J. Ontaneda, F. Viñes, F. Illas, and R. Grau-Crespo, Phys. Chem. Chem. Phys. 21, 10888 (2019).
[250] H. Lu, Y. Guo, and J. Robertson, ACS Appl. Mater. Interfaces 13, 47226 (2021).
[251] H. Lu, J. Robertson, and H. Naganuma, Appl. Phys. Rev. 8, 031307 (2021).
[252] Z. Zhang, Y. Guo, and J. Robertson, ACS Appl. Mater. Interfaces 14, 11903 (2022).
[253] Y. Guo and J. Robertson, Appl. Phys. Lett. 108, 233104 (2016).
[254] G. Kresse and J. Furthmüller, Phys. Rev. B 54, 11169 (1996).
[255] J. Klimeš, D. R. Bowler, and A. Michaelides, Phys. Rev. B 83, 195131 (2011).
[256] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. 102, 073005 (2009).