Atomic-scale study of electronic defects and ambient degradation in Van der Waals layered black phosphorus

Mark Wentink

A thesis presented for the degree of Doctor of Philosophy

Supervisor: Professor Anthony J. Kenyon
Final submission date: April 16, 2020

Department of Electronic and Electrical Engineering
University College London, UK
I, Mark Wentink, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

April 16, 2020
Abstract

Van der Waals heterostructures form a rapidly expanding field of study in physics, chemistry, material science and engineering alike. More and more exfoliable materials are discovered and the techniques to take them apart and build them back up are continuously refined. As stacking techniques become increasingly precise and reliable, potential applications of heterostructures in electronics and opto-electronics are unlocked.

Within Van der Waals materials, black phosphorus has the valuable combination of a direct, tunable bandgap and a high charge carrier mobility, making it a good candidate for heterostructure devices. On its own however, black phosphorus suffers from degradation in ambient conditions and a much lower mobility than predicted. Blue phosphorus, a high-pressure allotrope, shows similar promise with further opportunities as precursor to phosphorus nanoribbons and nanotubes, but is predicted to suffer from the same problems. For the materials to be success in nanostructure applications, both these shortcomings will need to be addressed.

In the following thesis, I set out to offer a well-rounded study of the properties and structure of black phosphorus in ambient conditions as well as a close examination of apparently inherent point defects at the atomic scale. The ambient degradation of black phosphorus was found to slow down significantly in a low-humidity environment, corroborating the theory that ambient degradation is a multi-step process involving oxygen and water. Chemical analysis of sample produced through both common synthesis pathways showed low concentrations of tin and iodine impurities. Atomic scale defects were studied both through STM and NC-AFM. The dominant defect has a characteristic double-lobed shape. DFT and Tight-binding calculations of tin substitutions are in good agreement with our STM and STS experiments, and find the double lobed shape to arise from a series of hydrogen-like states hosted in a potential well around the negatively charged tin defects. At high pressure, phase transitions between allotropes are found to be slow and exhibit a strong hysteresis, resulting in the coexistence of several phosphorus phases. A sufficient concentration of dopants may be able to stabilise blue phosphorus at ambient pressure.
Impact statement

Van der Waals heterostructures have the potential to provide a completely different approach to how we engineer compound materials. The stacking together of atomic layers of different materials is a consistent, reliable and reproducible way of tailoring compound materials to suit an extremely broad range of applications. The development of a set of materials of which the properties and the interactions are thoroughly understood, is essential for the successful engineering of these heterostructures. As a narrow gap semiconductor, black phosphorus is a valuable addition to this group of materials. The thorough study of its structural and electronic properties, as well as its common impurities as presented in this thesis, supports the successful implementation of black phosphorus in heterostructure applications.

Atomic point defects can dramatically change the electronic properties of crystalline materials. However, in contrast to conventional semiconductors such as silicon, our current knowledge of defects and impurity states in two-dimensional semiconducting materials is limited. The atomic-scale study of the fundamental physical properties of the well-known double-lobed defects seen in STM images of black phosphorus, as presented in this thesis, combined with theoretical calculations, improve the understanding of the p-type character consistently found in black phosphorus. Our insights pave the way towards the rational design of electronic devices based on black phosphorus by defect engineering.
3.5 Raman spectroscopy .................................................. 54
3.6 High pressure measurements using diamond anvil cells ................. 56
3.7 Density Functional Theory .............................................. 58
3.8 Tight-Binding calculations ............................................ 60

4 Sample origin and preparation ........................................... 62

5 Degradation and induced defects in black phosphorus .......................... 64
  5.1 Exposing atomically clean surfaces for UHV experiments .................. 64
  5.2 Chemical Analysis .................................................... 66
    5.2.1 XPS of pristine and contaminated black phosphorus .................. 66
    5.2.2 SIMS of pristine black phosphorus .................................. 69
  5.3 STM of samples as-received .......................................... 71
    5.3.1 Microscope and tip calibration on Si samples ....................... 71
    5.3.2 Lattice and bandgap measurements ................................... 72
    5.3.3 Non-double-lobed defects present in as-received samples .......... 74
  5.4 Non-contact AFM of samples as-received .................................. 75
    5.4.1 Lattice / step edges .............................................. 75
    5.4.2 Observed defects ................................................ 77
    5.4.3 Effect of frequency shift on resolution ............................. 80
    5.4.4 Atom row buckling .............................................. 81
  5.5 Degradation in ambient conditions: AFM and optical microscopy ............ 82
    5.5.1 Overnight AFM degradation ....................................... 82
    5.5.2 Low humidity AFM .............................................. 84
    5.5.3 Heating samples and washing to remove oxide ....................... 85
  5.6 Induced defects - Si and H₂O dosing in STM ................................ 87
    5.6.1 Water-related defects ......................................... 88
    5.6.2 Silicon-related defects ........................................ 90
    5.6.3 Discussion ................................................... 93
  5.7 Sputtering and annealing ............................................. 95
  5.8 Chapter Discussion ................................................ 97
  5.9 Chapter Summary .................................................. 98

6 Substitutional Sn defects in black phosphorus .................................... 101
  6.1 Double-lobes .................................................................. 101
    6.1.1 left- and right-handed ......................................... 103
    6.1.2 Depth of defects .............................................. 104
Acknowledgements

To me, this thesis symbolises four incredible years of academic and personal growth. To say it has been an easy ride would be a lie. I have felt at times lonely, lost, and inadequate. If I made it out the other side more open, more independent, more grounded in myself, it is thanks to the incredible support, help, and friendship of so many people.

I would like to thank firstly my supervisor, Tony, for his direction and mentorship, his honesty and his compassion. I always felt better leaving his office.

To my examiners, Matthew Halsall and Chris Howard, thank you for taking the time to read all this, and for being part of the process.

Thank you to Steven Schofield and Manuel Siegl, for introducing me to the world of STM, simultaneously an incredibly satisfying and frustrating experience.

Thank you to Julian and Alex Shluger, to Martik and Johannes, to John Proctor in Salford, and to Joseph and Alex Schwarz in Hamburg for showing me how it all fits together.

To all my friends in the CDT, 1111, and the LCN. There is nothing more heartwarming than to be able to commiserate, and then laugh together. Without you, I would have lost my sanity long ago.

I would like to thank the teaching team at the faculty, and all the undergraduate students I have met and taught, without whom I may have finished this thesis 6 months earlier, but would have missed out on a great passion in my life.

To my parents, for keeping an eye out while setting me free upon the world, and to my sister, for clearing the way and showing me that it can be done.

Finally, to GV, for being by my side for most of these four years. Your love and support have been invaluable. Out of all the lessons these years held, yours have changed me the most.
1 Introduction

1.1 Van der Waals layered materials

In the pursuit of new discoveries and understanding, the science community has relentlessly explored the limits of our physical world. We explored the extremely fast, found our interpretation of physics to break down, and developed special and general relativity. We explored the extremely small, and similarly found a need to refine our understanding to include quantum mechanics. These entirely new fields of physics became accessible to us thanks to the development of experimental techniques with ever-increasing accuracy and resolution. Every so often, an apparently small or incremental improvement in techniques opens up huge opportunities to explore new science.

When in 2004, Andre Geim and Konstantin Novoselov developed a simple technique to pull apart graphite into individual atomically thin sheet of carbon[1], they gave us access to an area of material science that has since exploded in research interest and has brought together physicists, chemists and engineers[2, 3]. In the years following their discovery, Geim and Novoselov's technique was used on a family of materials called Van der Waals layered materials. What these materials have in common is that the atoms they consist of have strong covalent bonds in single atom thickness sheets, but have weak Van der Waals bonds between the sheets[4, 5]. Like pages in a book, gentle exfoliation techniques can pull individual sheets apart without breaking in-plane bonds. The resulting sheets have macroscopic lateral dimensions, but thicknesses on the nanometre or even angstrom scale. As such, they are called two-dimensional.

2D materials continue to amaze us by exhibiting exotic physics and are proving to be fertile ground for new device applications[6, 7]. Together, these materials span a huge range of electronic, optical and structural properties and many new applications rely on stacking together single layers of different materials to create 2D heterostructures[8, 9]. Thanks to the weak interlayer bonding in these materials, common problems such as chemical interaction and lattice mismatch are much less prevalent and creating atomically smooth interfaces is easier[2].

Beyond the appeal of combining as many materials as desired to fine-tune their properties, 2D layered materials are excellent platforms for the study of new physics thanks to the dimensional confinement of charge within the sheets, the clean interfaces between different electronic materials, and proximity effects layers have on each other. Many layered materials exhibit phenomena such as superconductivity and charge density waves thanks to their two-dimensional nature[10, 11]. Put together, a large range
of interplays, such as charge redistribution and structural changes, becomes available. These interactions can further be influenced by choice of material and crystal orientation alignment[12]. Charge and spin density waves have been observed in sheets with transition metals, as well as induced into neighbouring sheets through proximity effects. In graphene, interactions with neighbouring layers has induced secondary Dirac points, the opening of a bandgap, and spin orbit coupling while the insulating properties of hexagonal boron nitride dramatically improves the performance of graphene based devices[13, 14]. Thin film perovskite based solar cells already show power conversion efficiencies of 20%[15] and transition metal dichalcogenides lend themselves well to the engineering of quantum dots[16].

2D material applications are often sensitive to defects. As they specifically rely on the dimensional confinement of charge and on perfect atomic interfaces, even low concentrations of charge traps or scattering centres are detrimental to their properties. In graphene, atomic vacancies provide nucleation sites for dislocations[17] and damage from electron beams[18] during measurements or device fabrication, and the resulting defect structures favour the zigzag edge, strongly affecting the local electronic structure[18]. In h-BN, single atom defect sites near edges allow interlayer bonding[19], allowing h-BN metallic edges to recover the bulk insulating properties[20].

On the other hand, the intentional doping of the materials can induce significant changes in electron and hole concentrations causing the material to change from n-type to p-type or vice versa, or can create nucleation sites for the onset of charge and spin density waves.

There are several examples of defect engineering in Van der Waals materials. The use of PDMS stamps[21] to pick monolayers up and stack them together introduces impurities at the interfaces. Annealing the structures afterwards has been shown to remove those contaminants and achieve higher interface quality. On the flipside, the introduction of p-type dopants in MoS$_2$ greatly improved its photoluminescence[22], while doped graphene is expected to exhibit chiral superconductivity[23]. Be it for improving the purity of materials or for harnessing defects to alter their properties, a thorough understanding of the present defects and their properties is essential.

Black phosphorus is a recent addition to the field that has displayed interesting properties by itself in bulk and monolayer form, as well as in heterostructures. Its direct bandgap varies with thickness within the near-IR and visible frequency range[24, 25] and many of its properties show an anisotropy arising from its buckled crystal structure[26]. In heterostructures, black phosphorus has demonstrated high field effect mobility and on/off ratio[27], as well as femtosecond optical switching[28].
A closely related allotrope, dubbed blue phosphorus based on its photo-absorption range is predicted to be equally stable as black phosphorus[13], but has yet to be produced in ambient conditions. The allotrope is accessible at high pressure[29, 30] and has been produced by chemical vapour deposition on gold[31], similarly to pseudo-2D materials such as silicene and germanene.

1.2 Motivation of study

BP has shown a lot of promise for 2D applications. In particular its adjustable direct bandgap, which varies inversely with the number of layers, ranging from 1.6 eV in a monolayer to 0.3 eV in bulk[24, 25], and its very high predicted charge carrier mobility make BP an excellent material for opto-electronic applications. However, its ambient degradation makes it a difficult to work with. The precise degradation mechanism[32, 21], and consequently possible passivation or encapsulation techniques[33] have only been actively studied recently. While encapsulation in thick alumina layers has already been shown to stabilise the material, a better understanding of the degradation process will support less invasive passivations such as sandwiching black phosphorus between other layered materials[34, 35].

Beyond the process of its degradation, black phosphorus presents a second question. Although black phosphorus devices have already demonstrated carrier mobilities on the order of 1000 cm²V⁻¹s⁻¹, theory studies predict black phosphorus has the potential of an order of magnitude increase in this value[36, 37], suggesting current devices hold a large number of scattering centres in the black phosphorus lattice. Since it was first made in 1914[38, 39], black phosphorus has always been measured to be p-doped, despite continuous attempts to make perfectly pure samples. While intentional doping can be used to make BP n-type[40], its default state appears to be p-doped. This p-type nature seems to go paired with atomic scale double lobed defects that have been ubiquitous in STM studies of the black phosphorus surface[41, 42, 43]. The origin of these defects and the remarkable effect they have on the anisotropic black phosphorus structure and its electronic properties is an area of active study.

1.3 Outline

The second chapter of this thesis offers an overview of the relevant literature to date. I first introduce Van der Waals materials as a family, and specifically black and blue phosphorus, their properties, and the phase transition between them. I then go into
more detail on the synthesis processes of black phosphorus and what is known so far on its purity and its degradation in ambient conditions.

In chapters 3 and 4, I introduce the characterisation techniques that were involved in this project along with details of sample origin and preparation.

Chapter 5 describes the black phosphorus samples as received, as well as their ambient degradation, studied through chemical characterisation, scanning probe microscopy, and electron and optical microscopy. I further study defects induced through dosing with water and Si.

Chapter 6 specifically discusses the double-lobed defects seen in STM studies of black phosphorus. I compare STM/STS experiments to DFT and tight-binding calculations of their structural and electronic properties, as well as the charge state switching that can be induced in them through a locally applied field.

Chapter 7 addresses discrepancies in the literature regarding the phase transition between black and blue phosphorus, and discusses the role of impurity defects in the stability of blue phosphorus and phase coexistence of the two.

Chapter 8 summarises the findings of my project and offers an outlook and suggestions for further work.
2 Literature

In this chapter, I give an overview of previous research relevant to my project. I start with introducing the field of Van der Waals layered materials, highlighting its key members and the progress studies have made towards heterostructures. Next, I briefly discuss phosphorus allotropes before looking in more detail at black and blue phosphorus, the two layered versions of the element. Beyond their structural and electronic properties, I specifically address studies of the BP surface and the defects that have been identified through Scanning Probe Microscopy. Next, I give an overview of the synthesis processes for BP, as well as of various studies evaluating the purity of the resulting samples. The discussion of impurities continues into a section on the ambient degradation of BP. Finally, I discuss applications of BP both by itself and as part of heterostructures, and techniques that have so far made progress towards protecting devices from oxidation and degradation.

2.1 Introduction to Van der Waals layered materials

The field of layered materials has gone through a rapid growth since the isolation of graphene by Geim and Novoselov in 2004[1]. With great advances in large scale production of graphene sheets[44], alternative exfoliation and synthesis methods[45], passivation techniques, and methods of restacking sheets into heterostructures, many applications are being realised and an extensive field of interface physics is being opened up.

As new layered materials are discovered and exfoliated, the family of 2D materials is becoming increasing diverse. In 2014, an attempt by Huang et al[46] to classify the family yielded the table shown in Fig.2.1. Layered materials were split up into graphene related materials, chalcogenides, and oxides. Materials shown in blue are stable in air, compounds in green are predicted to be stable in air, while materials in red are unstable in ambient conditions. The absence of black phosphorus from this classification goes to show how recent the interest in its structure and properties is.

Black phosphorus and graphite have so far been the only mono-elemental species found to adopt a layered structure and of which stable monolayers were isolated. There are several more pseudo-2D elements, including silicene and germanene, which have been produced atomically thin, but only on a suitable substrate such as Ag(111) to stabilise the layer[47, 48]. 4-Valent graphite consists of layers of flat sp² hybridised hexagonal rings of carbon atoms with three bonds each, all 1.4 Å long. The fourth elec-
Figure 2.1: Classification of the major families of 2D materials. Compounds in blue or green are known to be stable or can be made stable. Compounds in red are unstable while materials in grey require further investigation. Taken from Huang, 2014[46]

electron sits in a p-orbital sticking in the out of plane direction normal to the surface. Layers are stacked in AB or ABC stacking order with an interlayer gap of 3.3 Å. Graphene has a charge carrier mobility 10-100 times greater than silicon[49]. It has excellent mechanical strength and flexibility and shows little change in electrical resistance with strain[50], which are essential properties for flexible electronic devices. While its lack of a bandgap limits its application in opto-electronics, its exceptional conductivity makes it an appealing backgate material for devices. It is extensively studied as anode material in fast charging batteries[51] as well as in flexible and printable devices[52]. In heterostructures with other 2D materials, in particular h-BN, it has shown success in FETs[53] and infrared optoelectronics[54].

Various derivatives of graphite, functionalised with oxygen, carbonyl or hydroxyl groups, generally termed graphite oxides, have found applications in water purification[55, 56], protective coatings[57], and as precursor material to large scale graphene sheet production[58].

Hexagonal BN has the same number of free electrons as graphene, and adopts the same hexagonal planar structure, with alternating boron and nitrogen atoms with bond lengths of 1.44 Å[59]. It is highly resistant to mechanical and chemical changes and has a large bandgap in the UV range. Its interlayer gap and bond lengths are practically identical to graphene, making it a perfect pair to stack together[60]. Graphene and h-BN, excellent conductor and insulator respectively, form the cornerstones of many 2D heterostructures[13, 14].

Transition metal dichalcogenides(TMDs), with formula MX_2, where M is a transi-
tion metal and X a chalcogen (S, Se, Te), all have hexagonal structures with each monolayer comprising three subplanes of X-M-X atoms. They span a broad range of electronic properties, with Ti, Hf, Zr, Mo and W dichalcogenides showing insulating or semiconducting properties\[61, 62\], while V, Nb and Ti dichalcogenides are metallic or semi-metallic\[63, 64, 65\]. Bandgaps, when present, are typically indirect other than at monolayer thickness, at which point a direct bandgap forms at the K point\[66\]. The subplanes can be arranged in a trigonal prismatic (MoS$_2$, WSe$_2$) or an octahedral structure (TiS$_2$, VSe$_2$)\[67\], the first resulting in reduced inversion symmetry, which significantly changes the electronic structure and gives rise to piezoelectricity. TMDs commonly have complex phase diagrams including superconductive phases and charge density waves\[10, 11\].

The last category, 2D oxides, contains a broad range of materials, typically less stable, that have been predicted or shown to be exfoliable. Transition metal oxides such as V$_2$O$_5$ and MoO$_3$\[12\] typically contain interlayer interactions beyond Van der Waals, either in the form of weak covalent bonds, or sparse intercalants. They also tend to have poor crystallinity, making them less suitable to mechanical exfoliation. More well-studied examples of 2D oxides are hybrid perovskites, which adopt the formula ABX$_3$ or AB$_2$O$_7$, where A and B are cations and X is an anion. These materials have excellent optical absorption thanks to the confinement of excitons within the layers and have found success in the production of thin film solar cells which can be dye-printed to reduce production costs.

2.2 Black and blue phosphorus

Phosphorus has an extremely rich chemistry albeit with a rather bad reputation. Even in elemental form, phosphorus has many allotropes, summarised in Fig.2.2. The poor reputation primarily stems from white phosphorus, P$_4$, which is extracted from phosphate rock. Prolonged exposure to white phosphorus fumes causes liver damage as well as the infamous phossy jaw, which is the necrosis of tissue around phosphorus deposits in the jaw, common in miners in the second half of the 19th century\[68\]. White phosphorus is also highly flammable and pyrophoric, auto-igniting in ambient conditions\[69\].

When heated in the absence of air, the P$_4$ tetrahedrons, originally in a body-centered cubic configuration, adopt an amorphous network forming red phosphorus, which is an effective flame retardant in thermoplastics\[70\]. Prolonged heating of red phosphorus results in a polymeric chain dubbed the violet phase which is unreactive under normal conditions\[71\]. The presence of water during the heating process produces highly toxic
phosphine gas, $\text{PH}_3$. Under high pressure and high temperature, white phosphorus transforms to black, which is the lowest energy configuration of the element. Recently, a second pathway to black phosphorus was discovered by Kopf et al.[72], through vapour transport of red phosphorus in the presence of SnI$_4$. Under further pressure, the orthorhombic black phosphorus will adopt a rhombohedral structure, blue phosphorus. As of yet, the blue phase is only stable at high pressure.

Black phosphorus is thermodynamically the most stable form of phosphorus[73]. It typically grows in clusters of needles up to a centimeter long but less than a milimeter wide. Recent improvements in growth methods have started to yield samples with larger crystals of several square millimeter. It has an orthorhombic structure with space group Cmca ($D_{18h}^1$)[74]. The 5-valent phosphorus atoms are arranged tetrahedrally, with three single bonds to neighbouring atoms and the remaining two electrons forming a lone pair in the out-of-plane direction. The tetrahedral structure results in a buckling of the 2D sheets, creating the ridged structure shown in Fig. 2.3a. There are two crystal directions which are similar to the ones seen in graphene and labelled accordingly as zigzag and armchair. Lattice constants are 4.37 Å in the armchair direction, 3.31 Å in the zigzag direction, and 10.47 Å in the out of plane direction[21] with an interlayer spacing of 5.5 Å, allowing the crystal to be exfoliated with methods developed for other 2D materials such as graphene and TMDs. The out of plane lattice constant is double...
Figure 2.3: a. Lattice structure of black phosphorus, b. Calculated band structure for 1, 2 and three layers highlighting the variable bandgap, c. Bandgap measurements and calculations from literature. Figure taken from Castellanos-Gomez, 2014 [21]

the Van der Waals spacing to account for an AB stacking order of the sheets. DFT calculations found a significant charge redistribution between the in-layer and interlayer regions, suggesting interlayer interactions are more involved than simple Van der Waals forces[75]. The AB stacking order, and consequently the interlayer interactions, are preserved by the buckled structure[76], which ensures phosphorus layers always align at the same angle, as opposed to graphene, which readily forms turbostratic graphite[77].

Due to its layered structure, black phosphorus can sustain significant compression out of plane. While the armchair axis is also readily compressed, the zigzag axis is not[78].

Bulk black phosphorus is a direct bandgap semiconductor with a bandgap of 0.3 eV[24]. The bandgap increases with decreasing number of layers to reach a maximum of 1.5-2.1 eV for a monolayer[25]. Fig. 2.3c shows several theoretical and experimental studies of the bandgap for different number of layers. As the figure demonstrates, there is a large discrepancy between studies regarding the bandgap. In particular DFT calculated values seem to heavily underestimate the bandgap. Nevertheless, even experimental measurements don’t agree on the bandgap value of a monolayer. Where photoluminescence measurements have quoted values in the 1.3 to 1.7 eV range[79], a scanning tunnelling spectroscopy experiment identifies the monolayer bandgap as 2.05 eV[80]. This STS study however is an outlier even among other scanning probe studies, none of which have identified a monolayer bandgap on bulk samples.

Interestingly, since the first studies of black phosphorus, the material has always been observed to be p-doped[41, 42], despite extensive attempts to optimise the purity of samples. The Fermi level sits about 0.02 eV above the valence band edge, suggesting a high number of acceptor dopants in the crystal[43, 81, 82].

Due to the buckled and layered structure of black phosphorus, there is a large
anisotropy in electron and hole effective mass and mobility along the zigzag (along the ridges) and armchair (perpendicular to the ridges) directions, in particular observed in field-effect mobility and angle-resolved optical extinction spectra[79]. Transport measurements have found two orders of magnitude difference in the charge carrier mobility, favouring the zigzag direction[83]. Optical response to polarised light varies approximately by a factor of two, favouring the armchair direction[26].

It has been suggested this anisotropy can be harnessed in a design where transport goes along the AC direction with highest mobility, and channel width along the ZZ direction with highest density of states[84]. Applying strain was found to allow for the control of the anisotropy. With 5% biaxial strain, the preferred direction of the electron effective mass can be switched through 90°[85]. The hole effective mass does not change direction with strain. The anisotropy extends into Raman spectra, where the relative intensity of Raman peaks depends on the alignment of linearly polarised light and the crystal orientation. This offers a fast way to determine the lattice directions of thin samples[86, 87].

Several theoretical studies have suggested the high pressure phase of phosphorus, rhombohedral blue, may be exfoliated similarly to black phosphorus. It would have comparably tuneable properties and would be energetically stable in the form of nanoribbons and nanotubes[88, 89, 90]. Blue phosphorus was originally discovered by Jamieson[78] in 1963 through the high pressure transition from black phosphorus at 8.3 GPa. At the time, its layered structure was unknown. More recently, blue phosphorus has experienced a comeback when it was rediscovered as a 2D material and a potential synthesis pathway to phosphorus nanotubes[91]. DFT calculations have been used to simulate the black to blue transition[92] and have shown blue phosphorus to be almost as stable as its black counterpart.

Blue phosphorus exhibits a hexagonal structure similar to graphite, but maintaining the buckled character of black phosphorus[13]. It is calculated to have a 2 eV indirect bandgap which is inversely proportional to the number of layers of a sample[91]. A semiconductor to metal transition is predicted upon the application of an electric field[93]. Blue monolayers are predicted to make excellent anode materials for lithium ion batteries thanks to the low Li diffusion barrier on their surface, and their high charge capacity[94].

While this and other theoretical studies[24][90][13] show great promise for blue phosphorus, it has yet to be produced at ambient conditions. Recently, a blue phosphorus monolayer was successfully grown on a Au substrate[31] in vacuum, although it is not clear whether the layer can be lifted off, or electronically isolated from the sub-
strate, a similar issue as that experienced by silicene and germanene. The bandgap of a blue monolayer was measured as 1.1 eV\cite{31}, significantly smaller than predicted values. The authors suggested this could result from the strain induced by the lattice mismatch between Au and blue P.

Various nanostructures of black and blue phosphorus have been demonstrated or calculated to be stable. Blue phosphorus nanoribbons, if they could be isolated at ambient conditions, would have bandgaps that vary inversely with ribbon width\cite{90}. 8 Å wide ribbons would have a 2.3 eV gap in the armchair direction and a 2.7 eV gap in the zigzag direction\cite{90}, with wider ribbons showing decreasing bandgaps tending towards the monolayer limit. These values are very similar to what a nanotube is predicted to have, 2.67 eV\cite{95}. While many of the calculations and applications of blue phosphorus remain theoretical, the recent success in producing monolayers\cite{31} might be a pathway to a CVD production mechanism at atmosphere.

Recently, black phosphorus nanoribbons were produced by the research group of Chris Howard at UCL\cite{96}. They employed a top-down process from bulk black phosphorus crystals to create liquid dispersions of BP nanoribbons 4-50 nm in width yet up to 75 micron in length.

The phase transition mechanism between black and blue phosphorus has recently been studied to explore potential pathways of producing blue phosphorus that is stable at ambient conditions. The transition relies primarily on out of plane compression flattening the buckled structure of black phosphorus, and reducing the interlayer gap until P-P bonds are formed across the gap. Chains of 3-5 bonds are made in the out of plane direction, causing areas of both blue and black character to coexist. The lone pairs are reoriented through a rotation of 60°, breaking black phosphorus armchair bonds\cite{92}. The result is a new layered crystal, rotated by 60° compared to the original and with zigzag character on both axes. Further details on the transition mechanism between black and blue phosphorus are given in chapter 7.

2.3 STM studies of black phosphorus

The effective use of BP in devices and heterostructures requires a good understanding of the density of states at its surface, as this will determine its interactions with other layers in the heterostructure. A scanning probe approach also offers insights in the defect and impurity structures in the material, offering atomic resolution topographic and electronic information.

The BP surface was first imaged in STM by Zhang et al in 2009\cite{41}. They measured
lattice constants in agreement with other characterisation techniques and calculations, 3.33 Å in the zigzag direction and 4.33 Å in the armchair direction[41]. They further identified the bulk bandgap as 0.4 eV, slightly bigger than previously reported values. More recently, Liang et al[80] suggested the top layer is electronically completely decoupled from the bulk, and as such behaves as a monolayer. Their scanning tunnelling spectroscopy measurements quote a bandgap value of 2.05 eV, consistent with a monolayer. They suggested the perceived density of states is exclusively that of the top layer. Normally however, we would expect the density of states of the top layer to be a convolution of contributions from several layers. Liang et al suggested the distance between layers is so large that there is no electronic interaction between the top layer and the bulk, hence the perceived density of states from STM measurements is that of an independent monolayer. There are as of yet no further studies discussing this electronic isolation of the top layer. In fact, several other studies have discussed the interlayer interactions in black phosphorus[42, 43], and STM studies since then, including our own, measure a bulk bandgap on bulk black phosphorus samples.

Surface atoms are arranged nearly identically to the bulk, with a slight relaxation in the AC direction. The relaxation results in a slight buckling of 0.02 Å between atoms on each side of the zigzag ridges[41, 42]. Under the right conditions, this can be observed in STM and AFM topography scans as brighter and dimmer spots.

Studies of the black phosphorus surface have focused on specific defects that seem ubiquitous to samples irrespective of source[41, 42, 43]. The observed defects are characterised by bright areas of higher current, which extend up to 10 nm and show anisotropic contrast with a characteristic double-lobed shape. Scanning tunnelling spectroscopy revealed that these defects induce a bound state in the band gap near the valence band maximum[42]. Depending on the apparent brightness of the feature, the defects were attributed to reside in up to four layers deep into the crystal. The defect density was found to be constant with depth and it was deduced that the defects are evenly distributed throughout the crystal, suggesting that they may be ingrown or intrinsic to the crystal[42].

Several previous studies of these defects have attributed the double-lobed defects to single phosphorus vacancies [42, 43, 97]. They perform tight-binding calculations of single vacancies that produce anisotropic defect states which extent across phosphorus atom rows. These studies found the lowest energy state of a mono-vacancy to be negatively charged, and could hence account for the p-doped nature of the material[42, 36].

It should be noted that these tight-binding calculations were performed on an ideal
BP lattice, with simply one atom removed. This is not a good representation of the defect structure, as the lattice would be expected to relax around the defect, forming new bonds to replace the broken ones and adopting new structure that may or may not leave dangling bonds that can host a charged defect. Recently, DFT calculations were performed to map out in detail the possible reconfigurations around a mono-vacancy[82]. It was found that the lowest energy configuration is neutral in monolayer BP, and state switching to a negatively charged state is only possible when the Fermi Level is significantly raised. In bulk BP however, the negatively charged configuration was lowest in energy.

When creating the vacancy, three P-P bonds are broken, at a cost of 6.51 eV. In the neutral configuration, the neighbouring atoms relax into the space created and compensate two of the missing bonds, leaving one P atom undercoordinated (2- rather than 3-valent). The resulting dangling bond creates an unoccupied state in the monolayer BP bandgap approximately 0.4 eV above the valence band edge.

The negatively charged defect adopts a very different structure. One P atom becomes sp$^3$d$^2$ hybridised, bonding to four neighbours at roughly 90° angles, straddling the upper and lower zigzag rows. In this configuration, no defect state is found in the bandgap. Instead, there is an anti-bonding state in the conduction band, and a non-bonding state near the valence band maximum. The bonding state sits deep in the valence band.

The most stable di-vacancy in BP was calculated to have a formation energy of 8.02 eV, which is significantly lower than the sum of two mono-vacancies, 10.02 eV [82]. In the divacancy, neighbouring atoms relax so that each atom again ends up with three covalent bonds as in the pristine surface, creating a neutral defect. As the divacancy is lower in energy than two separate mono-vacancies, it is expected that two single vacancy will coalesce when encountering each other. Indeed, calculations of the diffusion barrier of mono-vacancies found them to be highly mobile at room temperature as well as at liquid nitrogen temperatures. Hopping rates on the order of $10^3$ s$^{-1}$ were found. At the observed defect concentrations of 80 ppm, this would mean two vacancy would find each other and coalesce within 2 seconds at room temperature[81, 98, 82, 95].

A possible alternative assignment for the double-lobed defects was put forward in reference [43] where it was suggested that these impurities might be induced by negatively charged substitutional Sn atoms. The paper does not go into detail of what a substitutional Sn defect would look like and did not perform any calculations on Sn-specific defects. Instead, they consider a tight-binding model of an undefined negatively
charged point defect and are able to reproduce the observed features by considering the Coulomb interaction of charge carriers with a charged acceptor state. The local charge creates a potential well around it that is host to a hydrogen-like series of states, distorted due to the anisotropic carrier effective mass in the lattice directions. The anisotropy further breaks the degeneracy of states with same quantum number (in particular $p_x$, $p_y$, $p_z$). The exact energy ordering of the states depends strongly on screening. In particular, it has been predicted that the 2s state lies higher in energy than the 2p states in this system. The hydrogen-like states are discussed in more detail below.

While the presence of Sn impurities is expected for vapour-grown BP, where Sn is used as a catalyst, doubled-lobed defects have also been observed in high-pressure grown BP which does not utilise Sn directly in its synthesis[40], suggesting the impurities may be introduced at a precursor stage before BP synthesis.

2.4 Applications

The excellent charge transport properties of black phosphorus, along with its tunable bandgap have spurred several studies of potential applications. While FETs have received by far the most attention, gas sensors[99], battery anodes[100], saturable absorbers[28] and memory devices[101] have been demonstrated as well. FETs have been made with hole mobility of up to 1000 cm$^2$V$^{-1}$s$^{-1}$ and on/off ratios ranging $10^2$-$10^5$. The highest values are quoted for unencapsulated devices, typically with tens of nanometres thickness, but these decay within hours in ambient conditions. Capping layers such as alumina can prolong the lifetime of FETs but typically have adverse effects on device performance[33].

As a gas sensor, BP is significantly more effective at detecting ppb concentrations than its main competitor in 2D materials, graphene. 12 nm flakes are encapsulated in ionophore solutions, which allow molecules to selectively permeate. BP device performance dropped by less than 10% after a week. In particular in the response rate and the detection of Pb$^{2+}$ did phosphorus outperform graphene, with a detection limit of 1 ppb, versus 5 ppb for graphene[99].

A phosphorene-graphene hybrid material has been demonstrated as a high capacity anode for sodium ion batteries[100]. The material consists of phosphorene sandwiched between graphene layers and is readily intercalated with sodium ions, which alloy with the phosphorene to form Na$_3$P. The graphene serves as flexible backbone to accommodate the phosphorene’s volumetric expansion upon alloying as well as providing good conductivity throughout the material. The anode achieved a capacity of 2400 mAhg$^{-1}$
with a density of $0.05 \text{ Ag}^{-1}$ and a 83% retention after 100 cycles\[100].

To highlight its potential for optoelectronics, black phosphorus has been demonstrated as a saturable absorber to generate ultra short laser pulses. A 300 nm thick flake of phosphorus was sandwiched in a fibre laser connector and used to generate a 272 fs pulse centered at 1560.5 nm\[28]. The equivalent electron and hole effective mass, along with high mobility and on/off ratio make BP well suited for ambipolar charge trap devices with multibit storage\[101]. Many of these applications show greatly improved characteristics as thinner and thinner phosphorus flakes are used. Thin BP can make a key difference in near and mid-infrared optoelectronics thanks to its direct bandgap and improved specific surface area compared to bulk, yet especially devices based on few-layer BP suffer from the degradation the material undergoes in ambient conditions. Most devices break down on the scale of hours or days as the material oxidises and thinner flakes dissolve.

### 2.5 Synthesis

When initially discovered by Bridgman in 1914\[38], black phosphorus was synthesised from white phosphorus under medium pressures and temperatures using kerosene as hydrostatic medium. The bulky Bridgman pressure anvil was able to apply a pressure of 1.2 GPa and temperature of 200 °C for 30 min to convert white to black phosphorus\[39], but resulted in polycrystalline samples and had very limited access for in situ measurements to refine the process. Nevertheless, the same method was further developed by Jacobs\[102] and Keyes\[103] to obtain grain sizes of 0.1 mm. Later studies by Brown and Rundqvist\[74] identified the puckered orthorhombic structure of black phosphorus sample produced through high pressure. They still found the material grew in needles with poor crystallinity.

A significant improvement in quality and crystal size was achieved when red, rather than white phosphorus was used as precursor\[104]. In 1982, Shirotani produced millimeter sized crystals by heating red phosphorus to 270 °C under 3.8 GPa pressure and slowly decreasing temperature\[105]. The high pressure, high temperature method is still used in commercial black phosphorus samples available from 2DSemiconductors, USA\[106].

Recently, it was shown that the high pressure synthesis route of BP readily lends itself to the incorporation of substitutional dopants, as demonstrated with S, Se, and Ge\[107]. As much as 5 at% dopants were incorporated in the BP lattice simply by mixing the dopant with the red phosphorus precursor before applying heat and pressure.
In 2007, Lange developed a vapour transport method to produce black phosphorus at ambient pressure\cite{108}. The method involved heating red phosphorus along with Au, Sn, and SnI$_4$ mineralisers in a silica ampoule at 600 °C. The method was refined by Kopf\cite{72,109} in 2014 to only rely on Sn/SnI$_4$ catalysts.

The precise growth mechanism is still disputed. It has been suggested that BP crystals precipitate from a molten alloy of red phosphorus and Sn\cite{110}. Another study suggested violet phosphorus acts as intermediate state with phase transitions driven by P-Sn-I compounds\cite{111,112}. Commercial black phosphorus produced through the vapour growth method is available from HQGraphene\cite{113}, and SmartElements\cite{114}.

2.6 Purity

Both synthesis pathways produce samples with highly crystalline orthorhombic structures\cite{21,115} but the purity of the produced phosphorus is not always consistent\cite{40}. The mechanism of the vapour growth method has been extensively studied with the aim of increasing crystal size, but its resulting purity has only gained attention recently when Mayorga-Martinez\cite{40} found samples to contain Sn and I contaminants, used as catalysts in the synthesis. The study by Mayorga-Martinez prepared their own black phosphorus through both methods and evaluated the chemical composition. Metal impurities (Mg, Cr, Fe, Ni, Cu, Zn) were all found present in the hundreds of ppm in samples from both synthesis methods. The red phosphorus used as precursor was found to contain these impurities at the tens of ppm level. The VG samples further contained Sn and I at concentrations of 2000 ppm. The HP samples were not found to contain Sn or I.

Metal impurities have been found to be quite common in layered materials. They have been observed in graphite\cite{116,117}, where they dramatically affect its catalytic\cite{117} and electronic\cite{118} qualities. Similarly, MoS$_2$ has shown improved catalytic properties after purification processes\cite{119} but also through intentional doping with Ni, Fe, or Co\cite{120}. While the high pressure method does not specifically require catalysts to produce black phosphorus, it has been shown that the method readily lends itself to the incorporation of substitutional dopants. In the case of graphene and carbon nanostructures, metal impurities arise from the catalysts used to dissolve and dissociate carbon containing gas (e.g. CH$_4$) into the desired structure, as well as from the mining process of graphite ores\cite{116}. Transition metal impurities are rarely completely removed in ore purification processes, and persist in the precursors to CVD and chemical vapour transport growth of Van der Waals layered materials.
Both synthesis methods use red phosphorus as a precursor, which itself is a highly reactive material. It can be produced by heating white phosphorus to 300 °C in the absence of air. Red P exists in 5 types\cite{121}, each varying in crystallinity and extent of polymerisation. The fifth type is now commonly known as violet, or Hittorf phosphorus. Red phosphorus is commonly contaminated with small amounts of yellow phosphorus (oxidised white) and is found to take on a darker colour under prolonged storage as it slowly produces phosphoric acid in the presence of oxygen and water. Taking one step further back, the white phosphorus used to produce red is sourced from fluoroapatite (‘phosphate rock’), $3\text{Ca}_3(\text{PO}_4)_2\cdot\text{CaF}_2$, which commonly contains Ca, Mg, Fe, Al, and Si impurities. Finally, KI is used as catalyst in the conversion from white to red.

Beyond accidental impurity, black phosphorus has been successfully doped to up to 5 at% during the high pressure synthesis with Se, S, Ge and Te\cite{107}. Surface doping was also achieved by depositing K in vacuum. The K ions donate their charge, creating a 2D electron gas at the BP surface. Doping with K, Se, and Te turned the material from its usual p-type to slightly n-type, and resulting in a small increase in the bandgap, ranging from 0.3 eV to 0.6 eV\cite{50}.

It is not clear where the natural p-doping of black phosphorus comes from. While it was suggested that intrinsic vacancies produced acceptor states that could account for the shift of the Fermi level, there have since been numerous theoretical studies showing that monovacancies are highly mobile on the black phosphorus surface and are unstable as compared to di-vacancies, preferring to pair up upon encountering each other. Impurities resulting from sample degradation are consistently observed as increased oxygen and water content. The oxidation mechanism has been actively researched in both theoretical\cite{122, 123, 32} and experimental studies\cite{97, 124, 115} and is discussed in the next section.

2.7 Degradation

Since 2014, there have been extensive studies of the ambient stability of black phosphorus. While there has been an overall consensus that BP degrades in ambient conditions, precise mechanism have only been described recently.

In AFM and optical microscopy, thin flakes of BP are found to develop droplets on their surface over the course of hours, although these disappear again when samples are stored in vacuum. Left in ambient conditions, flakes shrink in size both vertically and laterally, leaving behind deposits in the form of large droplets. Flakes of 10 nm thickness were found to completely disappear over the course of 2 weeks\cite{124}. 

24
Stability in air was found to be dependent on sample thickness, with bulk BP stable for several months in ambient conditions, while monolayers degraded over the course of hours[115]. The fast degradation of thinner flakes is explained by their larger bandgap, which is closer to the energy levels of oxygen acceptor states, enhancing the charge transfer and therefore oxidation rate[115].

The electronic qualities of devices made from few-layer BP degrade quickly in air. The on/off ratio and mobility in field effect transistors was found to drop by 4 orders of magnitude over the course of 50 hours, at which point devices broke down[33].

Several theoretical studies have described the interaction mechanisms between BP and water and O$_2$, and experimental studies have evaluated exposure to each independently, as well as dependencies on light exposure and layer thickness.

Physisorbed molecular oxygen is metastable and dissociates to two oxygen atoms[122, 123] occupying the lone pair sites on adjacent P atoms. The dissociation energy of 0.54 eV is well below the thermal energy available at room temperature, 0.9 eV.

Water does not adsorb readily onto the pristine BP surface. The adsorption energy of water on BP was found to be slightly smaller than the intermolecular interaction energy of water, suggesting the BP surface is slightly hydrophobic[122, 32]. In ambient conditions, water aggregates in droplets rather than coating the surface. Water does adsorb on defective sites, such as step edges, which have incomplete bonding[125]. BP-based FETs immersed in normal water were found to decompose within 2 hours, while FETs immersed in deoxygenated water were stable for over 100 h after an initial drop in $I_{sd}/I$ to 60% [122]. The decomposition of black phosphorus in normal or DI water was attributed to dissolved oxygen in the water.

Samples exposed first to O$_2$ and then to water showed enhanced reactivity compared to the reverse process or exclusive exposure to one or the other[122]. Isotopic oxygen experiments found regions of O$_2$ concentration to coincide with H$_2$O concentration suggesting that if an area is already oxidised, water does readily chemisorbs onto it[122]. Hence water by itself does not seem to degrade the material, but rather the chemisorption of water is part of a multi-step degradation process[126]. Indeed, it was found that a water molecule can dissociate on two oxygen atoms in lone pair configurations, creating two OH groups in lone pair positions, and one O atom bridging P atoms on the top and bottom zigzag rows within a layer[32]. This configuration lowers the dissociation barrier for further oxygen molecules, accelerating the degradation. The resulting defective site attracts further oxidation and water adsorption. H$_3$PO$_4$ is produced and released from the surface, expanding the defective site to host fur-
ther oxidation[32]. This suggests that water and oxygen in large concentrations should cause the degradation of the material by forming nucleation sites and craters.

Light was found to accelerate oxidation reactions by promoting the charge transfer between lone pairs and superoxides[32]. Light does not cause degradation by itself, nor does it facilitate the dissociation of water.

2.8 Current passivation techniques and their limitations

To obtain long-lasting functional electronic devices, several capping layers and passivation methods have been explored[33]. FETs capped with 30 nm alumina deposited through ALD have an on/off ratio of 250 and mobility of 44 cm²V⁻¹s⁻¹ after 1 week ambient exposure after an initial dip in quality within the first 20 hours[124]. FETs made with Ni/Au contact rather than Ti/Au were found to have on/off ratio of 3000 and mobility 53 cm²V⁻¹s⁻¹ after two weeks in ambient conditions. Kim et al[127] found that flakes covered in thin capping layers of 2-3 nm were structurally stable, but still experienced electronic degradation, manifesting as a two orders of magnitude drop in sheet resistance along the perimeter of the flake. It was suggested the thin capping layer did not provide good side-wall coverage of the BP flake[127].

Heterostructures of two dimensional materials can be made with significantly improved stability[34]. Parts of thin flakes of black phosphorus covered with a graphene monolayer were stable for 48 hours before degradation started to creep under the graphene from the sides. Completely covering flakes in graphene may further improve stability. Perhaps the most elegant passivation mechanism is to harness the oxidation to form a saturated phosphate (P₂O₅) glass covering the flake. This protects it from further degradation and is transparent to IR and UV, which is critical for opto-electronic applications[128]. The oxide growth saturates after 2 days resulting in a 0.4 nm layer of oxide. Edmonds et al. do note that within the first 2 days, approximately 10-15 nm of phosphorus is etched away by P₂O₄ oxidation before the oxide saturates and passivates the surface[128]. One big XPS peak is associated with P₂O₅ while two smaller peaks represent dangling O and bridging O in P₄O₂ configurations. Higher energy XPS (probing deeper) shows significantly smaller peaks indicating the bulk is mostly pristine phosphorus.

A recent study by Zhao et al[129] showed ligands could be used to passify the lone pairs responsible for its degradation. Functionalising the surface with titanium sulfonate through mixing in NMP caused no changes to the BP structure. Flakes showed no signs of degradation after 72 hours exposure to ambient condition.
3 Characterisation techniques and instrumentation

STM/STS data, along with DFT and tight-binding calculations are presented in chapter 5 and 6 of this thesis. AFM, in vacuum and ambient conditions, and chemical analysis using XPS and SIMS is presented in chapters 5, while Raman spectroscopy at high pressure is described in chapter 7. This chapter offers descriptions of each of these techniques, along with experimental considerations relevant to the data presented.

3.1 Scanning Tunnelling Microscopy theory

The development of the Scanning Tunnelling Microscope (STM) in 1981, gave researchers access to the atomic and electronic structure at material surfaces, which were known to exhibit distinct characteristics from bulk but could up till then not be probed at the nanoscale.

Binnig and Rohrer presented an initial design in 1982[130] and were awarded the Nobel prize in 1986 for opening up the field of atomic resolution surface science. The original patent[131], shown in Fig. 3.1, is still very close to what is used nowadays.

Figure 3.1: The original STM patent published by Binnig and Rohrer[130]. A UHV chamber(1) is pumped(2) and cooled(3). The tip(5) is operated by three piezo drives (6,7,8) for lateral and vertical motion. The tip and sample(4) feed into a measuring unit(9) which controls the scanning parameters and transfers readings to an analysing unit(11) which plots(12) and visualises(13) data. A feedback control unit(10) adjusts the vertical piezo drive to maintain tunnelling contact.
Its key components include a sharp metal tip mounted on a piezo-electric tube which is used for highly controlled motion in three dimensions. An electronic feedback loop measures interaction between the tip and sample, and controls the piezo tube. The original design used a magnetic stage in a superconducting bowl to isolate the microscope from mechanical vibrations.

The extraordinary height and lateral resolution of a STM arises from quantum tunnelling, where the electron density of the tip and surface interact across a small vacuum gap[132]. The principle of a tunnelling junction is discussed in detail below.

### 3.1.1 Quantum Tunnelling

In classical mechanics, the momentum $p_z$ of an electron with energy $E$ moving in a region with potential $U$ can be described by:

$$\frac{p_z^2}{2m} + U(z) = E \tag{1}$$

where $m$ is the mass of an electron $(9.1 \times 10^{-28} \text{g})$.

As the $p_z^2$ term cannot be negative, the electron is unable to cross a region where $U(z) > E$. We say the electron cannot penetrate the potential barrier.

In a quantum mechanical approach, the same electron can be described as a wave, and has a wavefunction $\psi(z)$ that obeys the time-independent Schrödinger equation[133]:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dz^2} \psi(z) + U(z)\psi(z) = E\psi(z) \tag{2}$$

where $\hbar$ is the reduced Planck constant $(1.054 \times 10^{-34} \text{Js})$. We can represent a tunnelling junction by considering three regions, each with constant but different $U$. The equation can then be solved to yield solutions for each region individually.

Fig. 3.2 illustrates how this applies to a scanning tunnelling microscope. A metallic tip is brought to within a small distance (0.5 nm) of a sample, with a vacuum in between. The electron energy $E$ is less than the vacuum level, and in a classical model, the electron is unable to cross the vacuum.

In the classically allowed regions (tip and sample), where $E > U$, the Schrödinger equation has simple wave solutions:

$$\psi(z) = Ae^{ikz} + Be^{-ikz} \tag{3}$$
where the wave vector $k$ is described by:

$$k = \sqrt{\frac{2m(E - U)}{\hbar}}$$

(4)

In the vacuum region, where $E < U$, the Schrödinger equation can also be solved, and gives:

$$\psi(z) = \psi(0)e^{-\kappa z}$$

(5)

where we define the decay constant $\kappa$:

$$\kappa = \sqrt{\frac{2m(U - E)}{\hbar}}$$

(6)

For small energy difference $eV$ between the tip and sample Fermi levels, the potential barrier can be approximated as square and $U-E$ is equal to the sample work function, $\phi$.

The probability of finding a particle near a point $z$ is given by the square of the wavefunction:
which is non-zero inside the barrier. Hence there is some probability that an electron can be found in, or past, the potential barrier.

Clearly, the wider the vacuum region (bigger \( z \)), and the greater the barrier height, \( U-E \), the less likely an electron is to cross the barrier. In experimental terms, the further the tip is away from the sample, the smaller the tunnelling current.

As seen in eq. 6, the decay constant \( \kappa \) only depends on the work function of the material. A typical value for work function of 4 eV, results in a decay constant of 1 Å\(^{-1}\). We then find that the tunnelling current decays by a factor of \( e^2 \), or 7.4, per angstrom tip-sample separation.

This extraordinary sensitivity to vertical displacement also accounts for the atomic lateral resolution of STM. If we describe the STM tip as an inverted pyramid of metal atoms, there is a single apex atom that predominantly interacts with the sample, the next layer of atoms on the tip sits 1-2 Å further away from the sample, making contributions from these atoms more than an order of magnitude weaker.

3.1.2 Local density of states

Moving on from a quantum mechanical toy model, let’s set the electron tunnelling in a solid state context.

By applying a bias \( V \) between tip and sample, we create an energy difference \( eV \) between the tip and sample Fermi levels, and a tunnelling current can flow. For electrons to tunnel from sample to tip, there need to be available states in the tip at the same energy as the sample state. Hence the tunnelling current depends on the empty tip states within \( eV \) from the Fermi level:

\[
I \propto \sum_{E_F}^{E_F+eV} |\psi_n(0)|^2 e^{-2\kappa w} \tag{8}
\]

We can introduce here the local density of states (LDOS) which by definition is the normalised sum of all state probabilities within an energy range \( \epsilon \) as the range tends to 0. The LDOS is dependent on the probed energy \( E \), as well as the position \( z \).

\[
\rho_s(z, E) = \frac{1}{\epsilon} \sum_{E-\epsilon}^{E} |\psi_n(z)|^2 \tag{9}
\]
For small bias $V (< 100\text{mV})$, there is no significant variation in LDOS, and we can consider $\rho_s(z,E)$ independent of energy.

The tunnelling current at a position $z$ then reduces to:

$$I \propto \rho_s(z,E_F)e^{-2\kappa w}$$

(10)

Note that at this stage, we have made several approximations:

- The tip apex is sufficiently small that its LDOS is constant.
- There is no thermal excitation of electrons beyond the Fermi level.
- The tip-sample bias is small enough that the sample LDOS does not vary across the energy range.

For a more formalised description of STM tunnelling, we need to have a closer look at the work of Bardeen as well as the Tersoff-Hamann approximations.

### 3.1.3 Extended STM tunnelling models

Bardeen studied a model of two metals separated by a thin oxide layer[134], and used Oppenheimer’s solution to the time-dependent Schrödinger equation[135] to derive the transition probability (or tunnelling matrix element) using only the Hamiltonians of the two metals, which are known.

He defined the tunnelling matrix element between two states, $M_{\chi^-\psi}$, as the overlap between the wavefunctions of those states, as illustrated in Fig. 3.3 and described as:

$$M_{\chi^-\psi} = \frac{\hbar^2}{2m} \int_{\text{surface}} \chi^* \frac{\partial \psi}{\partial z} - \psi \frac{\partial \chi^*}{\partial z} dS$$

(11)

In one dimension, evaluated at a point $a$ within the overlap region, this reduces to:

$$M_{\chi^-\psi} = \frac{\hbar^2}{2m} (\chi^* \frac{d\psi}{dz} - \psi \frac{d\chi^*}{dz})|_a$$

(12)

Tersoff and Hamann used Bardeen’s work on the tunnelling matrix element to describe the tunnelling current[136][137] as:

$$I = \sum_{\text{spin states}} \frac{2\pi e}{\hbar} \sum_{s,t} |M_{s-t}|^2 \left[ \rho_s(E) f(E) \right] \left[ (\rho_t(E + eV)(1 - f(E + eV))) \right]$$

(13)
Figure 3.3: In one dimension, the transmission probability or tunnelling matrix element is simply defined as the overlap between two wavefunctions.

where we sum over all sample and tip states, the Fermi function \( f(E) \) is used to describe the distribution of electrons near the metal oxide interfaces, and \( M_{s-t} \) is the tunnelling matrix element between sample and tip states.

At low temperature, the Fermi function approximates to a sharp cut-off and we only need to consider the vacuum region of the tunnelling junction:

\[
I \approx 4\pi e \bar{\hbar} eV \int_0^{eV} |M_{s-t}|^2 \rho_s(\epsilon)\rho_t(\epsilon + eV) d\epsilon
\]

If we assume the tip can be represented as a isotropic point state, the tip DOS is constant with energy and \( \rho_t(\epsilon + eV) \) can be taken out of the integral. Furthermore, Bardeen showed that for a wide enough junction, where the overlap between wavefunctions is small, the matrix element does not vary significantly with energy.

The tunnelling current then further simplifies to:

\[
I \approx \frac{4\pi e}{\bar{\hbar}} eV |M_{s-t}|^2 \rho_s(\epsilon + eV) \int_0^{eV} \rho_s(\epsilon) d\epsilon
\]

Everything outside the integral is constant with energy, which brings us to the key point that the tunnelling current \( I \) is proportional to the integrated sample LDOS from the Fermi level up to the scanning bias energy.

While convenient for calculating tunnelling currents, the assumption of an even tip
DOS is not quite realistic, and there are examples of observations that can only be explained taking into account the tip DOS, such as the case of negative differential resistance, observed by Lyo and Avouris, where variations in IV measurements were caused by the tip DOS resonating with dopant states[138].

One specific feature of the tip DOS that is regularly considered is that the DOS is greater near the tip Fermi level. In experiments, this results in an increased sensitivity to states that line up with the Fermi energy of the tip, causing variations in feature intensity as different biases are used.

### 3.1.4 Tip-induced band bending

The work presented in this thesis involves the observation of charged defects, which require an understanding of local band bending.

The applied bias between tip and sample creates a potential difference between the two. This potential difference is partially dropped in the sample, creating a charge depletion region (in the case of a positively charged tip) directly under the tip[139, 140].

This charge variation due to the tip-sample potential difference is responsible for several features visible in topographies and spectroscopy.

![Figure 3.4: Potential energy profiles at 0.5, 1.0, and 2.0 eV for a probe tip radius of 4 nm, a tip–sample separation of 1 nm, and an n-type doping concentration of $10^{18}$ cm$^{-3}$. Figure reproduced from Feenstra, 2003[139]](image-url)
The total charge density in a semiconductor as a function of Fermi level position is given by

$$\rho(E_F) = e[p(E_F) + N_D^+(E_F) - n(E_F) - N_A^-(E_F)]$$

(16)

where p and n are the hole and electron densities in the valence and conduction bands, and $N_D^+(E_F)$ and $N_A^-(E_F)$ the densities of ionised donors and acceptors respectively.

An electrostatic potential $V$ produces a rigid energy shift $\phi = -eV$ to the Fermi level. Its position relative to the energy bands then becomes $E_F - \phi$ and the resulting charge density is given by $\rho(E_F - \phi)$.

This effect is well described by Feenstra in his 2003 publication[139], where he calculates the potential energy profiles for a tip of radius 4 nm and tip sample separation 1 nm. His calculations are illustrated in Fig. 3.4 and show that the potential difference between tip and sample create a quadratic electron depletion area. This manifests as a local shift in the Fermi level respective to the electron energy bands.

### 3.1.5 Defect-induced band bending

Alongside the STM tip, charged defects also generate an electrostatic potential and contribute to band bending.

The left panel in Fig. 3.5 illustrates two effects around a charged defect, in this case a Si substitution in Al/Si(111) as imaged by Hamers in 1988[141]. Within 5 Å of the defect, the tunnelling current is dominated by the LDOS of the defective site, and a sharp trough can be seen. On the scale of nanometres, an electrostatic effect is superimposed as carriers screen out the defect charge. This results in a local protrusion or depression depending on the polarity of the defect charge and scanning bias.

The potential around a charge defect creates localised band bending and can be described as

$$\phi(r) = -\frac{e}{\epsilon r} e^{-r/R_S}$$

(17)

where $r$ is distance from the defect, $\epsilon$ is the dielectric constant of the material, and $R_S$ is the characteristic screening length defined as:

$$R_S = \sqrt{\frac{\epsilon k_B T}{4 \pi n e^2} \frac{F_{\frac{1}{2}}(\frac{E_F}{k_B T})}{F_{\frac{1}{2}}(\frac{E_F}{k_B T})}}$$

(18)
Figure 3.5: a. STM topography line scan of a Si substitution in Al/Si(111). The defect state creates a deep local depression, while the positive charge on the defect creates an extended protrusion. b. Illustrated defect-induced band bending. The vertical axis corresponds to energy while the horizontal is tip-sample separation. A positive charge locally bends the bands down, which makes more states available in empty state tunnelling, producing a protrusion centered on the defect. Figure taken from Hamers, 1988[141]

where \( n \) is the density of free carriers and \( F_K \) is the Fermi integral of order \( K \), which describe the overlap between the Fermi function and the DOS of the material.

The right panel in Fig. 3.5 highlights how this potential locally pulls the valence and conduction bands down, which makes more states available to tunnel into, resulting in the protrusion seen in the left panel.

### 3.1.6 Scanning Tunnelling Spectroscopy (STS) and Current-Image Tunnelling Spectroscopy (CITS)

While topographies give a representation of the integrated LDOS at a specific energy across an area, we can inversely vary the energy difference between tip and sample at a fixed point.

The height and position of the tip are fixed, and the tip-sample bias is swept. The resulting changes in tunnelling current are mapped on an I(V) curve.

As detailed in section 3.1.3, for low temperature and constant matrix element, tunnelling current can be expressed as proportional to the available tunnelling states at the tip and sample between the Fermi level and the applied bias:

\[
I \propto \int_0^{eV} \rho_s(E_F - eV + \epsilon)\rho_t(E_F + \epsilon)\delta\epsilon
\]  

(19)
If we then again assume that the tip LDOS, \( \rho_t \) is constant, we find that

\[
\frac{dI}{dV} \propto \rho_s(E_F - eV)
\]  

(20)

Hence I(V) curves can be differentiated to obtain a spectrum proportional to the sample LDOS.

By sweeping the bias range at each pixel of a map, a three-dimensional current image can be produced, with two spatial directions and an energy dimension. When differentiated, these current-imaging tunnelling spectroscopy (CITS) maps offer an approximate visualisation of the LDOS of the surface.

A direct comparison is limited by two factors. As previously mentioned, tip LDOS variations may distort the I(V) features. Further complications arise from convolution of electronic and topographic data in the CITS map as at each pixel, the STM tip adjusts its height to achieve a predefined current setpoint. This convolution can be removed by calculating the decay constant of the material (ideally at each pixel) by taking I(z) measurements and scaling the I(V) spectra accordingly.

CITS with high spatial and energy resolution is very time-consuming and the spatial stability of the system should then be considered.

3.1.7 Thermal drift and piezo creep

A low resolution topography image can be obtained in a few minutes once a tunnelling current is established. Large-area high-resolution topographies and CITS maps, however, can take several hours to complete. At this timescale, two phenomena affect the spatial stability of the microscope.

Thermal drift arises from temperature changes in the environment, either from variations in the lab temperature, or from the microscope itself, for example when cryogen is depleted and the microscope starts warming up\[142, 143\]. Not all parts of the microscope have the same expansion coefficients, which typically vary in the range \(10^{-5} - 10^{-6} \text{ K}^{-1}\). Hence a 10 cm long component will expand/contract by up to a micron as temperature is changed by 1 °C. While components are made to have similar coefficients, there is still a translation of the tip relative to the sample and in smaller area scans can cause the tip to move away from the feature of interest.

Further instability is introduced by piezo creep, which is the overshooting in the motion of a piezo-electric scanner after a large voltage change. This effect is dependent on the size of the voltage change and dies down exponentially\[144, 145, 146\]. The effect is most noticeable after changing the scanning area or zooming in or out. If high
spatial stability is needed, several hours can be required to allow the piezo creep to fully settle.
3.1.8 STM design

The work presented in this thesis was performed on an Omicron LT-STM system operating at $10^{-10}$ mbar pressure and liquid nitrogen temperature.

The vacuum system is presented in Fig. 3.6(a) and consists of a microscope chamber and a preparation chamber which hosts evaporators and a sputter gun for sample preparation, as well as a residual gas analyser. Samples and tips are introduced through a separate loadlock. The entire system sits on active damping legs as an initial barrier to mechanical vibrations.

Beyond the active damping legs of the system, the microscope itself is further mechanically isolated through suspension from springs and eddy current damping along the sides of the microscope (Fig. 3.6(b) and (c)). Tips are placed in a holder on the piezo-actuator and samples are slotted in upside down above the tip. Scanning biases are applied to the sample while the tip remains grounded.

A cryostat surrounding the microscope allows cooling with liquid nitrogen or helium to 4 K. A ceramic pyrolytic BN heater can be used to counterheat the microscope to any desired temperature.

Si flux for sample dosing was obtained from a EFM 3-cell evaporator. The evaporator was loaded with a 99.999% Si rod of 2.0 mm diameter. A heated filament produces electron flux which locally heats the tip of the rod to evaporate the Si[110]. The Si flux is controlled by the filament current (intensity of electron bombardment) and by approaching or retracting the Si rod from the heating filament. A shutter allows for the controlled exposing of samples.

Scanning resolution strongly depends on the sharpness and stability of tips used.
Figure 3.7: Schematics illustrating the etching process of W tips. (a): The density difference between WO₄²⁻ and OH⁻ molecules creates a vortex at the air-solution interface, locally accelerating the etching process. Figure taken from Ju, 2011[151]. (b): Optical microscope image of W tip after chemical etching. (c): Field emission curve of tip after e-bombardement in vacuum, with applied bias on the horizontal axis and measured emission current on the vertical.

Tips are prepared in two stages: an initial chemical etching of tungsten wire and a further in-situ e-bombardement to remove the oxide layer tungsten accumulates in ambient conditions.

The chemical etching is done in a 3M KOH solution and follows the reactions:

\[
6H_2O + 6e^- \Rightarrow 3H_2(g) + 6OH^- \quad (21)
\]

\[
W(s) + 8OH^- \Rightarrow WO_4^{2-} + 4H_2O + 6e^- \quad (22)
\]

0.25 mm tungsten wire is centered in a circular filament cathode and inserted into the solution by about 2 mm. Surface tension creates a meniscus around the wire where the anodic reaction takes place. The produced tungstate (WO₄²⁻) dissolves in the solution and flows down the submerged part of the tip. This flow reduces the etching rate of the submerged wire, creating a bottleneck at the air-solution interface. As the wire thins, the current detected between the wire and the filament drops off linearly until the wire is too thin to support the weight of the submerged part, and breaks. The sudden drop in current is detected using a high speed microcontroller which immediately cuts the bias between cathode and anode and stops further etching.[147, 148, 149, 150]

The sharpness of the produced tip strongly depends on the shape of the meniscus at the wire and on how quickly power is cut when done. Keeping the tip position and the meniscus static produces the desired hyperbolic shape, while an immediate power cut ensures the tip apex is not blunted by further etching.

The procedure used at UCL routinely produces tips with radius of curvature around
20 nm.

After chemical etching, new tips are introduced into the vacuum system, outgassed at 150 °C overnight, and then further treated through e-bombardement to remove the insulating tungsten oxide layer.

Before and after e-bombardement, the quality of the tip is evaluated through field emission[152]. A negative potential is applied to the tip, and it is placed in close proximity of a grounded electrode. As the bias is swept from 0 to -1100 V, an emission current can be read between tip and electrode. The curvature of the emission curve can be related to the radius of curvature of the tip as described in [153].

3.1.9 Constant height vs constant current operation

A scanning tunnelling microscope is commonly operated in one of two modes, which relate to the use of the feedback loop during scanning.

All the data presented in this thesis is taken using constant current imaging. In this mode, the tip is approached to the sample until a desired tunnelling current is achieved. This current is used as a setpoint. As the tip is rastered across the surface, electronic and topographic features on the surface will change the tunnelling current. The feedback loop reads this change and adjusts the tip height to return the current to the setpoint. The required height adjustment is measured and used to map out the surface. The resulting image is called a topography, despite it being a convolution of both structural and electronic information. Rather than tracing out the actual structure of the sample, the tip follows the LDOS at the sample surface.

The other operating mode is constant height, and produces what is referred to as spectroscopy maps. In this mode. The tip is again approached until a current setpoint is achieved. However, during a single line of the image, the feedback loop is turned off, so that the tip height is constant, and variations in the tunnelling current are read instead. At the start of each new line, the current setpoint is re-established to maintain consistent height.
3.2 Atomic Force Microscopy theory

The development of the STM was a huge breakthrough in that it demonstrated for the first time that it’s possible to get atomic resolution on the surfaces of metals[154][130] and semiconductors[155]. This was a great boon to surface science, which quickly expanded as a field of study[156].

Only a few years after publishing the STM design, Binnig published his design for a theoretically even stronger microscope, the AFM[157][158]. AFM promised two great advantages. It could be used on any sample regardless of conductivity and it offered even finer spatial resolution.

STM and AFM are very similar techniques in that they both trace out a contour map of a surface by scanning a sharp tip across it, moderating its position through the interaction between tip and sample. However, where STM relies on a tunnelling current between the tip apex and sample atoms, AFM is sensitive to all force interactions between tip and sample. Specifically, where STM is only sensitive to the most loosely bonded electrons near the Fermi level, AFM observes all electrons, including core electrons. This should lead to even better resolution as the wavefunctions of core electrons are more confined than Fermi level ones.

During its development, AFM was quick to establish what is called apparent atomic resolution[159], reproducing the correct lattice spacings of the measured samples, but not resolving atomic defects or step edges[160][161]. The forces measured in these studies were well beyond what a single atom tip apex should be able to withstand[162]. It was found this resolution came from many tip atom-sample atom interactions.

A common macroscopic example is used to illustrate: if we trace a single egg along an egg box, its path will resolve the corrugation of each individual slot, as well as the variations from missing or imperfect slots or humps. However, if we trace the same egg box with a second egg box, the overall periodicity is still resolved, but individual humps and defects are missed. Hence a tip that is not atomically sharp may still resolve the atomic lattice, but not individual point defects. Nevertheless, with the development of sharper tips and more advanced operating modes, atomic resolution was eventually achieved. Along the way, 4 key challenges had to be met[163].

3.2.1 Short- and long-range forces

The first challenge to overcome is that the short-range chemical forces AFM relies on for its resolution are non-monotonic with distance. Quantum tunnelling interactions have
Figure 3.8: left: schematic of tip apex and sample surface. STM tunnelling and chemical interaction only occurs between the dark blue/red atoms while long range forces occur over several thousands of atoms. Right: in red, the exponential dependence of quantum tunnelling; in green, the inverse dependence of long range forces; in dark blue, the Lennard-Jones chemical potential; in black, the combination of short- and long-range forces in AFM. Reproduced from Giessibl, 2005[164]

an exponential decay with distance, as shown in red in Fig. 3.8 and the log of the signal lends itself well to a linear feedback system.

In AFM, interactions are less straightforward. The force between tip and sample, $F_{ts}$, has several components. There are long-range background forces which can act up to 100 nm away (light blue atoms and green curve in Fig. 3.8a. and b.) as well as the short-range chemical interaction with the apex atom (dark blue atoms and curve), which has a range of 1-2 Å. Other than very close to the sample, the long range interactions wash out any atomic resolution, creating our second challenge: the isolation of short and long range forces.

The long-range background is principally made up of two components: Van der Waals forces and electrostatic interactions. Van der Waals contributions come from the fluctuations in the electric dipole moment of atoms and their polarisation. Molecules that aren’t perfectly symmetrical have an electronic centre of mass which slightly shifts as the bonds in the molecules stretch and bend. Pulling the electric dipole moment off-centre polarises the molecule, allowing it to interact with neighbouring molecules.
The Van der Waals potential can be described as:

\[ V_{vdw} = -\frac{A_H R}{6z} \]  

(23)

where \( R \) is the radius of a spherical tip and \( A_H \) is the Hamaker constant which depends on the material of the tip and sample and is usually approx 1 eV. For a typical tip radius of 100 nm and distance 0.5 nm from the sample, the VdW energy is -30 eV, and the associated force -10 nN.

Electrostatic forces are interactions between charged objects and are described by:

\[ F = \frac{\pi \varepsilon_0 R U^2}{z} \]  

(24)

For the same tip radius and distance, and a potential \( U \) of 1 V, this amounts to -5.5 nN.

Both these contributions have an inverse dependence on distance, and are represented by the green curve in Fig. 3.8b.

Short range chemical interactions between individual atoms are described by the Lennard-Jones potential:

\[ V_{LJ} = -E_{bond}(\frac{2}{\sigma^6} - \frac{12}{\sigma^{12}}) \]  

(25)

where \( E_{bond} \) is the bonding energy and \( \sigma \) is the equilibrium distance.

Contrarily to STM, where the exponential decay of tunnelling current can be controlled through a linear feedback mechanism, the chemical potential is non-monotonic, consisting of an attractive and a repulsive regime.

3.2.2 Contact-mode (static) AFM

Contact-mode AFM operates in the repulsive regime. The AFM tip is brought in contact with the sample, and repulsive forces on the tip are measured. The tip is mounted on a flexible cantilever, with known spring constant \( k \). The deflection of the cantilever, \( q' \) is proportional to the applied force:

\[ F_{ts} = \frac{q'}{k} \]  

(26)

The cantilever must be softer than the sample atomic bonds (less than 10 nN/m)[165] to avoid the tip deforming the sample. For further noise and stability considerations, cantilevers with spring constants less than 1 nN/m are typically used.
This gives rise to the next challenge in AFM: as the tip is initially approached to the sample, the attractive forces will pull soft and flexible cantilevers towards the equilibrium position in the force curve. This 'jump to contact' with the sample results in an uncontrolled landing that damages atomically sharp tips[166][167].

The effect is reduced on chemically inert samples, on which perfectly sharp tips and atomic resolution have been achieved, but this method is not feasible for reactive surfaces with stronger short-range interactions.

The final hurdle to overcome is to measure the tiny forces involved with sufficient signal to noise ratio. In practice, what is measured is the deflection of the cantilever, hence we need to minimise deflection noise.

In summary, the four big challenges[164] to achieving atomic resolution in AFM are:

- Non-monotonic imaging signal
- Isolating short- from long-range forces
- Jump-to-contact upon landing
- Instrument noise in measuring small deflections

This first challenge is addressed by operating on either the attractive (non-contact mode) or the repulsive part (contact mode) of the force curve, and locally approximating the curve as linear.

### 3.2.3 Dynamic (non-contact) AFM: amplitude modulation

The development of dynamic mode AFM addresses the jump-to-contact challenge. In this mode, the cantilever is made to oscillate either with fixed frequency (amplitude-modulated[168]), or with fixed amplitude (frequency-modulated[169]), above the sample. This avoids jump-to-contact as long as the oscillating force at the lowest point of the oscillation, $A\kappa$, where $A$ is the amplitude of the oscillation and $k$ the spring constant of the cantilever, is stronger than the maximum attractive force experienced from the sample[170].

Initially Si cantilevers were used with a spring constant of 10 N/m which requires an oscillating amplitude of 10 nm. The introduction of much more rigid quartz cantilevers (1 kN/m) allowed for subnanometer oscillations.

While the subnanometer oscillations would be preferred in terms of singling out the short-range interactions, in practice, these deflections are too small for a reasonable signal to noise ratio.
3.2.4 Frequency-modulated AFM

In 1991, Albrecht[169] developed frequency-modulated AFM, which brought an order of magnitude improved sensitivity to the measurement of chemical forces and addresses the last two challenges.

Rather than measuring tiny changes in the oscillation amplitude, we look for changes in the oscillation frequency, which are much easier to resolve.

The strength of FM-AFM is the $1/f$ dependence of noise intensity. Fig. 3.9 illustrates how the noise density decreases as oscillating frequency is increased, until a base white noise level is reached in the kHz range. By picking cantilevers with natural frequencies beyond this, noise is minimised and the frequency resolution, and therefore force resolution, is improved.

This then leaves the final question of how to isolate short range from long range forces.

![Figure 3.9: The noise in cantilever deflection shows an inverse dependence on cantilever oscillation frequency, until a minimum white noise level is reached in the kHz range. Reproduced from Morita[163]](image)
In FM-AFM, a cantilever is driven at its natural frequency $f_0$ through positive feedback to maintain a set amplitude.

The natural frequency of a cantilever is given by:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (27)$$

where $k$ is the spring constant and $m$ the effective mass of the cantilever.

Through interaction with the sample, this frequency is shifted by $\Delta f$:

$$f = f_0 + \Delta f = \frac{1}{2\pi} \sqrt{\frac{k + k_{ts}}{m}} \quad (28)$$

where we label $k_{ts}$ the force gradient of the tip sample interaction.

For $k_{ts} << k$, we can expand the square root to obtain:

$$\Delta f = \frac{f_0}{2k} k_{ts} \quad (29)$$

and we find that the frequency shift is proportional to the force gradient. Short range chemical forces have a much steeper force gradient than Van der Waals or electrostatic forces. By using small oscillating amplitudes that stay within the range of chemical forces, we can maximise their contribution to the signal. The effect can be enhanced by using higher harmonics of the frequency shift, which are proportional to the second derivatives of forces, at which point the signal comes almost exclusively from short range forces.

Hence using small amplitude FM-AFM has two key advantages: there is a good signal to noise ratio from the high oscillating frequency, and it is primarily sensitive to short range forces.

This simple expression of the frequency shift as proportional to force gradient is only valid if the force gradient $k_{ts}$ is constant throughout an oscillation cycle. For chemical interactions, this limits us to oscillations amplitudes of about 1 Å.

There are two problems standing in the way: at these small amplitudes, the withdrawing force $kA$ is not enough to avoid jump-to-contact for most cantilevers, and this close to the surface, shear forces tend to damage tips. Both of these problems can be addressed by using harder and stiffer cantilevers, recently developed in the form of quartz tuning forks, although this has to be balanced with the trade off of stiffer cantilever experiencing smaller frequency shifts.

In practice, larger oscillation amplitudes on the order of 1-2 nm are typically preferred, as they are more easily detected. In the case of large amplitudes, $k_{ts}$ is not
constant and we instead describe the oscillation using a normalised frequency shift $\gamma$ obtained from first order perturbation theory.

$$\gamma = \frac{kA^2}{f_0} \Delta f$$

(30)

3.2.5 AFM design

![Diagram of AFM design](image)

Figure 3.10: Block diagram of the different components of a frequency-modulated AFM system.

To a large extent, STM and vacuum AFM designs are very similar. There is an overall vacuum system and mechanical vibration damping mechanism. The sample is mounted on a piezo-actuated stage, while the tip is mounted on an actuator maintaining the oscillation of the cantilever. The back of the cantilever is exposed to a deflection measuring scheme reproducing the cantilever oscillation as an electrical signal. A bandpass filter isolates a region around the eigenfrequency of the oscillator to filter out noise. The signal is send through a phase shifter which offsets the phase by $-\pi/2$. An automatic gain control unit amplifies the signal which is used to drive the cantilever oscillation. The same signal is used for processing and plotting.
Cantilevers are characterised by their spring constant $k$, eigenfrequency (natural frequency of free vibrations) $f_0$ and their quality factor $Q$. Resonators with high quality factors have low damping and require lower driving voltages to maintain their oscillation. The reduced damping allows for better resolution in the amplitude and frequency measurements. For micromachined cantilevers, $Q$ ranges $10^2$ in air and $10^5$ in vacuum.

In the first AFM, deflection was measured with an STM positioned on the back of the cantilever. This had the drawbacks of applying a force on the cantilever and the difficulty of approaching the STM tip to the cantilever. Modern AFMs rely on photodetection or interferometry mechanisms. In ambient AFM, this typically consists of a laser reflected from the back of the cantilever onto a quadrant segmented photo-diode. This not only resolves the up and down motion of the cantilever, but also torque applied to it by friction effects.

An alternative deflection detection system uses Fabry-Perot interferometry[171]. A single mode optical fibre is closely approached to the back of the cantilever. A monochromatic light source shines onto the back of the cantilever and is reflected back into the optical fibre. The interference between incident and reflected light offer very high resolution to small cantilever movements but is prone to misalignment and signal loss as the back of the cantilever is small and may be rough. The fibre needs to be approached to within one or two wavelengths (less than 1.5 micron) to maximise the returning signal[172]. The distance between fibre and cantilever is then further adjusted in such a way that the cantilever oscillation operates on the steepest part of the interference wave, for maximum sensitivity to changes in $z$.

Figure 3.11: interferometer alignment: As the optical fibre is approached to the back of the cantilever, signal strength improves until it makes contact and bends the fibre. The fibre is then retracted until we are operating on the steepest part of the first rising or falling edge. Figure from Nanoscience Group Hamburg.
3.3 X-ray Photo-electron Spectroscopy

X-ray photo-electron spectroscopy (XPS) is a surface sensitive chemical characterisation technique. It can probe the first 10 nm or less of a surface and quantify chemical composition, as well as the chemical state of the components, to a parts per thousand level. In 1954, the research group of Kai Siegbahn produced the first high resolution XPS spectrum[173], for which he was awarded a Nobel prize in 1981.

XPS relies on the photo-electric effect, illustrated in the left panel of Fig. 3.12, which was initially observed by Hertz in 1887[174], and later explained by Einstein in 1905[175], for which he was eventually awarded a Nobel prize (1921). The effect describes how a photon with sufficient energy can excite an electron beyond its ionisation energy and eject it from its atom. The remaining energy is converted to the kinetic energy of the electron.

The experimental set-up of XPS is relatively simple. As shown in Fig. 3.13, the sample is illuminated by a monochromatic x-ray source (or more commonly a general x-ray source equipped with a monochromator) of known photon energy.

![Figure 3.12: Left: The photo-electric effect. An incident photon is absorbed by a core electron (here K), which is ejected from the sample. The kinetic energy of the electron is the photon energy minus the binding energy of the core state. Right: The Auger effect. The hole left by the ejected electron is filled by an electron from a higher energy shell. The transition energy can be transferred onto a third electron, which is also ejected.](image)
Radiation is produced using an electron gun on an Al target. The Al $k_\alpha$ excitation is isolated by irradiating a thin disc of quartz, and has well known energy 1487.6 eV. The x-rays ionise core electrons from the sample, which escape the surface. The ejected electrons are collected by an electron lens, separated by kinetic energy and then counted. A plot of electron count over kinetic energy is then produced.

The principle of measuring the kinetic energy of the electrons dates back to 1914[176], when Rutherford showed that the measured kinetic energy is the difference between the energy of the incident photon and the energy spent escaping the sample, the binding energy:

$$E_{binding} = h\nu - E_{kinetic}$$ (31)

The binding energy is specific to each electronic shell in each element. Consequently, the combined observation of specific peaks (from different shells of the same element) allows the elemental characterisation of the sample. In practice, an extra term needs to be added to this equation. To avoid the build-up of charge from electrons leaving the sample, both the sample and spectrometer are grounded, so that their Fermi levels align. We then need to account for the difference in work function of the sample and the spectrometer, described as $\phi$. The practical binding energy is then:
\[ E_{\text{binding}} = h\nu - (E_{\text{kinetic}} + \phi) \] (32)

\( \phi \) is commonly used as a correction factor to shift the spectrum to align reference peaks. Note that for insulating samples, using a common ground on instrument and sample is not effective. A low energy electron gun can then be used to provide a steady state potential at the surface. Chemical bonds, and oxidation states, can change the binding energy of specific shells, and hence cause the corresponding XPS peaks to shift. Matching the exact peak position to a database can indicate the bonding and oxidation state of the observed element.

As well as primary electron peaks, an XPS spectrum will contain Auger peaks. The Auger effect is summarised in the right panel of Fig. 3.12. After a core electron is ejected through the photo-electric effect, the hole left behind is filled by a higher energy electron decaying[177]. The transition energy can be transferred onto a third electron, which is also ejected and can be detected with an energy:

\[ E_{\text{Auger}} = E_L - E_K - E_{\text{binding}} \] (33)

where \( E_K \) and \( E_L \) are the energy levels of the photo-electric electron and the decaying electron respectively. Auger peaks are labelled by the energy level quantum numbers of the three electrons involved. A KLL Auger peak comes from a 1s electron being ionised. A 2s or 2p electron decays into the hole left, and transfers its energy to another 2s or 2p electron, which is ejected.

The energy resolution of XPS is limited by the FWHM of the produced x-rays, along with the energy resolution of the spectrometer. Typically resolutions of 0.5 eV are achieved in labs while synchrotron radiation can offer energy resolutions of 0.2 eV[178].

The surface sensitivity of XPS arises not from the penetration depth of x-rays, but from the inelastic collisions, electronic recombination, and electron trapping electrons are exposed to as they escape the surface, resulting in an exponential decrease of signal intensity with depth. To further enhance surface sensitivity, the incidence angle can be reduced. Using angles of less than 1°, the sensitivity can be refined to the first 1-2 nm of the surface[179]. The background of an XPS spectrum is primarily made up of electrons that lost energy to inelastic scattering but did manage to escape the sample. To accurately compare the intensity of peaks and deduce sample composition, we need to consider transitions from different elements, and even from different electronic states of the same element, will vary in intensity. We can normalise peak intensity using standardised relative sensitivity factors (RSF) for each transition.
3.4 Secondary Ion Mass Spectrometry

Secondary Ion Mass Spectrometry (SIMS) is an extremely sensitive destructive surface characterisation technique able to identify elemental and isotopic composition with a detection limit of ppm to ppb, depending on how readily ionised the target is.

Fig. 3.14 illustrates the principle behind SIMS. A primary ion beam is produced from gas and accelerated to a desired energy. \( O_2^- \) and \( Cs^+ \) are commonly used primary gases and enhance the generation of positive and negative ions respectively. The incident beam can be focused to a spot size less than a micron, and sputters the first 1-2 nm of a surface[180]. Sample molecules, as well as their derivatives, and combinations produced from the incident beam (e.g. oxides), are ejected from the surface. Neutral molecules are not detected, but charged ones are identified and counted in a mass spectrometer. To avoid detecting anything other than the sample, SIMS are contained in vacuum systems. A low energy electron gun is used to offset the charge from the primary beam.

The sputtering yield, given by the number of scattered particles per incident particle, depends on the primary beam energy and typically ranges 5-15[181]. Not all these scattering events result in ejected particles. Following the collision cascade model, particles can get dislodged and pushed deeper into the sample, resulting in a 10 nm penetration layer where molecules from the immediate surface and the bulk are mixed[182]. As well as surface molecules being pushed deeper into the sample, the primary ions are implanted in the surface. Eventually, the concentration of primary ions reaches a steady state, where they are implanted at the same rate as they are sputtered from the surface.

As well as offering a full spectrum of present species in a sample, SIMS is able to produce ion images, mapping out the distribution of a specific species within the scanning area. While SIMS is extremely sensitive to dopant species thanks to its low background, quantifying dopant concentrations is difficult. The secondary ion yield strongly depends on the chemical species, but also on the chemical environment of those species, along with sputtering parameters such as primary ion, energy, and incident angle. Typically species with low ionisation potential (for positive species), or high electron affinity (for negative species) have high relative ion yields. The implantation of negative \( O_2^- \) or positive \( Cs^+ \) ions further alters the work function of sample species. Similarly to XPS, relative sensitivity factors (RSFs) are used to convert particle count to concentration. However, as mentioned previously, these factors vary strongly on the chemical environment of the species in the sample. Consequently, in practice, SIMS can only be used quantitatively if reference samples of known composition are
Figure 3.14: Schematic of SIMS operation: Primary ions (green) are fired at the sample, and sputter the top 1 nm of the surface (sampling layer). Surface particles and their derivatives are ejected from the surface and are identified and counted in a mass spectrometer. Primary ions are implanted in the surface, and cascade collisions cause a mixing of the surface atoms and the sample up to 10 nm deep (penetration layer).

A SIMS set-up can be operated in two modes, depending on the energy of the primary beam. A very low energy beam is able to achieve a full mass spectrum while sputtering less than a tenth of a monolayer and can be used to identify thin films without mixing with the substrate. This is called static SIMS, in contrast to dynamic SIMS, where a higher energy primary beam produces a sputtering rate of 0.5-5 nm/s. In this way, depth profile information can be obtained. For depth profiles, it is important to consider that sputtering rate is dependent on crystal orientation, and in our case, having different crystallites in the sputtering area results in uneven sputtering. The roughness of the crater can further be enhanced by pre-existing lattice defects.
3.5 Raman spectroscopy

Raman spectroscopy is a commonly used technique that provides a fingerprint spectrum of vibrational excitations in a molecule. It can be used as an identification tool, as each spectrum is unique to one molecule, or to track changes in vibrational modes within a sample. Raman spectroscopy relies on the inelastic scattering of laser light as it interacts with lattice vibrations, or phonons[183, 184].

Fig. 3.15 shows a schematic of a typical Raman spectrometer. Monochromatic light in the near-IR, visible, or near-UV range is generated typically by a He:Ne laser or a laser diode and focused onto the sample through an optical microscope. Radiation from the illuminated area is collected with a lens and passes through a notch filter, which stops photons that did not experience a change in energy. The remaining radiation is dispersed onto a CCD detector to count the number of photons with specific energies.

Figure 3.15: Schematic of a Raman spectrometer. In our set-up, a 532 nm laser is focused through an optical microscope with microscope objectives x5-x20.
Fig. 3.16 shows a schematic of the three possible ways of scattering. The incident light excites electrons in the sample, raising them to virtual excited states. The majority of electrons are Rayleigh scattered and decay back down to their original state without interactions, emitting a photon with the same energy as the one absorbed.

About 1 in 10 million photons experience a slight shift in energy due to interactions with phonons, vibrational excitations of the lattice. In the case of Stokes scattering, the excited electron loses energy in the interaction with the phonon, promoting the phonon to a higher vibrational state while losing $\Delta E$. As the electron decays back to the ground state, it emits a photon. The lost energy results in a decrease $\Delta \nu$ in the photon frequency. In the case of anti-Stokes scattering, the electron gains energy from the interaction. This can only occur if a lattice phonon was in an excited state and the interaction causes it to relax. This results in an energy gain $\Delta E$ and hence an increased frequency of the emitted photon. This anti-Stokes scattering is weaker, because for a system in thermal equilibrium, the populations of excited phonons decrease exponentially. Other than their intensities, the Stokes and anti-Stokes spectra are symmetrical. Consequently, it is common to only record the more intense Stokes spectrum.

A plot of photon counts with energy shift, called Raman spectrum, allows us to map out the vibrational modes that are unique to the material. While Raman spectroscopy is
a very convenient and quick technique, not all lattice vibrations are Raman active. Only excitations between vibrational states with different polarisability are allowed and produce a Raman peak. Hence only vibrations that move the electronic centre of mass of a molecule can be identified. In practice, this implies that materials with higher symmetry produce fewer peaks. The simple cubic phase is perfectly symmetrical and doesn’t have a Raman spectrum at all.

3.6 High pressure measurements using diamond anvil cells

Diamond Anvil Cells (DACs) were popularised in the 1960’s as effective tools for high pressure spectroscopy studies when Weir et al [185] at the NBS labs produced a cell capable of reaching pressures of 3 GPa (30 Kbar). Previously, the Bridgman anvil was used for high pressure electrical measurements up to 10 GPa, but the unwieldy apparatus didn’t allow for accurate pressure reading, and lacked access for spectroscopic measurements. The use of diamonds to focus the pressure onto the sample not only allowed for much higher pressures thanks to the hardness of the material, but also produced a transparent optical path through the sample.

While there have been several variations of the DAC[186], all designs rely on a few key components. All cells consist of two high-purity diamonds, typically synthetic, with their cuvet ground down to 300-500 micron octagonal faces. The diamonds are aligned and pressed together to exert pressure on the sample between them. The force-generating mechanism, and the alignment method vary between DACs.

In our experiments, the diamonds sit in a stainless steel casing, which is tightened by screws and Belleville spring washers. The lower diamond is fixed in place, while the upper diamond can be translated with three adjusting screws at 120° angles to align the diamonds. The sample is held in place by a stainless steel or rhenium gasket with a 200 micron hole. The gasket improves the structural integrity of the cell and helps distribute pressure evenly. The casing has cut-outs above and below the diamonds for transmission measurements, as well as in plane with the sample for diamond alignment.

The sample chamber produced by the hole in the gasket and the two diamond faces is filled with a pressure-dispersing medium, which maintains hydrostaticity, homogeneous pressure from all directions. In our experiments, we have used solid NaCl and liquid $N_2$ as pressure media, which offer hydrostatic conditions at low and high pressures respectively. Further details of sample loading are given in the results chapter.

The use of ruby crystal chips as pressure gauge was developed in 1972 by Forman et al[188], also at NBS labs. It relies on the well-documented shift of photoluminescence
peaks in the ruby spectrum with pressure. Cr impurities replace Al in the \( \text{Al}_2\text{O}_3 \) lattice and produce two decays from a split 3d orbital, with wavelengths 692.8 and 694.3 nm in ambient conditions[189].

The change with pressure of these peaks is best modelled by an adapted version of the Murnaghan equation:

\[
P(GPa) = \frac{A}{B} \left( \left[ 1 + \frac{\Delta \lambda}{\lambda_0} \right]^B - 1 \right)
\]

which expresses the compression of a body when exposed to pressure. A and B are calibration factors. \( \lambda_0 \) represents the optical excitation wavelength of ruby at ambient pressure. For pressures below 80 GPa, Mao et al[190] show that a linear approximation is sufficient and the equation simplifies to

\[
P(GPa) = \frac{\Delta \lambda}{0.365}
\]

The reliability of ruby as a pressure gauge with linear dependence in nitrogen and NaCl was shown by You et al [191]. The strong peaks, as well as the convenience of photoluminescence measurements in parallel with Raman spectroscopy make ruby the pressure gauge of choice in DACs.
3.7 Density Functional Theory

Density functional theory is a quantum mechanical modelling method that can compute the electronic ground state in many-body problems[192, 193].

In their essence, what these calculations seek to obtain is the energy for a specific configuration of atoms, and how that energy changes when the atom positions change.

The first step to tackling this problem is to use the Born-Oppenheimer approximation[194], which states that atom nuclei are stationary on the timescale of electron movement, to separate the problem into two steps: we first fix all nuclei in a specific position and calculate the lowest energy configuration of the electrons around them, and secondly we explore changes as we move nuclei around.

To find the lowest energy configuration, the ground state, for a specific set of nuclei, we need to solve the time-independent Schrödinger equation:

\[
\psi = E\psi
\]  

where the three terms on the left correspond, in order, to the electron kinetic energy, the interaction between each electron and the combined potential of nuclei, and the interaction between each pair of electrons.

\(\psi\) here corresponds to the overall electronic wavefunction incorporating all N electrons. With 3 dimensions per electron, and many electrons per nuclei, this quickly becomes impossible to solve, especially considering the third term of the equation, which shows that the wavefunction of each individual electron is dependent on the wavefunction of every other electron.

What DFT computes instead, is the electron density of the entire surface, rather than the individual wavefunctions. Where the Schrödinger equation has 3 dimensions to compute for each electron, the electron density only has 3 for the surface as a whole. By focusing on the electron density, we give up information of each individual electron, but seeing as electrons are indistinguishable anyway, this is no big price to pay.

The problem is then how to describe the electron density of a surface. For this, two key theorems are used, proven by Hohenberg and Kohn[195]. The first states that The ground-state energy from Schrödinger’s equation is a unique functional of the electron density. This means there exists a one-to-one mapping of the ground state wavefunction to the ground state electron density. While this proved that determining the electron density was sufficient to determine the system energy, and that there exists a functional to do this, it gave no clue as to what that functional might be.
The second theorem states that *The electron density that minimises the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger equation*. This implies that rather than solving the functional absolutely, we can try lots of different densities and consider the one with the lowest energy the density the best. The energy functional is typically approximated by a sum of several known factors: electron kinetic energy, Coulomb interaction between electrons, and Coulomb interaction between electrons and nuclei. There further is an exchange-correlation functional, which combines all remaining effects in a correction factor. Kohn and Sham[196] furthered showed that finding the right electron density can be broken down into solving a set of single-electron equations:

\[
\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) + V_{\text{H}}(r) + V_{\text{XC}}(r) \right] \psi_i(r) = \epsilon_i \psi_i(r)
\]

(37)

The first two terms on the left correspond to our 'known' factors. \(V_{\text{H}}\) is the Hartree potential[197] and describes the Coulomb repulsion between the electron considered and the overall electron density.

This leaves us with a problem. To solve the Kohn-Sham single-electron equations, we need to define the Hartree potential, for which we need to know the overall electron density, which is exactly what we are trying to find by solving the equations.

To break the cycle, we take an iterative approach with the following steps:

- Define a trial electron density \(n(r)\) and use it to calculate \(V_{\text{H}}(r)\).
- Solve the Kohn-Sham equations to find the single particle wave functions \(\psi_i(r)\).
- Calculate the electron density by combining and squaring the single particle wave functions.
- Compare the calculated electron density to the trial electron density. If they are the same, we have the ground state electron density. If not, the trial density is adjusted and we repeat the process.

Exactly how to adjust the trial density varies for each implementation of the method. Many approximation to the unknown \(V_{\text{XC}}(r)\) term have been developed. Most commonly used are the local density approximation (LDA) and generalised gradient approximation (GGA), which include the Perdew-Burke-Ernzerhof functional (PBE) used in the calculations presented here.
3.8 Tight-Binding calculations

With today’s computing power, DFT calculations can simulate systems of several hundred to thousands of atoms. For bigger systems, we need to use more approximate techniques.

Tight binding is a semi-empirical approach to calculating the band structure of a material, fast enough to calculate systems of tens of thousands of atoms [198, 199].

By assuming electrons are ‘tightly bound’ to their nuclei and have limited interaction with others, their wavefunction can be approximated as atomic orbitals of a free atom. The limited interaction between electrons is incorporated as a small correction to the hamiltonian:

\[ H(r) = \sum_{R_n} H_{at}(r - R_n) + \Delta U(r) \quad (38) \]

where \( R_n \) are the nuclei positions.

Assuming small interactions between electrons also allows us to treat this many body problem as the sum of single-electron wavefunctions.

Solutions to the time-independent Schrödinger equation are then approximated as linear combinations of atomic orbitals \( \phi \):

\[ \psi_m(r) = \sum_{R_n} b_m(R_n) \phi_m(r - R_n) \quad (39) \]

The Bloch theorem states that wavefunctions in a crystal can change only by a phase factor when translated, so that the coefficients have to satisfy:

\[ b_m(R_l) = e^{i k \cdot R_l} b_m(0) \quad (40) \]

Normalising the wavefunction to unity yields:

\[ \psi_m(r) = \frac{1}{\sqrt{N}} \sum_{R_n} e^{i k \cdot R_n} \phi_m(r - R_n) \quad (41) \]

This form of the wavefunctions, along with the assumption that only the nth energy level contributes to the nth energy band, allows the calculation of the Bloch energies. The expression for the energies has three matrix element components.

\[ \epsilon_m(k) = E_m - \beta_m + \sum_{R_n \neq 0} \sum_l e^{i k \cdot R_n} \gamma_{m,l}(R_n) \left( 1 + \sum_{R_n \neq 0} \sum_l e^{i k \cdot R_n} \alpha_{m,l}(R_n) \right) \quad (42) \]
\( \alpha_{m,l} \) denotes the overlap integral between atomic orbitals \( m \) and \( l \). \( \beta_m \) denotes the atomic energy shift due to the potential of adjacent atoms, and \( \gamma_{m,l}(R_n) \) is the interatomic matrix element, or bond energy[200]. If the tight bonding model is a good representation of a structure, then there is little influence from adjacent atoms and \( \beta \) matrix elements are small. Similarly, the overlap integral \( \alpha \) is negligible in a situation where a tight binding model is appropriate. This reduces the problem to determining \( \gamma \). While it is possible to calculate exact bond energies if the atomic wavefunctions and potentials are known, typically \( \gamma \) matrix elements are parametrised either from look up tables, or, as in our case, using results from DFT calculations.
4 Sample origin and preparation

The experiments in this thesis were performed on black phosphorus samples prepared through the two available synthesis pathways: a vapour transport method and a high pressure method.

Vapour grown black phosphorus was purchased from SmartElements, Austria[201]. The purity of the samples is quoted as 99.998% based on SIMS characterisation. High pressure black phosphorus crystals were obtained from 2DSemiconductors[106], based in the US. The material is quoted with a 99.9995% purity based on SIMS characterisation. The most common contaminants were identified as iodine at 10 ppm and calcium at 7 ppm. Due to their sensitivity to ambient degradation, we kept the black phosphorus crystals in an argon glovebox. When a pristine surface was required, samples were prepared in ambient conditions, and a post cleaving technique was used to split flakes open to reveal unoxidised surfaces.

Figure 4.1: top left: high pressure black phosphorus clusters as-received, approximately 1 cm across. The top right and bottom panels show the isolation of individual crystals, the fixing onto sample plates, and the completed sample with cleaving post glued on. The flake shown is approximately 2.5 by 1.5 mm in size.
The top left panel of Fig. 4.1 shows the high pressure black phosphorus as received. Samples were received as clusters of crystals, typically 0.5 to 1 cm in diameter. Crystal size and orientation varies strongly within the clusters, with many crystals growing as needles up to 0.5 cm in length but perhaps 0.1 cm wide. The inconsistent crystal orientation in particular makes the cleaving and isolation of single crystals difficult.

The top right panel shows an isolated flake, 2x2 mm in size, cut with a razor from the clusters. A razor or scalpel is a good method for isolating useful flakes, as there is a distinct difference in the force required when cutting in the out of plane direction as opposed to cutting through crystal layers. Hence careful probing with the corner of a razor can reveal easy cleaving planes. The isolated flake is then glued onto a stainless steel sample plate for STM and AFM measurements using a UHV compatible silver epoxy, EPOTEK H21D[202]. When gluing the samples, care is taken to ensure epoxy covers at least one side of the sample and contacts the top surface, to ensure a good conductive pathway. The epoxy is cured on a heating plate at 120°C for 1 hour.

The bottom panels show a top and side view of the flake glued onto the sample plate, and a top view of the finished sample, with a 2 mm diameter ceramic post glued on top of the sample using the same epoxy.

Once in the vacuum system, the cleaving post is knocked off the sample, either manually with the sample manipulator arm, or with a spring-loaded metal rod mounted onto a UHV chamber port. Thanks to the weak interlayer bonds, the cleaving post rips part of the sample with it, exposing a pristine surface, which has not suffered ambient degradation. The thickness of the flake is not of great consequence. Even thin samples can be cleaved, as the epoxy connecting to the sample plate and the cleaving post makes a much stronger connection than the interlayer interactions in black phosphorus.

What is essential, is that the flake is properly aligned so that its layers sit horizontally. When this is not the case, and the sample is cleaved, the cleaving plane shears through several layers, lifting up layer edges so that they stick out from the sample. These flakes sticking out of the surface are hard to see once the sample is in the UHV system, and are too bendable to land the STM or AFM tip on. Upon approaching the sample, the tip tends to bump into these loose flakes, ruining the atomically sharp tip apex.
5 Degradation and induced defects in black phosphorus

In this chapter, I discuss the characterisation of the ambient degradation process in black phosphorus and its mechanism, as well as an atomic scale study of inherent and induced defects on its surface. I address the chemical purity of both vapour grown and high pressure grown BP samples to inform the atomic scale study in AFM and STM of the defects. I used ambient AFM to study the degradation of BP in air and in a reduced humidity environment. Using UHV STM, I induced further defects to compare to inherent defects and to simulate sample degradation.

The discussion of the charged double lobed defects consistently observed in STM studies is reserved for chapter 6.

5.1 Exposing atomically clean surfaces for UHV experiments

The successful study of pristine BP samples, especially in UHV microscopy such as STM and NC-AFM, relies on the ability to expose atomically flat and clean bulk surfaces. There are several common ways of cleaving layered materials[9, 203, 204]. To test the effectiveness of different cleaving techniques on BP, I took SEM images of samples prepared through tape, razor, and post cleaving.

Fig. 5.1 a. shows a sample that was cleaved repeatedly with sticky tape. While regular office tape is effective, it is more common to use blue fluortape, which has a milder adhesive making it easier to deposit flakes from the tape onto a substrate once exfoliated. This method is by far the gentlest, and results in smooth terraces in areas where the cleave is successful. Tape cleaving, sometimes also achieved with a PDMS stamp as opposed to scotch tape, has a lot of success with other 2D materials, such as MoS$_2$[203], of which samples can be obtained with much bigger single crystal sizes. Because BP samples tend to contain several flakes with varying crystal orientation in the same cluster, the tape does not adhere evenly across all of them, and it often takes several cleaves to completely clean the surface. This is not practical in a vacuum system where you only have one opportunity to pull off the tape. Furthermore, the tape lifts some smaller flakes up, as seen in white contrast on the SEM image. The adhesion with the tape is not strong enough to remove an entire flake or layer, but instead leaves it dangling from the surface, which is a problem for scanning probe measurements. The lifted flake may be poorly conductive due to line defects and crystal edges, may be pushed down like a spring, yielding irregular force interactions, and is covered in adhesive from the tape.
Figure 5.1: SEM images of BP samples cleaved using three common methods. While post-cleaving yields the smoothest surface, the force of the exfoliation breaks up crystals, resulting in smaller useful areas. All images were taken using a 20 kV accelerating voltage.

Panel b. shows a sample cleaved with a razor blade, which is used in vacuum systems to cleave hard crystals such as NaCl and perovskites. In the case of BP, the cleave however results in a very rough surface, leaving lots of debris as it scrapes the sample. The resulting surface is in no way smooth enough for STM or AFM measurements with atomic resolution. The cleaving method is only suitable for brittle materials that snap along a lattice direction when pressure is applied, the way a silicon wafer does.

Panels c. and d. both show post cleaved samples. The post used to cleave this sample was only 1 mm in diameter, and did not cover the entire sample, which was found to be essential for a smooth cleave. The top right of image c. shows a roughly circular crater where the post was ripped away. As the image shows, the post only removes the part of the flake directly under it, leaving a jagged edge. The newly exposed surface however, shown in panel d., contains areas of over 100x100 micron that are smooth and clear of debris. These areas are large enough to approach visually with an STM or AFM tip.
5.2 Chemical Analysis

As shown in chapter 4, the high-pressure samples as-received showed odd discoloration reminiscent of iodine. To evaluate the purity of the black phosphorus samples we procured, and to support our study of its defects, we performed chemical analysis using lab-based XPS and SIMS.

5.2.1 XPS of pristine and contaminated black phosphorus

Figure 5.2: XPS analysis of vapour-grown and high-pressure grown black phosphorus. Left: survey spectra resolving P and C peaks. Right: high-resolution spectra of present impurities.

XPS survey scans were taken with measurement step sizes of 1.0 eV while high-resolution scans were taken with a step size of 0.1 eV. The energy resolution however is limited by the FWHM of the incident x-ray beam, which for lab-based spectrometers is about 0.5 eV. For survey scans, the step size of 1.0 eV will be quoted as uncertainty, while for high resolution scans, an uncertainty of 0.5 eV will be used. Fig. 5.2 shows survey scans (left panels) and high resolution scans for oxygen, tin and iodine (right panels) for vapour-grown BP and high pressure-grown BP. Both survey spectra are dominated by the phosphorus 2p and 2s peaks, at 130 eV ± 1 eV and 188 eV ± 1 eV, respectively, and their regularly spaced satellites[205], which correspond to the exciton.
binding energies. The spectra further show a small carbon peak, likely arising from the common use of carbon tape in XPS to stick down powder samples. This practice causes a permanent carbon background in the vacuum system. This carbon contamination also accounts for the small C KLL Auger peak visible at 1200 eV. No other peaks were found at survey scan resolution.

The high-resolution scans reveal the presence of small amounts of O, Sn and I. High resolution scans for metal impurities (Fe, Cu, Ti, Zn), observed in ref. [40], were also taken but no peaks were resolved. The presence of oxygen is likely due to the requirement to cleave our samples ex-situ for the XPS measurements as the vacuum system does not have a cleaving stage. This results in approximately one minute timeframe as the sample is placed into the vacuum system loadlock and pumped down for water to adsorb onto the surface and a small amount of oxidation to occur. For pressure-grown BP, the oxygen peak has a single component at 532.5 eV ± 0.5 eV, which is commonly associated with O-H bonds and which we attribute to physisorbed water[126]. The vapour-grown BP also has this component but further shows a second smaller peak at 530.2 eV ± 0.5 eV, which arises from forming POx complexes[126]. Vapour-grown BP has significantly smaller crystals and we suspect that the higher density of crystal edges and defective areas allows for more rapid surface oxidation than in the high-pressure samples[125].

The Sn 3d doublet peaks are identified in both types of samples at 487.0 eV ± 0.5 eV and 495.2 eV ± 0.5 eV, as are the iodine 3d doublet peaks, at 619.2 eV ± 0.5 eV and 930.1 eV ± 0.5 eV. Tin and iodine are known impurities in vapour-grown BP due to their use as catalysts in the synthesis [40]. The integrated peak area, adjusted with the respective sensitivity factors, suggest atomic concentrations on the order of 0.1% for both Sn and I. This however is also the detection limit of lab-based XPS, and it would be more reasonable to say qualitatively that there is a small amount of Sn and I present, rather than attempting to specify a concentration.

Another feature illustrating the limitations on quantitative analysis are the FWHM of the high-resolution peaks, which are not all the same. Specifically, they vary between two peaks of the same doublet, as seen in iodine of the HP-BP sample, which we would expect to show identical FWHM. This is an indication of the difficulty in fitting peaks to the relatively noisy data.

Interestingly, samples produced through both synthesis pathways show similar contaminations, suggesting these contaminants originate from the precursor rather than the synthesis method. Contrary to previous reports[97], we thus find Sn impurities in both vapour-grown and pressure-grown BP samples.
Figure 5.3: XPS survey spectrum of a black phosphorus sample glued onto a sample plate and left in ambient conditions. Right: high-resolution spectra of the P 2p and O 1s peaks

To compare to the cleaved surface, spectra were also taken on oxidised samples. Figure 5.3 shows an XPS survey spectrum of a bulk black phosphorus sample previously used for STM measurements, where it was cleaved perfectly clean, and then left in ambient conditions for a week.

The spectrum shows a C 1s and O 1s peak stronger than the P 2s and P 2p peaks, suggesting the first 10 nm of the surface have accumulated a significant amount of organic contaminants and oxide. We can further identify Ag and N peaks. These likely originate from a small patch of silver epoxy, which is used to glue crystals onto sample plates, within the scan area. Weak Sn and I peaks can also be seen.

The panels on the right show high-resolution spectra of the P 2p and O 1s peaks. There is a small bump to slightly higher energies of the P 2p peaks at 133.5 eV ± 0.5 eV, indicated by a black arrow, which is associated with oxidation states II to V of $\text{P}_x\text{O}_y$ complexes[126]. The main component of the O 1s peak, at 533.7 eV ± 0.5 eV, is attributed to P-O-P bonds in oxide complexes. The peak has a shoulder around 531 eV. The shoulder is likely a combination of a water peak at 532 eV and a P=O peak at 530 eV, but during fitting, two separate peaks were not a better fit to the shoulder than a single component. As mentioned previously, we found that cleaved samples primarily showed the P=O peak, suggesting oxygen dissociated onto the surface, but no
significant degradation had yet occurred. The exposed samples primarily show P-O-P peaks, indicative of more complex oxides and phosphoric acid, showing the degradation is more advanced.

Beside the direct photo-electric peaks, several Auger peaks can be seen at higher energies. Auger electrons arise from multi-step processes and tend to have lower kinetic energies when they are ejected from the sample. The binding energy plotted on an XPS spectrum is calculated by measuring the kinetic energy of the observed electrons, and subtracting that from the incident photon energy. Hence, Auger electron are plotted at high binding energy. This binding energy does not represent a specific electronic orbital, but rather the whole multi-step Auger process.

The overall background shape on the oxidised sample is very different from the cleaved pristine samples. The background in XPS primarily arises from inelastic processes the electrons undergo on their way to the sample surface. These processes cause the electrons to lose random amounts of energy and the spectrometer reads this as electrons having random binding energies. In Fig. 5.2, almost the entire sample is pure P, and the background is made up of P 2s and P 2p electrons that have lost energy. As such the background is strongest near the phosphorus peaks, and decreases from there. On the oxidised surface, there is a much bigger contribution from C and O electrons, resulting in a more even background. the background can still be seen to decrease from the O peak onwards. The background would reach zero when the incident photon energy is reached, as the incident energy would then not be enough to release electrons from the surface.

5.2.2 SIMS of pristine black phosphorus

We complemented the XPS measurements with SIMS to spatially resolve the distribution of impurities in each sample. Figure 5.4 shows SIMS measurements taken on a cleaved HP-BP sample using a O\textsubscript{2}\textsuperscript{2-} beam. The panel on the left shows the counts for the most prevalent species as a 5 micron crater is sputtered into the sample. The highest counts are unsurprisingly measured for P, P\textsubscript{2}, and P\textsubscript{3}, as well as PO species, which arise from the O beam mixing with the sample. The graph further shows a small amount of Sn, in green, and a brief spike of Na.

The panels on the right show how these counts are distributed over the scan area. As expected the P and PO distribution show no variations, showing the sample is evenly sputtered across the scan area. The Sn content is also evenly distributed, in contrast to the brief Na spike, which can be seen to be concentrated in the bottom left of the scan.
Figure 5.4: SIMS of cleaved BP. Left: Counts of most prevalent species with depth (sputtering rate approx 6 nm/s) Right: Spatial distributions of prevalent species area.

It is not immediately clear where the Na comes from, but because Na is very easily ionised, only a small amount would be needed to provide this signal. The localised nature of the impurity would suggest it is not intrinsic to the material, but arose either from somewhere else in the vacuum system.

The Sn signal shown only accounts for the Sn-120 isotope, which makes up 1/3 of naturally occurring Sn. Even considering this fact, the Sn counts are too small to effectively extract atomic concentrations from. Due to the dependence of SIMS sensitivity on the chemical environment of the dopant, this would also require a reference sample with known Sn content. Spectrometer counts were not adjusted for relative sensitivity factors.

What we can say, is that the even distribution of the Sn is consistent with the impurity being introduced during or before the growth process of BP, rather than an environmental contamination afterwards.
5.3 STM of samples as-received

5.3.1 Microscope and tip calibration on Si samples

Figure 5.5: left: Si(111) surface just after cleaning. Right: Si surface after one week in vacuum system. Both topographies were taken at -1.8V tip bias and current set-point of 75pA. Si(111) has well documented lattice constants and an atomic step height of 3.1 Å.[206, 207, 208]

The Si(111) surface, a staple in STM, was measured on a regular basis for two reasons. The vertical and lateral distances that the STM outputs as images are calculated from the voltages applied to the piezocrystal holding the scanning tip. The deformation of the piezocrystal per volt depends strongly on temperature.[142, 143] At cryogenic temperatures, a much higher voltage is required to produce the same displacement. While there are calibration factors in the microscope software, the displacement the microscope thinks it’s producing does not always match real lattice dimensions. The Si(111) surface is an extremely well studied and easy-to-measure surface and its lattice constants are precisely known.[206, 207, 208] By imaging the surface and using Fourier analysis to extract the lattice constants, it was found our STM set-up overestimates distances by 10% at liquid nitrogen temperatures, which was subsequently used to adjust measured lattice constants for black phosphorus.

The second reason for imaging Si(111) has to do with testing the quality of the scanning tip. The black phosphorus surface can be unpredictable, and it is often difficult to distinguish whether a failed approach of the surface is due to poor tip quality, or due to an uneven sample surface. Preparing STM tips is very time-intensive. As such, it
is essential that we are able to assess whether it is the tip or the sample that needs replacing when we are unable to measure. By scanning the Si surface, which can be sputter annealed repeatedly to pristine conditions[209], we can assess the quality of the tip before attempting to scan a phosphorus surface. Figure 5.5 shows the Si surface on two occasions. The panel on the left was obtained on a freshly cleaned Si surface. Step edges can be resolved and the dimer rows can be seen to change orientation by 90° on each atomic layer. A relatively low density of point defects, as well as the two dimer row configurations can be identified.

The panel on the right shows the Si surface after a week in vacuum. The highly reactive surface is starting to degrade despite the clean environment and the concentration of defects is much higher. In this scan, we can see the STM still resolves defects well, but struggles with bigger defects where the shear forces on the tip are strong enough to rearrange the apex atoms, resulting in bright scratches across defects. Furthermore, each protruding defect appears to have a smaller version of itself just below it, suggesting we might be scanning with a double tip.

5.3.2 Lattice and bandgap measurements

The buckled structure of black phosphorus appears in STM as parallel zigzagging rows of atoms, as shown in figure 5.6. These correspond to only the upper sub-lattice of
the surface layer, marked in red in panel (a). Lattice constants were determined through Fourier analysis, where fast Fourier transforms of the images are taken to produce reciprocal space distributions. Periodicities in the lattice appear as bright spots on these transforms. Lattice constants were found to be $3.5 \pm 0.1 \ \text{Å}$ and $4.2 \pm 0.1 \ \text{Å}$ in the zigzag and armchair directions respectively, in agreement with our density-functional theory calculations and previous studies [42, 43, 82]. The lattice unit cells consists of 4 atoms, two on each sub-plane of a BP layer, and is marked with a blue rectangle in figure 5.6(a) and a white rectangle in (b).

Figure 5.7: Bias sweeps taken at fixed points on vapour grown BP. The curves on the left were taken at three different locations with scanning parameters $\pm 2.0 \ \text{V}, 50 \ \text{pA}$. The measured bandgap is 1.1 eV. On the right, scanning parameters $\pm 0.4 \ \text{V}, 50 \ \text{pA}$ were used. Because at smaller biases, the tip is closer to the sample, a more accurate reading of the bandgap is achieved, and we measure 0.25 eV.

Early STM studies of black phosphorus reported bandgaps greater than 1 eV on bulk surfaces, leading to speculation that the top layer is electronically decoupled from the rest of the surface, and what was in fact being measured was a monolayer band structure.[80]

Figure 5.7 shows that this measurement can be reproduced if certain scanning parameters are chosen. The $\text{d}I/\text{d}V$ curves on the right shows the LDOS observed at three points on a clear area of the surface while scanning at $\pm 2.0 \ \text{V}$ and 50 pA, and measures a slightly p-type bandgap of 1.1 eV. The high tip voltage and low current setpoint result in the tip being relatively far away from the surface. While the bias is swept, the tip height is fixed, leading to a measurement of the LDOS several angstrom away from the surface.

In contrast, the red curve on the right was taken while scanning at $\pm 0.4 \ \text{V}$ and 50 pA, much closer to the surface. We then measure a strongly p-type bandgap ranging
from 0 V to 0.25 V. The valence and conduction band edges are more sharply defined and the bandgap is not overestimated.

For a perfectly accurate bandgap measurement, several measurements would need to be taken with the exact distance from the surface known, at which point the bandgap vs distance from surface plot could be extrapolated to obtain the bandgap at zero distance.

5.3.3 Non-double-lobed defects present in as-received samples

A small number of other point defects that do not give rise to extended features in the STM image were also observed. These defects, which we assign to divacancies or other small charge-neutral vacancy complexes, appear as localized dark spots or small reconstructions in the STM image.

Figure 5.8: The STM topography on a vapour-grown BP sample (a), simulated atomic potential (b) and calculated LDOS (c) of a Stone-Wales defect in monolayer BP.

One of these is a Stone-Wales defect, which involves no adding or removing of atoms, but only the rotation of a p-p bond. DFT calculations yield a formation energy of 1.8 eV, slightly higher than a phosphorus vacancy[82]. The left and centre image show a 4 nm STM topography and a simulated DOS isosurface respectively. The topography, taken at -0.4 V, shows a diagonal distortion across one atom row. The two adjacent atom rows are also affected. On each of them, one atom is pulled towards the defect, and the rest of the row relaxes down into the created space. The defect is less than 2 nm, across, and shows no further long-range Coulomb effects.

The simulated image shows a similar diagonal distortion, slightly protruding from the surface, which creates space on either side of it for the adjacent layers to relax into. Consequently, the adjacent rows appear slightly darker. Both adjacent rows show a slight enhancement on opposing sides of the defect.
The third panel shows the calculated DOS of the Stone-Wales defect, showing a band structure very similar to the pristine surface, and no in-gap states that could host charge. The calculations were performed on a monolayer of BP, and hence the calculated bandgap is much larger than bulk. The absence of a dielectric contribution from lower layers might also account for the small difference between the topography and the calculated image.

As there is no change in the number of atoms or bonds, the defect is neutral and does not produce states in the bandgap. The simulated DOS plot in Fig 5.8 shows the edge of the monolayer valence band and conduction band at 0.1 eV and 1.2 eV respectively, but no states in between.

5.4 Non-contact AFM of samples as-received

Similarly as in STM, AFM samples were cleaved from bulk crystals and glued onto stainless steel or molybdenum sample plates using conductive epoxy (Epotek H21D[202]). A ceramic cleaving post with diameter 2 mm is glued on top of the sample. After curing, the sample is introduced into the loadlock of the UHV system and outgassed at 150 °C overnight until pressures of $10^{-8}$ mbar are reached. The cleaving post is tall enough that it can be knocked against the top of the transfer valve, or when turned sideways, against the side of the loadlock, to cleave the sample at vacuum. Samples were immediately transferred to the microscope, where they were kept at $10^{-10}$ mbar and 8 K. In all AFM experiments, vapour-grown BP was used.

To minimise signal contributions from electrostatic interactions between tip and sample, a bias sweep is taken before scanning, as shown in Fig 5.9. The tip is held in position and the frequency shift experienced by the cantilever is measured as bias is swept. The frequency shift is lowest when the tip and sample Fermi levels are aligned and there is no potential difference between the two. In our experiments, this usually required a contact potential of less than 100 mV. Even at its minimum, the graph shows a frequency shift of just under 1 Hz, demonstrating we can’t completely eliminate electrostatic force contributions.

5.4.1 Lattice / step edges

The bare cleaved surface is shown in Fig. 5.10. The left panel shows a large area overview image of a series of step edges. the inset line profile shows the smallest step is 2.5 Å, with other steps multiples of that number.
Figure 5.9: Frequency shift of cantilever oscillation as bias is swept with AFM tip in close proximity of the sample. At higher biases, increased electrostatic interactions result in a bigger frequency shift. The red and white line show the backward and forward sweep across the bias range. The small difference between them arises from vertical relaxation of the tip.

The interlayer distance of BP is typically quoted as 5.2 Å [75, 210], suggesting our height measurements could be off by a factor of 2. What is measured in NC-AFM is not a cantilever deflection or vertical adjustments of a piezo stack, but a change in the oscillation frequency of the cantilever. Relating the frequency shift of the cantilever to absolute distance is difficult as it is strongly temperature dependent and varies with the shape of the tip [163]. To relate this meaningfully to height changes, we would need to measure the normalised frequency shift which requires several measurements with each new cantilever of slowly retracting the cantilever from contact and measuring how the forces decrease with distance, as described in the AFM theory section in eq. 30. As the normalised frequency shift is constant within the decay length of a force, having several frequency shift measurements on the same area yields the absolute oscillation amplitude, which is used for the z-calibration.

Due to limited time on the equipment, and the regular replacement of AFM tips due to crashes into the surface, we did not perform these measurements. Bright and dark areas on the NC-AFM images hence do not correspond directly to height, but only reflect stronger or weaker forces applied to the cantilever. Some of these come from height variations in the surface, but others from chemical interactions or undercoordinated bonds.

The line profile inset in Fig. 5.10 shows small protrusions at the edges of layers. Due to the incomplete bonding at step edges, these are common sites for adatoms.
or impurities and other defects to aggregate. The slope across the line profile on terraces that are expected to be flat is an artefact of an imperfect background subtraction in signal post-processing. The two common background subtractions, plane and line subtraction, don’t deal well with sudden changes in height, such as at step edges, as they attempt to fit a straight line across the whole scanning area.

The right panel of Fig. 5.10 shows an atomically resolved image of the BP surface. The rows of atoms in the zigzag direction are readily identified. Lattice constants were extracted through Fourier analysis: 3.6 Å in the zigzag direction and 5.2 Å in the armchair direction. Both values are 10-15% bigger than literature values, which again is due to the temperature calibration of the system. The movement per applied voltage of the piezo-actuators that move the cantilever in x and y varies strongly with temperature. Commonly, a reference sample of well known lattice constants is measured first to calibrate the voltage to distance conversion, but this was not prioritised in our study due to time pressures and difficulties in obtaining atomic resolution.

5.4.2 Observed defects

Fig. 5.11 shows a high-resolution image of one of the two most commonly observed defects in our AFM study. A deep depression suggests at least one missing atom. The four nearest neighbours to the depression are slightly brighter than the pristine surface.
It should be considered that a protrusion in an AFM topography does not necessarily imply a permanent feature. In particular around a defect, as is the case here, atoms have missing or longer bonds due to the structural reconfiguration. This makes them more weakly bound in the surface than other atoms. All atoms feel an attractive force from the scanning tip, but on a pristine surface, this interaction is not detected as all atoms are equally attracted to the tip. However, the more weakly bound atoms, when feeling the same attractive force from the scanning tip, are pulled further out of the lattice, and hence appear as protrusions.

The topography thus shows one or two missing atoms, surrounded by more weakly bound neighbours.

It was recently shown that single vacancies are not stable on the BP surface [82, 98]. They have low diffusions barriers and it is energetically favourable for them to combine into di-vacancies.

The right panel of Fig. 5.11 represents the defect topography in schematic form. Due to the height difference between the sub-planes, only the orange atoms are visible in AFM. The two atoms highlighted in red are the missing di-vacancy. The nearest neighbours affected by the altered bonds are highlighted in darker hues of their respective colours. As only the orange atoms are visible, the protruding atoms appear in a square around the defect.

Most other defects in AFM appear as small local protrusions (Fig. 5.12), often only
The limited extent of the lattice deformation and protrusion suggest this is most likely a single substitution. Other defects such as adatoms, vacancies, or interstitials result in much larger structural rearrangements.

The right panel in Fig. 5.12 shows a DFT calculated model of a Sn substitution in the BP lattice. The impurity slightly protrudes from the lattice, causing a small distortion of its two nearest neighbours, but otherwise leaving the lattice unaffected.

Both the hole defect and the small protrusions were detected in similar concentrations, although the overall number of defects observed was not sufficient to derive relevant statistics. Due to the sensitivity of the surface to tip-sample interaction, and the required fine control of AFM tips, atomic resolution as required to observe these defects could only be guaranteed on slow small scans, scanning no more than 25 x 25 nm at a time. As a result, the overall area observed in many times smaller than in STM, where scans of 200 x 200 nm will still yield atomic resolution. This, compounded with the fact that AFM is only sensitive to the very surface layer defects meant that generally, observing defects was rare. A more statistical analysis of the defect concentrations would require significantly more scanning time to cover a sufficient surface area.

The final type of defect observed on the BP surface is a long-range amorphous protrusion, shown in Fig. 5.13. The extent of the protrusion, and the absence of specific shape, suggest a subsurface electrostatic potential, perhaps a cluster of charged impurities. These clusters are not normally detected on the pristine surface unless there has been tip-sample interaction. However, the fact that the observed protrusion is smooth and relatively flat, suggests this potential arises from deeper within the material, ruling out tip-sample interactions.
5.4.3 Effect of frequency shift on resolution

The appearance of the pristine surface, and of defects, is further complicated by the chosen frequency shift setpoint on the NC-AFM tip oscillation. A frequency shift setpoint of 0 would imply the tip is oscillating at its natural frequency, without any interaction with the sample. An increasingly negative frequency shift suggests increasing tip-sample interactions[163].

The lowest setpoint we were able to scan at without damaging the tip or surface was -125 Hz, at which point the buckling of the zigzag rows is strongly emphasised as we get increasing contributions from local effects. As the setpoint is decreased, resolution of the zigzag row buckling is lost until around -100 Hz, both sides of each row look equivalent. As the setpoint is further decreased, local interactions contribute less to the overall signal, and atomic resolution is eventually lost around -60 Hz, below which the bulk of the signal comes from long-range electrostatic and Van der Waals interactions.

The setpoint variation suggests there is a relatively small parameter space within which the atomic resolution is optimal while creating the least distortion on the atom rows and without damaging tip and sample. We found the range between -100 and -120 Hz to be most representative of the surface as observed in STM and previous studies.
5.4.4 Atom row buckling

In both STM and AFM, the atom rows on the phosphorus surface can strongly vary in appearance depending on the shape of the tip, and on applied bias. Fig. 5.15 shows 3 AFM topographies taken with different tips and biases. The left image shows a buckling of the atom rows, with the left side of each zigzag row protruding compared to the right. This buckling has been seen in previous STM studies, and has been the topic of some investigations[42, 43]. Recently, it has been suggested that regularly spaced defects could produce such buckling[211], but it seems to us unlikely that defects would be spaced perfectly regularly throughout the material nor does this model justify why the buckling is only seen in some cases.

More recently, it was shown that BP is more sensitive to compression in the armchair direction than in the zigzag[212], which is consistent with the buckling direction.

A different study showed several stacking orders are stable in BP, and it may be that variations in stacking order produce sufficient local strain to cause the buckling[76].

After changing the contact potential from -30 mV to -80 mV, the middle image was obtained. The surface buckling is no longer apparent and instead, atoms appear in diagonal pairs along the rows. Adjusting the contact potential is intended to reduce
electrostatic forces by aligning the Fermi levels of the tip and sample. It may be that the change in applied bias caused a change in the tip shape, or in the tip-sample separation that no longer resolves the surface buckling.

The last image was obtained with a different tip, presumably blunter, which does not resolve variations within the atom rows, but instead shows an almost completely uniform orthorhombic matrix.

5.5 Degradation in ambient conditions: AFM and optical microscopy

Flakes were exfoliated with Fluortape onto gold patterned silicon in a cleanroom with transmission microscopes available. The samples were then transported to the AFM lab in a sealed jar with low oxygen environment (achieved by blowing N₂ into the jar before sealing).

5.5.1 Overnight AFM degradation

Fig. 5.16 shows AFM topographies taken on a freshly cleaved BP sample monitored for 24 hours. The amplitude error rather than the absolute height is shown, as this is most sensitive to changes in height, highlighting sharp features and step edges. Images were taken approximately every 35 minutes and the growth features on the surface was followed. The first panel shows the first image taken of the observed area, 30 min after cleaving. The area is already starting to show small features covering its surface. Small droplets can be seen to line up on crystal step edges. The step edge in the centre of the
Figure 5.16: AFM amplitude error images of a cleaved sample left overnight. The change in oscillation amplitude is rather than height to better show features. Feature heights are discussed in the text. Times since cleaving are indicated in the top left corners. The first overview image is 4x4 micron, while the rest are 2x2 micron.

scan is about 10 nm tall, while the small lines of droplets in the bottom left are tracing out a single atomic step edge. The area squared in blue is 2x2 micron in size and was followed for 24 hours.

2 Hours after cleaving, the area is thoroughly covered in small droplets, 3-4 nm tall. The four bigger droplets are 10-15 nm tall. The steep step edge in the centre of the image starts to accumulate a lot of uneven debris. Curiously, the next image sees a large area cleaned of droplets. It seems they have aggregated into one big droplet, about 30 nm tall. We confirmed that we are able to move droplets around with the AFM tip during a normal scan, but it seems surprising that the cleaned area would be round as opposed to the square scan area. From 3h40 onwards, after being almost completely cleared, the surface seems to repopulate with droplets. The droplets that form now are much more amorphous than the previous ones, likely a sign that the surface has become rougher or more hydrophilic. The area circled in green can be seen to start highly amorphous, and as it gathers more material, adopts a rounder droplet shape. After 8 hours, the green droplet is 30 nm tall and 200 nm wide. After about 18 hours, the surface is once again saturated in round droplets, with the biggest ones 45-50 nm tall. The size and distribution of these droplets suggests they are adsorbed water rather than phosphorus oxide. On bulk BP, oxidation results in an oxide layer of tens of nm[21, 32, 126], and while phosphoric acid is produced in the process, these droplets are much too big for the acid to account for them.
Studies have found that the oxidation process makes the BP surface more hydrophilic[122], which can explain the difference in droplet shape throughout the scans. Early on, while the surface is still relatively hydrophobic, droplets are mostly round, to minimise their contact area with the surface. Later on, they are more amorphous and only become round once their size makes that shape favourable to minimise their surface energy. The debris around the step edges is likely oxidation, as step edges are well known nucleation sites to oxygen and water chemisorption due to their irregular bond structure.

Note the shift in orientation between the 2h00 scan and the next is due to the scan direction. As the AFM rasters the surface up it produces one image, then on the way down, it produces the next, but there is a slight distortion between the two images.

5.5.2 Low humidity AFM

![Figure 5.17: Left: Humidity measured inside the AFM set-up under dry air flow. Right: AFM amplitude error images of cleaved samples, in ambient and dry air conditions, over 18 hours. The change in oscillation amplitude is shown to better show features. All images are 2x2 micron in size.](image)

We sought to separate the effect of oxygen and water on the degradation of the BP surface by repeating the overnight AFM measurements with a new sample under dry air flow. The AFM set-up was sealed as much as possible and dry air (80% pure N2, 20% pure oxygen) was blown into the set-up. The left panel of Fig. 5.17 shows the humidity measured inside the set-up with time at the start of the experiment. Humidity
decreases exponentially from ambient (30%) until it stabilises around 7% after 3 hours. The dry air was left flowing and humidity dropped further to 3% overnight. The sample was introduced when humidity reached 11%, resulting in a small increase seen on the graph at 110 min.

The right panels of Fig. 5.17 compare AFM images taken on samples exposed to ambient humidity and low (7%) humidity over the course of 18 hours. A small BP flake was been circled in green in the dry air data set to track drift. All images shown are 2x2 micron in size. After 1.5 hours, the dry air sample is only just starting to show a few small droplets along the step edges, while the ambient sample already has significant coverage of small droplets. Even after 18 hours, the low humidity surface is only sparsely populated with droplets, most of them under 10 nm tall. While we weren’t able to achieve a completely dry environment, we clearly see that the maintaining low humidity slows the appearance of droplets on the BP surface. The coverage as well as the size of the droplets is much smaller than in the ambient case. If the droplets were purely a result of the oxidation, and if the oxidation only required O\(_2\), we would expect the same coverage in both cases. Hence we see water clearly is involved in the appearance of these droplets. What we can not differentiate between at this stage, is whether these are simply adsorbed water droplets, or if they arise from oxidation and water is a critical component of the oxidation process. To resolve this distinction, we sought to remove the droplets, first through heating, then through washing the samples.

### 5.5.3 Heating samples and washing to remove oxide

The two left panels in Fig. 5.18 show optical microscope images of a BP flake before and after heating to 120 °C for 30 min. The droplets present on the surface shrink, leaving rings of deposited material showing their original size. The boiling point of phosphoric acid is 158 °C, so it is likely that these are in fact water droplets. After heating the flakes in an attempt to remove the droplets, they were gently washed with DI water and blown dry with N\(_2\). The middle two panels show a part of the flake before and after washing. The DI water does an excellent job of cleaning the surface, which optically appears almost pristine after washing.

The right two panels show AFM error amplitude images of a flake before and after washing. The flake is about 12 micron wide and 50 nm thick. Even at this scale, the gentle wash seems to have removed practically all the droplets, leaving the flake looking as new. Roughness measurements showed an RMS roughness on the flake of 8.7 nm before washing, and only 0.6 nm after washing. In comparison, the gold patterned SiO\(_2\)

85
substrate had an RMS roughness of 1.2 nm before washing and 1.1 nm after washing.
5.6 Induced defects - Si and H$_2$O dosing in STM

Defects were induced on the BP surface by exposing the cleaved surface to firstly water vapour, and later evaporated silicon. To dose the surface with water, a small flask of deionised water was mounted directly onto the microscope chamber of the STM vacuum system. A leak valve allowed for the controlled release of water vapour into the system. The surface was exposed to water four times, with increasing dosage. Only after the third dose, exposing the surface to 1.6x10$^{-7}$ mbar for 60 s, were defects detected in practical concentrations. A fourth dose, of 4.5x10$^{-7}$ mbar for 120 s, although 6 times more intense, only slightly more than doubled the density of defects found on the surface. It is worth noting that unless specifically deaerated, water will contain some amount of dissolved molecular oxygen. In fact, there have been several studies on the degradation of BP in normal vs deoxygenated water that show BP can be preserved in deoxygenated water, but degrades over time in normal water[122, 21]. As the water in our experiment was only deionised and not deaerated, we must consider the possibility the resulting defects are oxygen-related. Due to the difficulties in preparing workable samples in STM, a process which in the case of BP often takes weeks, the same sample was later used to induce Si-related defects. To dose the surface with silicon, a silicon evaporator, as described in the characterisation methods section, was used. The sample was exposed once to Si flux for 2 seconds in the preparation chamber of the vacuum set-up, before careful transfer to the microscope. For defect identification and counts, an in-house Java application was used: Feature Tracking App, v.1.0.0, developed by Adam Rahnejat, UCL.
Figure 5.19: 30 nm filled and empty state image of the vapour-grown BP surface after dosing with water. Double-lobed defects inherent to the surface are marked with purple ellipses, while three types of water-related defects are marked with circles. The brightest defects have an apparent height of 1.2 Å.

5.6.1 Water-related defects

Fig. 5.19 shows a 30x30 nm$^2$ filled and empty states STM topography of the BP surface after dosing with water. Three new types of defects can be identified. The double lobed defects inherent to the samples as received are marked with purple ellipses and are discussed in detail in chapter 6. Defects only observed after water dosing are marked with blue, green and yellow circles. The overall defect density was found to be 36 defects per 100x100 nm area after the small dose, although this was based on a single scan. After the bigger dose, a total of 820 defects were observed, at a density of 83 ± 7 defects per 100x100 nm area. Approximating the BP layer separation as 5 Å in the out of plane direction, this would translate to a density of 1.66x10$^{19}$ cm$^{-3}$.

The first type of defect appears as a small local depression on a single atom row, occupying the space of 1-3 atoms, as shown in the left panels of Fig. 5.20. The defect has the same appearance in both positive and negative bias. The defects further show an anisotropic Coulomb screening in negative bias, exclusively visible on one side of the defect. A Coulomb depression in negative bias typically indicates a positive charge, as this locally bends the valence band down, making fewer valence states available for tunnelling out of. However, using the same reasoning, this positive charge should create an enhancement in positive bias, where the bending down of the conduction
Figure 5.20: The three types of defects observed after dosing vapour-grown BP with water. The first appears as a local depression in both positive and negative bias, with a small amount of Coulomb screening favouring one side of the defect. The second appears as a strong protrusion in negative and a very slight protrusion in positive bias. The last appears as a small protrusion in negative, and a small depression in positive bias. All images are 5x5 nm. The ‘bright spot’ defect has an apparent height of 0.8 Å band makes more states available for tunnelling into. This enhancement is not resolved at the resolution we achieved.

The second type of defect appears as a small bright protrusion in negative scanning bias and a small dim protrusion in positive bias and is shown in the middle panels of Fig. 5.20. The protrusion in both biases may suggest a defect that physically sticks out of the surface, rather than a purely electronic effect. In negative bias, the defect seems to span across rows, although the rows themselves are not deformed, possibly suggesting an adsorbed defect bridging across atom rows.

The final type of defect, shown in the right panels of Fig. 5.20, appears as a small protrusion in negative bias and a small depression in positive bias. The defect is localised on a single atom row.

Table 1 summarises the statistics related to the observed defects. Holes are the most prevalent type of defect, accounting for 48% of cases. The bright and dim spots
appear approximately in equal numbers, making up 25% of cases each. Some other defects were observed, with only a few instances of each, which may have arisen from general degradation in the vacuum system, or from oxygen dissolved in the water.

5.6.2 Silicon-related defects

Silicon is 4-valent similarly to Sn. As we found Sn to be responsible for the double-lobed defects observed in BP, we explored whether Si adopts similar defect configuration and can produce similar charge states and defect states. We used the same BP sample for both the water dosing and the Si evaporation. Fig. 5.21 shows a filled and empty state topography of the sample. The double lobed defects are again marked as purple ellipses. We further distinguish 5 types of defects, one of which is quite rare and not present in this scan. The surface also contained bigger defective regions, 2-3 nm across. These are most likely groups of Si atoms that have clumped together on the surface. As they are too convoluted and varied in shape to tell us anything about exact defect configurations, they are not studied here.

High-resolution images of the defects are presented in Fig. 5.22. The most common three are labelled by their appearance in negative bias: two dots, three dots, and hole. Two more types were found in very small quantities: a variant on three dots, and a strongly charged defect. Two dots, marked with green circles in Fig. 5.21, are characterised in negative bias by two small protrusions on the same defect row. The two...
Figure 5.21: 50 nm filled (-0.4 V, 10 pA) and empty (+0.4 V, 10 pA) state image of the water contaminated vapour-grown BP surface after Si dosing. The double lobed defects are again labelled as purple ellipses while new defects are labelled in coloured circles. The brightest double-lobe has an apparent height of 1.0 Å.

Protrusions are 5.1 Å apart, about a third more than the zigzag lattice constant measured in the experiment, 3.8 Å. The defect further has a very small amount of Coulomb charging seen as a gentle depression extending about 5 atom rows on each side in the armchair direction. One of the two protrusions is unchanged in positive bias. The other is replaced by a narrow hole, extending across three atom rows in the armchair but only 2 atoms wide in the zigzag direction.

Three dots, marked with yellow circles, show in negative bias three bright protrusions flaring out from a central point like three petals of a four-leaf clover. The central point sits between atom rows so that the three protrusions straddle two atom rows. In positive bias, two of the protrusions remain straddled across atom rows, albeit slightly smeared out. The third now forms a depression covering one row plus the gap in between rows. One atom is still visible inside the depression.

The last common type of defect, labelled ‘hole’, is marked with blue circles and looks similar to ‘two dots’ in positive bias: a small protrusion next to a larger depression. In the case of ‘hole’, the depression is more square and centered on the atom row neighbouring the protrusion. The two types are more easily distinguished in negative bias, where ‘hole’ again shows a very similar structure, a protrusion next to a depression. The depression is less pronounced, and a small amount of Coulomb charging can be seen extending in the armchair direction.

Besides these three common types of defects, accounting for approximately 30% of
observed defects each, two other types have been found, accounting for 4% each. The first, labelled 'charge', is marked with a light blue circle and has two very bright localised protrusions close together on one atom row. In negative bias, the neighbouring atom row is distorted but does not show a strong protrusion or depression. In positive bias, this defect shows a very strong circular charging, distorted into an ellipse due to the anisotropy of the lattice. The actual defect can still be made out as a dark centre to the bright feature. The second rare defect, labelled 'three dots variant', does not appear in the overview scan, but is shown in the high-resolution images. It again presents three bright protrusions in negative bias, but rather than covering three corners of a rectangle, they are evenly spaced. In positive bias, this defect shows a fairly large depression covering two atom rows. One half of the depression appears distorted, while the other shows a clear atom at its centre, of similar brightness to the rest of the BP lattice.

Table 2 summarises the statistical analysis performed on 335 defects from different days taken with different STM tips. 'Two dots', 'three dots' and 'hole' account for 28%, 37%, and 27% of observed defects respectively. 'Charge' and 'three dots variant' were each observed in 4% of cases. The overall defect density was found to be $27 \pm 8$ defects per $100 \times 100 \text{ nm}^2$. 

Figure 5.22: High-resolution images of the types of defects observed on vapour-grown BP after silicon dosing. Top row: -0.4 V, 20 pA; Bottom row: +0.4 V, 20 pA. All images are 8 nm across. The left most images were cut off at the edge of a scan area but kept to the same scale for comparison.
5.6.3 Discussion

Two observations stand out immediately when we consider the same sample was used for water dosing initially, and then silicon dosing on top. Firstly, the defect concentration after the final water dose was $83 \pm 7$ defects per $100x100$ nm$^2$, while after silicon dosing, defects were only found at a concentration of $27 \pm 8$ defects per $100x100$ nm$^2$. Even if silicon only interacted with water defects and not the pristine surface, we would expect the defect concentration to at least remain constant. Secondly, none of the water defects are observed after silicon dosing. One possible explanation is that the water defects disappear over time, perhaps diffusing to step edges. However, on the limited timescale of this experiment, we did not observe a drop in defect concentration. If the diffusion were to occur at a longer timescale, this would not explain the sudden drop in concentration between water and silicon dosing. Instead, it seems that the silicon dosing process restored the BP surface, either through heating the sample from the silicon beam, or from the interaction of silicon with the water defects.

There have been no theoretical studies of Si on BP, but considering the valence of Si is the same as that of Sn, which is discussed in detail in the chapter relating to double-lobed defects, we may expect Si to also form substitutional defects hosting a dangling bond. However, this would result in a point defect with several possible charge

<table>
<thead>
<tr>
<th>Scan size (nm x nm)</th>
<th>2 dots (%)</th>
<th>3dots (%)</th>
<th>3dotted (%)</th>
<th>hole (%)</th>
<th>defects per 100x100 nm$^2$</th>
<th>total defects</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>50x50</td>
<td>0</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>100x100</td>
<td>22</td>
<td>44</td>
<td>6</td>
<td>11</td>
<td>17</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>150x150</td>
<td>19</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>100x100</td>
<td>28</td>
<td>44</td>
<td>2</td>
<td>0</td>
<td>26</td>
<td>22</td>
<td>50</td>
</tr>
<tr>
<td>50x50</td>
<td>38</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>50x50</td>
<td>30</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>50x50</td>
<td>50</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>48</td>
<td>12</td>
</tr>
<tr>
<td>75x75</td>
<td>25</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>75x75</td>
<td>57</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>29</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>75x75</td>
<td>29</td>
<td>24</td>
<td>6</td>
<td>0</td>
<td>41</td>
<td>30</td>
<td>17</td>
</tr>
<tr>
<td>75x75</td>
<td>21</td>
<td>37</td>
<td>5</td>
<td>16</td>
<td>21</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>150x150</td>
<td>21</td>
<td>37</td>
<td>5</td>
<td>8</td>
<td>29</td>
<td>28</td>
<td>62</td>
</tr>
<tr>
<td>75x75</td>
<td>18</td>
<td>35</td>
<td>6</td>
<td>0</td>
<td>41</td>
<td>30</td>
<td>17</td>
</tr>
<tr>
<td>75x75</td>
<td>24</td>
<td>41</td>
<td>0</td>
<td>12</td>
<td>24</td>
<td>30</td>
<td>17</td>
</tr>
<tr>
<td>75x75</td>
<td>17</td>
<td>25</td>
<td>8</td>
<td>8</td>
<td>42</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>75x75</td>
<td>37</td>
<td>37</td>
<td>5</td>
<td>0</td>
<td>21</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>60x75</td>
<td>50</td>
<td>17</td>
<td>17</td>
<td>0</td>
<td>17</td>
<td>13</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>average %</th>
<th>28</th>
<th>37</th>
<th>4</th>
<th>4</th>
<th>27</th>
<th>total defects</th>
<th>Defect density 27 ± 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>std</td>
<td>14</td>
<td>13</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>335</td>
<td>27 ± 8</td>
</tr>
</tbody>
</table>

93
states, which we do not observe in experiments. While it is possible that only the neutral state is stable on the BP surface, this would require theoretical calculations and band bending experiments to confirm. Studies of the oxidation of BP have shown the process to occur from nucleation sites and form craters[125, 32]. In our topographies however, we only see individual defects. It is possible that the total dose, or the amount of dissolved oxygen in the water, was sufficiently low to limit the oxidation process to only the first steps (oxygen dissociation and water attaching to the oxide) but we see no further evidence for the complete oxidation cycle[213]. The energetically most favourable configuration of adsorbed water, where water molecules form six-membered rings, would create large defects spanning across two atom rows, which we do not see[123]

One key difference between the water dosing experiment and the silicon dosing is the scanning bias used, and the overall resolution of the images. Typically, for the water-related scans, biases of ±0.7 to ±1.6 V were used, while Si-related scans were usually taken at ±0.4 V. It is well known that defects can appear significantly differently depending on bias. Double-lobed defects for example, while distinctly double lobed at ±0.4 V, are only barely visible as smeared out protrusions at ±1.0 V. Consequently, it is theoretically possible that some of the defects identified in Si-related scans are simply water-related defects, that appear differently at low scanning bias, although they were found consistent in appearance within the ±0.7-1.6 V range.
Figure 5.23: STM topographies of vapour-grown BP after Ar sputtering. a. shows a concentration of sputtering defects similar to the double lobed defects (+0.5 V, 10 pA). b and d. show close-ups of sputtering defects where dislodged atoms are clearly visible(b.) and where the top layer is intact(d.)(-0.5 V, 10 pA). The brightest defects in panel a. have an apparent height of 1.2 Å.

5.7 Sputtering and annealing

A popular hypothesis in the literature on atomic scale defects in black phosphorus is to attribute the double-lobed defects to single black phosphorus vacancies[81, 42]. To explore this hypothesis, we sought to induce vacancies through argon sputtering using a 99.999% argon source. Samples were sputtered at room temperature under an accelerating voltage of 100 eV and sputtering current 0.1 mA for 1 second, under 2x10⁻⁶ mbar Ar.

Panel a. in Fig. 5.23 shows that the sputtering process produces a small number of localised defect, with a concentration similar to that of the double lobed defects. The created defects appear to have a small number of atoms dislodged or removed, result-
ing in a small depression. In some cases, the missing atoms are easily recognised, such as in the case in panel b. where at least two atoms are missing in the central row. In other cases however, all atoms appear to be in place, as shown in panel d. The atom rows are intact and nothing appears to be dislodged on the surface. In cases where the top layer appears intact, the depression is likely due to missing atoms in the bottom half of the BP layer, which can’t be probed directly. This also creates issues in terms of definitively labelling a defect as a single- or N-vacancy. While the overall concentration of defects roughly doubled, the concentration of double-lobed defects remained constant throughout the sputtering experiment, suggesting our sputtering did not produce any further double lobes. Although we are unable to definitively identify our created defects as single vacancies, this is encouraging and suggests the double lobed defects may not be vacancy related. We note that the negatively charged single vacancy defect is expected to be sufficiently mobile on the surface to diffuse and coalesce into di-vacancies during the sample transfer from prep chamber to microscope[82, 98]. Hence we would not expect to see any negatively charged vacancies even if we created them. The neutral and positive configuration of the single vacancies however is stable at room temperature and should result in an increased concentration of double lobed defects if we had created any[81, 82].

Sample annealing is a common way of removing surface defects, either through desorbing them from the surface, or from allowing them to incorporate deeper into the material. On black phosphorus, annealing studies have been performed to study its degradation[214, 215, 216]. Samples heated beyond 300 °C started showing craters aligned with the zigzag direction where phosphorus had been removed[217]. Craters grew with increased temperatures and annealing times. In the hope of removing some of the defects we have studies on the surface, we gently heated a cleaved sample to 150 °C overnight, and imaged the surface.

STM tips would repeatedly fail to approach the sample, not detecting a current and crashing into it. When imaging did succeed, we observed very large areas of clustered contamination. The left panel of Fig. 5.24 resolves the phosphorus atom rows in the top left and centre of the scan, confirming that we are indeed imaging the surface. It does however also show large areas coated in an amorphous material, less than a nm tall, but spreading across vast areas of the sample. While scanning over the amorphous area, the STM tip is usually lost within a single scan, as shown in the top of the right panel.

There are not many candidates for what this amorphous material can be, especially when considering other groups have successfully annealed black phosphorus at much
higher temperatures and still see an atomically clean surface. I suspect this contamination comes from the silver epoxy we use to glue down our samples. We specifically cover even the sides of the sample in epoxy to maximise the conductive pathway.

The annealing temperature is slightly higher than the temperature the epoxy is cured at (150 °C vs 120 °C) [202]. It is possible this temperature is high enough for the epoxy to diffuse onto the surface. Annealing experiments performed by other groups used thin flakes of BP fixed on graphene without any epoxy, so would not have had this problem [217, 215]. The epoxy made scanning sufficiently unstable that no reliable measurements of defect densities could be obtained and the sample was replaced for a fresh one.

5.8 Chapter Discussion

Small amounts of tin and iodine were consistently found in our samples. The high-pressure grown and vapour-grown samples contained these impurities in similar concentrations despite their different synthesis methods. The vapour growth technique requires SnI₄ mineralising agents, and the resulting samples have been found to contain these impurities on the order of 0.1-0.3 at% in our own as well as previous studies [40, 21]. High pressure grown samples do not require this catalyst. Nevertheless, tin and iodine were observed in our chemical analysis. Correspondence with the manufacturing confirmed their in-house chemical analysis revealed similar impurity. 2DSemiconductors Co confirmed they found small amounts of iodine in their high-pressure black phosphorus despite not using iodine in any form in their facilities. They were not able
to confirm the presence of tin, but the analysis they performed through SIMS is known to have poor sensitivity to materials that are hard to ionise[218]. Their observation of iodine despite having iodine-free facilities suggests this impurity may arise from the red phosphorus precursor used in both synthesis methods. The high pressure method is known to readily incorporate pre-existing impurities in new lattices, and in fact has been used to intentionally dope black phosphorus samples with both positive and negative ions [107].

While there is no obvious origin for tin or iodine impurities in red phosphorus, it is a reactive compound[219]. Left in ambient conditions, it is observed to take on a darker hue over time as it slowly oxidises[220, 221]. The conversion process of white to red phosphorus is one of anaerobic heating and does not remove the commonly observed metal and alkali contaminants in white phosphorus. Ever since the first detailed studies of the electronic structure of black phosphorus, the material has been found to be p-type[30, 21, 42], suggesting the ubiquitous presence of negatively charged impurities. A close examination of these charged impurities, as reported in the next chapter, supports future efforts towards producing impurity-free black phosphorus that can more reliably be used in 2D heterostructure applications.

Both our STM and NC-AFM studies have observed a buckling of the atom rows in black phosphorus, resulting in one side of each zigzag row to slightly protrude from the surface. This has also been observed in other STM studies[42, 43]. One study from 2017[211] confirmed the presence of this surface buckling through LEED, but suggested it arose from the regular spacing of vacancy defects at the surface. Beyond the fact that there have been several studies showing vacancies to be unstable in black phosphorus, it is also not clear how these defects would arrange themselves in a sufficiently regularly spaced pattern to ensure even buckling. It seems more likely that this symmetry-breaking within the atom rows is due to surface relaxation[41, 42] or due to the AB stacking order[76] of black phosphorus.

5.9 Chapter Summary

In this chapter, I have presented a chemical analysis of the purity of black phosphorus samples produced through both common synthesis pathways. I have presented scanning probe microscopy in the form of STM and NC-AFM of the sample as-received, as well as ambient AFM of the degradation of these samples over time in ambient conditions. Finally, I described a range of defects induced in STM, in comparison to the double-lobed defects ubiquitous in STM studies of black phosphorus.
The chemical analysis, through XPS and SIMS showed samples from both synthesis methods contain small amount of tin and iodine impurities. The concentration of these dopants is on the order of 0.1%, the detection limit for lab-based XPS. SIMS showed the Sn impurities are evenly distributed throughout the material, suggested they were introduced during the crystal growth as opposed to as part of a degradation process of exposure to a certain environment after synthesis.

Scanning tunnelling microscopy of samples as-received identifies the lattice constants as 3.5 Å and 4.2 Å in the zigzag and armchair directions respectively, in agreement with our density-functional theory calculations and previous studies [42, 43, 82]. A bandgap of 0.25 eV is measured on the pristine surface. We have also observed that by choosing scanning parameters that place the tip further away from the sample, this bandgap can appear to be bigger, and even incorrectly be related to the monolayer bandgap of 1.1-1.5 eV. The samples as-received contained predominantly one type of defect, a double lobed protrusion, which will be discussed in detail in the next chapter. We further identified a small concentration of charge-neutral point defect appearing as localised dark spots. Some of these can be identified as Stone-Wales defects, while the remainder are likely different configurations of di-vacancies, with two missing phosphorus atoms allowing for various charge-neutral reconstructions.

Non-contact AFM was used for the first time on black phosphorus to obtain surface topographs with limited electronic contributions. Surface properties were found to be similar as observed in STM and two main types of defects were observed. The first type, small holes, likely correspond to the previously mentioned charge-neutral point defects and di-vacancies. The second is a small, single atom protrusion, which is consistent with DFT simulations of a substitutional Sn defect. The double lobed charge density surrounding these defects in STM was not identified in AFM as the technique is sensitive to all force interactions between tip and sample rather than to electronic interaction only. In both STM and AFM, a buckling of the zigzag rows was identified, recently described by [42] as arising from the AB stacking order in black phosphorus.

Ambient AFM and optical microscopy was used to follow the degradation of bulk samples in ambient conditions and in a low-humidity environment. Samples were found to develop water droplets at their surface, which grow over time. Droplets growing on a surface that is freshly cleaved minimise their contact area with the sample, attesting to the slightly hydrophobic nature of pristine black phosphorus. They are readily washed off the surface with deionised water. Droplets growing on an oxidised, but washed sample grow in amorphous shapes, suggesting the surface has become more hydrophilic, consistent with the current understanding of the role of oxygen dangling bonds in the
Finally, in vacuum, we dosed samples with water to study its role in the degradation at an atomic scale. Samples were also dosed with silicon, which shares the same valence as tin, in an attempt to replicate the double lobed defects. Water generally shows little interaction with the surface. Three types of defects were observed as a result of the dosing, at fairly low concentrations. Each type is charge-neutral and produces a localised reconstruction. After dosing with silicon, the water-related defects were no longer observed, suggesting the silicon may have combined with the pre-existing defects to create new configurations. Again, a negatively charged defect state was not the lowest energy state for the produced defects.

In the next chapter, we discuss the most commonly observed defects in black phosphorus in atomic studies. These appear as extended areas of enhanced charge density, in a double-lobed or p-orbital like shape. Through DFT and tight-binding calculations, we simulate the appearance and local density of states of tin substitutional defects and find these to be consistent with our observations.
6 Substitutional Sn defects in black phosphorus

The negatively-charged defects consistently observed in STM studies of black phosphorus were up till recently attributed to single phosphorus vacancies evenly distributed throughout the lattice\[42, 43\]. However, several studies, as described in the literature chapter, have since found monovacancies to be highly mobile in black phosphorus at room temperature, and even at liquid nitrogen temperature still have a hopping rate, the frequency at which the defect is able to move by one lattice position in the crystal, of several thousand hops per second\[82, 98\]. This would make them impossible to resolve in STM studies, as over the course of a scan (at least 10 min), a defect would have moved a significant distance, and appear as a streaky blur if resolved at all. What's more is that at the calculated hopping rate and the observed concentration, two monovacancies would meet each other within 3 seconds at room temperature. Upon meeting, it is energetically favourable for two vacancies to form a di-vacancy, which is stable and electronically neutral\[82\].

These studies have led us to reconsider the origin of the double lobed defects. To gain a better understanding of the atomic and electronic structure of these double lobed defects, we have combined scanning tunnelling microscopy and spectroscopy with theoretical modelling through density-functional theory and tight-binding calculations. Our calculations and observations support our theory that substitutional Sn impurities form charged defects and give rise to bound hydrogen-like states in the BP bandgap, including the observed double lobes.

6.1 Double-lobes

Black phosphorus samples were prepared as described in chapter 4 and cleaved in UHV to obtain surfaces without oxidation. Fig. 6.1 shows a 80 x 50 nm² empty state STM image of the BP surface with nine double-lobed defects. The right panel shows a higher resolution image of two defects superimposed on the rows of P atoms comprising the top half of the BP surface layer. The double-lobed defects are elongated in the armchair direction of the lattice along which conductivity is enhanced. Comparing the two defects carefully we see that the defect in the bottom left corner of the image has a bright structure at its centre, which is absent in the defect in the top right corner. This difference most likely arises from the defects sitting at different depths, the bottom defect sitting in the surface layer while the top defect resides deeper into the crystal.

Defects that do not sit in the top layer have a node at their centre, where the lat-
Figure 6.1: Left: Large area image showing intrinsic defects that appear as long range elongated protrusions on the vapour-grown BP surface. Right: Close up of a pair of defects. Both topographies were taken with a current setpoint of 20 pA. The bottom defect in the right panel has an apparent height of 1.0 Å.

The double lobed defects were observed in samples made through both the vapour growth and pressure growth synthesis methods. In each case, defect concentration was determined by counting defects over a large number of scans and then extrapolated to cm\(^{-3}\) using the known lattice constants of black phosphorus. Based on the work of Kiraly[42] and our own findings, the assumption was made that defects up to 5 layers deep can be observed at the surface. The resulting defect concentration were 80 ppm for vapour-grown samples and 60 ppm for pressure-grown samples. As these numbers were obtained by aggregating a large number of scans, no standard deviation information is available. Generally, defects were observed in both types of samples with comparable densities. The densities of the observed defects are an order of magnitude below what may be expected from XPS measurements, which detected Sn and I concentrations on the order of 0.1 at%. The XPS however also showed that there are strong variations in the dopant concentrations throughout the sample, and it is likely that the bulk of dopants aggregate at macroscopic defects such as crystal edges and dislocations, which are too rough for STM imaging.

The two observed defect central structures, as described in section 6.1.4, were observed in both types of samples, but we were not able to quantitatively describe their presence. The central structure can only be observed for defects in the top surface layer and only under very stable imaging conditions and a sharp tip apex.

This shows that in both samples, negatively charged point defects are present in
comparable densities, and, at least qualitatively, the same defects are observed in both samples. This leaves the question of what contaminants both synthesis processes may have in common. Our chemical analysis highlighted the presence of Sn in both types of samples. The use of Sn catalysts in the vapour growth method make Sn an obvious suspect for contaminations. While the high pressure synthesis does not specifically require catalysts, it has been shown that the method readily lends itself to the incorporation of substitutional dopants. Recently, pressure grown black phosphorus was successfully doped to up to 5 at% during synthesis with Se, S, Ge and Te\(^{107}\). While doping with Sn specifically has not yet been reported, it is not uncommon for layered materials to contain metal impurities. They have been observed in graphite\(^{116, 117}\), where they dramatically affect its catalytic\(^{117}\) and electronic\(^{118}\) qualities. Similarly, MoS\(_2\) has shown improved catalytic properties after purification processes\(^{119}\) or intentional doping with Ni, Fe, or Co\(^{120}\). Beyond the synthesis processes themselves, we can look for contaminants in the precursor material, red phosphorus. Red phosphorus is highly reactive, with commercially quoted purities ranging from 97\%\(^{222}\) to 99.99\%\(^{223}\). Its most common contaminants are Fe, Al, and Si. The exact origin of Sn dopants in high-pressure black phosphorus remains an open question. Experiments of intentional doping with controlled amounts of Sn would likely better inform the role and characteristics of Sn defects in the material.

6.1.1 left- and right-handed

The two lobes of the defects are not equally bright with some defects exhibiting brighter upper lobes, others having brighter lower lobes, as illustrated in Fig. 6.2. Consequently, we may assign the defects as left- or right-handed.

Figure 6.2: A line profile taken along the long axis of a double lobed defects shows one lobe protrudes more than the other.

In an unperturbed phosphorus lattice, we would expect left- and right-handed defect to be equivalent as the lattice is symmetric through a 180° rotation. However, distinc-
tions between the two have been described. Kiraly et al.[42] observed a preference for defects to one orientation over the other, with left-handed defects (on the protruding side of the atom row) to be twice as common as right-handed defects. It was suggested this imbalance arises from the location of the defect, whether it sits on one side or the other of a zigzag row. As discussed in the previous chapter, previous STM studies[41, 97], as well as our own STM and NC-AFM study, identified a buckling of the atom rows, leading to side of each row slightly protruding compared to the other. This symmetry-breaking would likely account for one side of the double lobe to be brighter than the other. The fact that defects are always distinguishable as left- or right-handed, even when the scanning resolution is not good enough to resolve the atom rows corroborates the idea that the buckling is always there, but simply is not resolved by a blunt or anisotropic tip.

The fact that defects in the protruded side of an atom row are more common implies the buckling of the atom rows makes that side of the zigzag rows energetically preferable over the other. Kiraly's team suggested surface relaxation led to one site being more stable, but did not present calculation to this end[42]. In our studies, we did not see a preference for defects to be of one orientation but rather observe similar amounts of left- and right-handed defects. An interpretation based on surface relaxation would also require the defects to jump positions when a crystal is cleaved. Before cleaving, the bulk crystal would contain equal numbers of left and right-handed defects. After cleaving and exposing a surface, a portion of right-handed defects would need to move lattice sites to become left-handed to explain the observed ratio.

6.1.2 Depth of defects

In the overview image, we can clearly see that some defects appear brighter than others. Kiraly et al.[42] showed that the apparent height of these defects is not a continuous range, but takes discrete values. It was deduced that defects could be observed up to 5 layers deep into the surface, with the apparent height of deeper defects reduced. While the extended electronic signature is detected for all defects, the central structure is only visible for those residing in the very surface layer. It should be noted that the defects used in Kiraly’s analysis were not all imaged with the same scanning parameters, even though especially the current setpoint is well known to affect the intensity of observed features.

Fig. 6.3 shows the apparent height of 70 defects observed with constant scanning parameters in our experiments. While our study does not resolve plateaus as clearly
Figure 6.3: The averaged apparent height of 70 defects, sorted in ascending order, shows a mostly continuous trend, with a discontinuity for very tall defects.

as is the case inKiraly’s study, we do see a clear distinction between very tall surface layer defects and dimmer subsurface defects.

The difference between the two studies likely arises from how the defect height is defined. In our experiment, we took the height to be the average height of the two lobes. In Kiraly’s study, height measurements were taken on the dominant lobe of left-handed and right-handed defects separately. While left-handed defects showed plateaus in their apparent height, right-handed defects just appeared as a continuous linear increase. As our study is a mix of the two, it is likely the plateaus are washed out.

It is not clear why left-handed defects would show clear plateaus while right-handed defects don’t. This distinction is again likely linked with the broken symmetry between the two zigzag lattice sites, but even if one protrudes from the other, we would expect to resolve plateaus for both positions as the discrete steps in height are related to the large gap between black phosphorus sheets, which affects both types of defect.

6.1.3 Size variations with scanning bias

Beyond variations in apparent intensity depending on the depth of defects, the shape and contrast of the defect is affected by the applied sample bias. Fig. 6.4 shows four topographies of the same defect, at increasingly small negative bias. The defect appears to stretch in the zigzag direction at lower biases. Where the brightest part of
the defect only spans 4 nm at -0.3 V, it stretches to 6 nm at -0.15 V.

Figure 6.4: Filled state bias dependency of a double lobed defect on vapour-grown BP. As bias is reduced, tunnelling current is enhanced along the long axis of the defect. All topographies were taken with a current setpoint of 20 pA. The defect shown has an apparent height of 1.0 Å.

This increase in size is partially due to the fact that at -0.15 V, we are very close to the valence band edge, and the density of valence band states starts to decrease. This means there are fewer pristine phosphorus states contributing to the signal, enhancing the contrast of the defect in comparison. Measurements closer to 0 V could not be taken as the tunnelling current on the pristine surface would disappear entirely, causing the tip to crash into the surface.

More interestingly, this enhancement is consistent with the defect state sitting in the bandgap close to the valence band edge. Due to the LDOS distribution of the tip, tip-sample interactions are particularly sensitive to energies similar to the tip Fermi level[132]. As the tip bias approaches the in-gap defect state, which is identified at +0.02 eV, we are increasingly sensitive to it, and it is enhanced in the topographies.

6.1.4 Defect centres

For double lobes sitting in the surface layer, a central atomic rearrangement could be identified. Throughout our experiments, we observed two types of central structures, presented in Fig. 6.5.

The first type appears as three bright spots in an inverted L shape in negative bias, straddling two atom rows. While two of the bright spots are on P atom sites, the last one appears to distort the zigzag row. In positive bias, only two protrusions are seen, apparently on normal P atom sites, with enhanced contrast.

The second type looks similar in positive and negative bias: two large protrusions are partially coalesced and almost appear to sit on top of the lattice, straddling two atom rows. The measured height of the defect is however not enough to account for
two adatoms, making this a LDOS effect rather than physical structure. The phosphorus lattice around the two protrusions appears more distorted than in the first defect type.

The four topographies shown all show another feature of the double lobed defects. For each zigzag row, the side that faces the defect is enhanced in tunnelling contrast, suggesting a higher density of states there.

The different defect centres likely suggest there are several origins giving rise to the double lobed defects. Indeed, our collaborators at Imperial College London found using tight-binding calculations that a series of hydrogen-like states can arise from a negative point charge, without defining what type of defect structure produces the charge. As such, it is possible that different defect structures produce similar hydrogen-like states as long as they are negatively charged.

As our chemical analysis has shown Sn and I as consistently present impurities in samples made through both synthesis routes, Julian Gaberle and Alex Shluger from UCL Physics have performed theory calculations to evaluate whether Sn or I defects could produce these double lobed defects.
6.2 DFT calculations of I- and Sn-related defects

DFT calculations were performed by Julian Gaberle from Prof. Alex Shluger’s group in the Department of Physics at UCL to calculate the electronic structure of iodine- and tin-related defects in monolayer and multilayer BP.

All calculations were performed using the CP2K code,[224] which employs a mixed Gaussian and plane wave basis-set (GPW). The DZ_MOLOPT_GTH basis set was used together with Goedecker-Teter-Hutter (GTH) pseudopotentials.[225] The plane wave cut-off was converged at 400 Ry, SCF convergence was set to $10^{-6}$ a.u. and residual forces on atoms were smaller than 0.01 eVÅ$^{-1}$. Since GGA functionals predict a metallic behavior for BP, the PBE0-TC-LRC hybrid functional was used with a cut-off radius of 2 Å and 10% HF exchange.[226] In order to reduce the computational cost of the hybrid functional calculations, the auxiliary density matrix method (ADMM) was used, which uses a reduced basis set for the HF exchange calculation and thus allowing for cells up to 1500 atoms large.[227] The optimized unit cell vectors were found to be $a = 3.3679$ Å, $b = 4.5484$ Å, and $c = 10.8699$ Å which is in good agreement with experiment and other DFT studies.[82, 98, 80, 228].

The bulk properties were calculated using a 1296 atom cell (9x9x4 supercell) and the monolayer consisted of 324 atoms (9x9) up to 1296 atoms (18x18) cells with 20 Å vacuum gap between periodically repeated images. In order to estimate the charge state of defects, the ionization potential of the slab was calculated with respect to the vacuum level. For a monolayer the VBM is at -5.5 eV, but increases to -5.2 eV for a bilayer and -5.1 eV for a 4-layer slab.

Adsorbed and substitutional iodine were calculated on monolayer black phosphorus, using a dielectric constant to account for interlayer interactions, while intercalated iodine was calculated in a 4-layer slab. The hydrogen-like states arising from a negatively charged Sn substitution were calculated on a monolayer of black phosphorus, and corroborated by tight-binding calculations performed by Martik Aghajanian and Johannes Lischner at Imperial College London.

A 4-layer slab has a slightly higher bandgap than bulk black phosphorus. The inverse decrease of bandgap with layer number however results in diminishing returns in accuracy when adding more layers, while significantly adding to computational complexity. Four layers thickness is commonly considered a good representation of bulk behaviour and allows for direct comparison with previous studies using similar parameters.[229, 98] The bandgap for monolayer black phosphorus was calculated to be 1.35 eV and decreased to 0.5 eV for bulk (4-layer) black phosphorus, close to the
bulk value observed in experiments, 0.35 eV. The difference in energy per P atom between bulk and monolayer black phosphorus is less than 0.05 eV, showing that calculated defect formation energies have little dependence on choice of black phosphorus reference.

6.2.1 I-related defects

![Image of DFT calculations of a substitutional iodine defect in a black phosphorus monolayer. Left: the reconstructed black phosphorus surface. The iodine atom forms a single bond to a phosphorus vacancy to create a neutral defect. Right: The band structure of substitutional iodine. No states are observed in the bandgap.](image)

Figure 6.6: DFT calculations of a substitutional iodine defect in a black phosphorus monolayer. Left: the reconstructed black phosphorus surface. The iodine atom forms a single bond to a phosphorus vacancy to create a neutral defect. Right: The band structure of substitutional iodine. No states are observed in the bandgap.

Julian assessed three possible iodine configurations and found substitutional I to be the least unfavourable, albeit with still a high energy cost.

Iodine adsorbs on the black phosphorus 2.74 Å above the P lattice with a positive adsorption energy of 0.3 eV per atom, indicating it is unfavourable to break up a I$_2$ molecule to physisorb the atoms to the surface.

The most suitable intercalated position would be along the channels created by the buckled rows, but the large overlap in electron density between the I atoms and the lattice make even this position unfavourable.

The atomic and electronic structure of a substitutional I defect is represented in Fig. 6.6. Creating a substitutional I defect has a significant cost of 1.39 eV (compared to the chemical potential of a P atom in the lattice). When a P atom is removed from the
surface and a 2-coordinated vacancy is created, the I atom supplies a third bond to create an electronically neutral defect without bandgap states[82].

6.2.2 Calculations of a substitutional Sn defect

Similarly to iodine, three likely configurations of Sn defects were calculated. In particular, Julian studied Sn atoms adsorbed to the surface of BP, Sn atoms intercalated in the van der Waals gap between BP layers and substitutional Sn atoms that are incorporated into the BP layers. Adsorbed and intercalated Sn structures were calculated on 4-layer slabs of BP. Substitutional Sn was calculated on both monolayer and 4-layer slabs. The 4-layer calculations were used to determine the formation energies of different charge states of Sn substitutions, while monolayers were used to visualise the extended hydrogen-like states that form around a negatively charged Sn substitution.

It was found that adsorbed Sn atoms bind to two surface P atoms in a bridging configuration with a formation energy of -2.76 eV. Intercalated Sn atoms bond to four P atoms: two in the upper BP layer and two in the lower BP layer. This defect configuration results in large atomic displacements in the BP layers and therefore has a higher formation energy of -1.82 eV. The most negative formation energy of -3.28 eV is obtained for substitutional Sn atoms suggesting that this is the energetically favourable defect configuration. This formation energy is comparable to the calculated cohesive energy of a P atom in BP, which is -3.25 eV, indicating that such defects easily form during the growth of BP.

![Figure 6.7: DFT modelling of substitutional Sn in 4-layer black phosphorus. (a): Substitutional Sn adopts the same configuration as P, replacing the lone pair with a dangling bond that can hold 0, 1 or 2 electrons. (b): Formation energy of charged Sn defects for varying Fermi level. 0.0 eV corresponds to the valence band maximum.](image)

Fig. 6.7(a) illustrates the configuration of a substitutional Sn atom. In contrast to
cubic semiconductors, such as silicon, the $sp^3$ bonded P atoms in BP are only three-fold coordinated and have one lone electron pair. When a P atom is replaced by Sn, which is one less valent than P, only one electron resides in the lone-pair orbital and the resulting dangling bond lies in the middle of band gap. This dangling bond is passivated by an additional electron from the valence band creating a negatively charged acceptor state.

To include the effect of the tip-induced electric field which leads to a local Fermi level shift in the vicinity of the defect, we have carried out DFT calculations of substitutional Sn impurities with a shifted Fermi level. The resulting defect formation energies from these calculations are shown in Fig. 6.7(b). In agreement with our measurements, we find that substitutional Sn impurities occupy a negative charge state for most Fermi levels. The pervasiveness of negatively charged defects in black phosphorus crystals also supports that it has been consistently measured as p-type ever since its discovery.

A charge transition level to a positively charged state is found when the Fermi level is very close (0.02 eV) to the VBM, suggesting the negative Sn defects can be ionised at high positive sample bias. The neutral state is never energetically favourable due to the high energy cost of leaving the dangling bond half occupied.

The ambielectronic nature of the double-lobed defect is corroborated by the obser-
vation of a charging peak near the conduction band edge in Ref. [43] and the observation of 1s and 2p states near the conduction band with an inverted energy ordering, see Fig. S7 of Ref. [43].

Fig. 6.8 shows the calculated charge distribution (wavefunction squared) of a negatively charged Sn substitution in a monolayer of BP. Panels 1 through 7 represents energy slices at the positions shown in the central panel, which shows the LDOS of the pristine lattice in red. There is a large (monolayer) bandgap which does not contain any states, with the Fermi level located at the valence band edge. The black curve, representing a substitutional Sn defect, leaves the conduction band mostly unchanged, but introduces a series of states in the bandgap and valence band.

The states that can be identified are reminiscent of a hydrogenic series. In fact, in isotropic two-dimensional semiconductors, like MoS$_2$, the defect states exhibit the eigenstates of the two-dimensional hydrogen atom. In anisotropic materials, such as BP, the hydrogen-like states are distorted and degenerate states are split. For example, the spherically symmetric 1s state assumes an ellipsoidal shape and the 2p$_x$ and 2p$_y$ states have different energies [43].

Despite the distortion from the lattice anisotropy, we readily recognise a 1s-like state in panel 1, ellipsoidal rather than spherical. The state is located roughly in the middle of the monolayer bandgap. Panels 3-6 show, in order, the 2s, 2p$_x$, 3s, and 3p$_y$ states. The other panel that stands out, panel 7, shows the 2p$_y$ state, with one lobe brighter than the other, as seen in experiments. The charge at the top and bottom of the simulation cell is an artefact from the limited cell size. To confirm this, we show tight-binding calculations performed on bigger cells that reproduce the same set without the artefacts in Fig. 6.9.

The state shown in panel 2 is not part of the hydrogenic series, but rather is the actual bonding state of the defect. The charge distribution is shown in more detail in the top left panel. The cross-like shape of the charge distribution likely arises from the lattice configuration. Due to the buckled structure, the diagonals are the easiest direction to propagate charge along[230, 24].

The 1s state is clearly isolated in the bandgap, and hence would be expected to be distinctly recognisable. The other hydrogen-like states are increasing close in energy, all gathering near the valence band edge. One thing to emphasise is that these calculations do not map one to one onto STM measurements. What is shown here are slices at precise energies, whereas STM measurements are integrated over a range of energy from 0 V to the sample bias, and should therefore contain contributions from all states with energies below the sample bias.

The large spatial extent of the defect wavefunctions becomes an issue considering
the limited cell size DFT calculations are able to produce. DFT calculations of the defect wavefunctions were limited to monolayer models, and even then appear to show some boundary condition artefacts.

Consequently, rather than directly comparing our experimental data to the limited-size DFT calculations, we used the relaxed structure parameters from DFT to inform a tight-binding model.

To assess artefacts and distortions in energy values due to boundary conditions, we reproduced the same calculations in tight-binding, which can calculate much bigger cells. The relaxed defect structure accurately determined from DFT was used to inform the choice of dielectric constant and interatomic potential parameters used in the tight-binding model.

Figure 6.9: Hydrogen-like states arising from the negative charge on a Sn substitutional defect simulated in tight-binding. Calculations and figure by Martik Aghajanian

In tight-binding, chemical information on the defect is not maintained, but rather the substitution is treated as an anonymous negative charge. As a consequence, we do not see the Sn substitution bonding state, as in panel 2 in Fig. 6.8. The calculations are however in very good agreement on the hydrogen-like series. Tight-binding is able to reproduce each of the states seen in DFT, corroborating that these purely arise from the negative charge. The TB further confirms that the 1s state would be expected in the middle of the bandgap, with the consecutive states on the valence band edge.

Two remarkable differences are the much more confined wavefunction of the 1s
state as compared to DFT calculations, and the reduced distinction between the left and right lobe of the 2p_y state.

6.3 Scanning Tunnelling Spectroscopy and tight-binding

![Figure 6.10: Comparison of experimental spectroscopy and tight-binding calculations.](image)

(a) and (b): experimental and simulated dI/dV curves on the bare surface and over top a double lobed defect. (c) and (d): Experimental and simulated dI/dV cross-section through a double lobed defect.

Fig. 6.10 compares the tight-binding calculations of a negatively charged defect to experimental scanning tunnelling spectroscopy. The black curve in Fig. 6.10(a) shows the measured dI/dV of the BP surface at a position far away from any defects, while the red curve shows the dI/dV in the vicinity of a double-lobed defect. The clean surface has a bandgap of 0.26 eV and the Fermi level is located in the vicinity of the valence band edge, indicating that the material is p-doped. The dI/dV of the clean surface exhibits a shoulder of low, but finite conductance in the band gap between 0.1 eV and 0.25 eV. Such shoulders are well-known from tunnelling spectroscopy experiments of cubic semiconductors and are a consequence of tip-induced band bending which lifts the
highest valence bands states above the Fermi level [140]. Our measurements therefore suggest that tip-induced band bending also plays an important role in BP. The pristine $dI/dV$ further shows a small peak at +0.02 eV, partially hidden by the valence band edge. Rather than a feature of the pristine surface, this is most likely a result of not taking the I/V far enough from the defect location, and as a result still weakly observing the defect state.

The red $dI/dV$ curve taken over the defect exhibits a peak at +0.02 eV above the valence band maximum, as well as a broad delocalised peak around 0.2 eV. Throughout the bandgap, the conductivity is non-zero, suggesting the defect has some metallic behaviour.

The $dI/dV$ curves just discussed were taken over many points on, and away from, the defect, and then averaged together. To preserve spatial information, we also took a CITS map, of which a cross section is shown in Fig. 6.10(c). The cross-section cuts through the middle of the defect along its long axis, has sample bias as its y-axis, and conductivity on its colour scale. Similarly as in the $dI/dV$ curves, we can identify a localised state around 0.02 eV, and see that it is in fact double lobed. The broad delocalised peak seen in the curves is visible here as a broad region of conductivity around 0.2 eV.

The CITS cross-section further highlights the bands bending upward near the defect. In the four corners of the image, the bands can be seen to curve upward. In the case of the conduction band, this is partially hidden by the added conductivity from the diffuse state.

Panels b. and d. show the equivalent Fig. as obtained through tight-binding calculations. The simulated $dI/dV$ curves again show a bandgap of approximately 0.25 eV for the pristine lattice, and two distinct states for the defective site. These states correspond to the 1s and 2p$_y$ states as visualised in Fig. 6.9. The predicted energies for the two states agree well with experiments. The main difference between the two is that calculations find the s-state to be much more localised than experiments, manifesting in a narrow strong peak as opposed to a broad diffuse region. This was also seen in the visualisations of the states, where the 1s state was much smaller than as calculated in DFT or as seen in experiments.

In panel d., the calculated energies, and the spatial extent of the p-state agree well with the CITS map, the main difference again being the localised nature of the s-state in calculations.

The band bending of the conduction band can be identified in the tight-binding calculations as it is in experiments, but the valence band appears flat. The valence band
is calculated slightly closer to 0.0 eV, and sits right up against the p-state.

![Figure 6.11: energy slices of a 3D CITS map of a double lobed defect. Scans are 8x8 nm and colour represents conductivity. The colour scale is set so that black shows zero conductivity. The current set-point was set to 50 pA and bias was swept from -0.4 V to +0.4 V.]

The CITS map in Fig. 6.10(c) was cut vertically to obtain one spatial and one energy axis. If rather than cutting a spatial cross-section through the 3D map, we cut along the energy dimension, we obtain the slices presented in Fig. 6.11.

The slices highlight several features of the spectroscopy. Bright areas in the slices have high conductivity, and so the bandgap can be identified as the black background in the slices at 0 V and +0.15 V. The fact that we still see the defect at these biases implies there are in-gap defect states available to tunnel into.

The two defect states discussed can be clearly resolved at 0 and 0.15 V. The first shows the distinct double-lobed shape of the p-like state. At +0.15 V, we see the ellipse of the s-like state distorted by the lattice anisotropy. Note that the spatial extent of the ellipse is bigger than predicted by calculations, in line with the s-like state in Fig. 6.10(c) and (d), where the experiment shows a much more diffuse LDOS than tight-binding calculations.

Another feature illustrated in the energy slices is the band bending around the defect, more thoroughly discussed in the next section. The slice at +0.35 V shows the defect darker than the general surface, suggesting fewer conduction band states are available there to tunnel into than expected. This comes from the negative charge on the defect locally bending the conduction band upwards.

The slice at -0.3 V requires some discussion. We would expect the upward band bending to produce a circular region of enhanced conductivity. We do see this region, but part of it is masked by the double lobed shape of the defect. Rather than an actual electronic feature, this arises from STM measurements always being a convolution of spatial and electronic information. The STM tip interprets the double lobed defect state.
as an actual protrusion, because it measures a stronger tunnelling signal when hovering above it. To compensate for this perceived protrusion, the tip moves further away from the surface, resulting in a reduced conductivity and a darker area in the CITS slice.

These artefacts are common in spectroscopy measurements, and can be compensated for by 'topography correcting' an image. This topography correction uses the recorded height of the tip at each pixel, and the exponential decay constant of the material, to normalise the value of each pixel, filtering out the height distortion. Unfortunately, the decay constant depends on the LDOS and ideally would be obtained through I(z) measurements, where the tip is retracted and the current decay with distance recorded, at each pixel of the image. As this was not done for this dataset, we chose not to correct these images.

### 6.4 STM topographies

![ STM topographies](image)

Figure 6.12: Explanation of defect appearance. panels a. through d. show different tip biases: -0.35 V, +0.35 V, +0.45 V, and +0.55 V respectively. For each bias, a schematic describes the tunnelling process and a topography is shown for a pressure-grown BP surface defect (left in each case) and a sub-surface defect (right). Topographies shown are 10 x 3 nm in size.

Fig. 6.12 presents STM topographies in filled and empty states biases of a double lobed defect located in the surface layer of BP (b) and located in a subsurface layer (c). A tunnelling model explaining the appearance of the defects at each bias is schematically presented in panel a. At each bias, the same features are identified in the
surface and the subsurface defect. Their key difference are the surface defect appearing brighter, and the surface defect showing a protrusion at its centre, while the deeper defect shows a node.

In filled states tunnelling, at -0.30 V, the surface defect exhibits the double lobed p-state superimposed on the valence band states. The bright central protrusion arising from the defect reconstruction is seen at all biases and only observed for defects in the surface layer. When imaging this defect close to the conduction band edge, at +0.35 V, we observe a combination of the p-like and s-like state, superimposing an elliptical protrusion on the double-lobed structure. We further observe a long-range enhancement in tunnelling current around the defect, spanning more than twice the length of the p-like state. At slightly higher bias, at +0.45 V, this long-range effect is not apparent and we only observe the combination of p- and s-like states. Finally, at high positive bias, +0.55 V, a long-range depression is superimposed on the p- and s-like states. The set of topographies in panel (c), corresponding to a subsurface defect, do not contain the bright central protrusion, and only weakly show the p-like and s-like states. This allows the long-range effects at each bias to be more clearly identified.

Our model, illustrated in panel (a), considers three contributions to the tunnelling current beyond the valence and conduction band states: tunnelling through the double lobed p-like state, tunnelling through the diffuse elliptical s-like state, and a long-range Coulomb contribution resulting from defect-induced and tip-induced band bending around the negatively charged defect. A combination of these three contributions is able to explain the appearance of the defect at each bias.

In filled state tunnelling, at -0.35 V, we tunnel out of the sample and can only observe states that hold electrons and we get the majority of our signal from valence band states. Beyond this, we can tunnel out of the p-state, which sits close enough to the Fermi level to be partially filled at liquid nitrogen temperatures, giving us the double-lobed shape. The upward band bending from the negative charge on the defect, and the downward band bending from a negative tip potential cancel each other out to keep the bands flat, resulting in no further protrusion or depression. When scanning near the conduction band edge, at +0.35 V, we are now tunnelling into the sample. We can access the p-state and the s-state, both sitting above the Fermi level. At this bias, potentials from the defect and from the tip both result in upwards band bending. This lifts valence band states over the Fermi level, at which point they become available for tunnelling into. Having more states available for tunnelling results in the overall protrusion seen in the topographies. As the bias is increased to +0.45 V, we now access more of the conduction band states, and the valence states gained from the band bending are offset
by conduction states lost, resulting again in no net protrusion or depression. At higher bias still, at +0.55 V, we are less sensitive to the p-state and valence states, now far away from the tip Fermi level, and more sensitive to the lost conduction band states, resulting in a depression superimposed on the defect.

The model shows that our observations of the double lobed defects in both filled and empty states tunnelling are consistent with a negatively charged defect with two in-gap states.

6.5 Charge state switching

As the DFT formation energy graph showed in Fig. 6.7, the substitutional Sn defects can switch between a positive and negative charge state when the Fermi level is shifted by an electric potential. Specifically, it showed that for most Fermi energies, the negatively charged Sn state is most stable, but for a Fermi energy within 0.02 eV of the valence band maximum, the positively charged state becomes favourable.

The Fermi energy can be shifted by the contact potential from the STM tip. The stronger this potential, the bigger the locally applied field[139].

Consider a strong positive sample bias. The grounded tip comparatively has a strong negative potential, creating a region of electron depletion under the tip. This manifests as an decrease of the Fermi level, which is the highest occupied electron energy at zero kelvin.

Note that while theorists tend to view this process as a shifting of the Fermi level, experimentalists have a perspective that focuses on the bending of electron energy band. Saying that a strong positive sample biases shifts the Fermi level down is equivalent to saying it bends the valence and conduction bands up. In both cases, the relative position of the Fermi level is shifted towards the valence band.

Figure 6.13: Pressure-grown BP STM topographies of two double lobed defects illustrating the charge state switching in positive bias.
Fig. 6.13 shows two double lobed defects at different bias. In negative bias, the defects can be seen in their usual p-like state. In the second panel, at +0.3 V, the defects appear with a very strong elliptical contrast superimposed. The contrast arises from the switching from a negative to a positive defect, which empties the dangling bond on the defect. The empty state becomes available for tunnelling into, enhancing the tunnelling current.

At higher biases, this contrast area can be seen to expand. At higher bias, the potential difference between tip and surface is stronger, and reaches farther as the tip moves around the surface. As a consequence, the Fermi level shift required to flip the defect from negative to positive is reached sooner as the tip moves towards the defect, resulting in a larger area of high contrast.

We can further see that the relative contrast between the positive charge state and the background decreases as bias increases, with the +0.3 V defect standing out more strongly than at +0.5 V. This is not related to the net brightness of the defect itself, but to the fact that the tunnelling current is an integration of all states from the Fermi level to the scanning bias. At +0.3 V, there are very few conduction band states making up the background, and the defect is comparatively strong. As the bias increases, more conduction band states are accessed, washing out the contrast from the defect.

Two more things to note in these scans are artefacts. The high contrast states have dark regions horizontally to either side. This arises from the post-processing background line subtraction, which takes the average value of a horizontal scan line and subtracts that from each point on the line. As the high contrast makes this average artificially high, the rest of the line appears overly dark compared to the pristine surface.

Finally, the higher bias images can be seen to have 'ghosts' of the defects to their left. This indicates a double tip, duplicating the defect. The distance between the real and ghost defects is the distance between the two tip apexes. Interestingly, the ghost defects are still show their double lobed shape. This suggests the secondary tip has a shape that applies less of a local contact potential, not shifting the Fermi level enough to induce the charge state switching.

The 3D representation of the +0.4 V image shown in Fig. 6.14(a) illustrates the abrupt switch to the high contrast state, as the dangling bond is depopulated and becomes available for tunnelling.

It further shows the original defect extending under the high contrast region. This is a point worth elaborating on as it supports the idea of a charged defect creating a potential well that hosts a hydrogen-like series of states. The negative or positive charge comes from the defect bonding state itself. The hydrogen-like series of states
arises from the potential well created by the defect but can be produced by any charged defect.

When the defect is flipped from negative to positive bias, there is still a potential well, this time inverted, which again holds hydrogen-like states. This is illustrated by the schematics in Fig. 6.14b.

This ambi-electronic nature to the defect has been discussed in other publications[43]. A difficulty arises from the unpredictable contact potential from the tip.

The contact potential is dependent on the Fermi level of the tip, which itself depends on the shape of the tip apex and the density of states arising from it. As it is impossible to determine the exact shape of the tip apex, there is little we can do to predict the contact potential, which is known to vary within a range of approximately one eV for different tungsten tip shapes.

As a consequence, the onset of the charge state switching can also vary in a one eV range, making it hard to pinpoint the energies of the states involved and develop a definitive model of the process.

6.6 Chapter summary

Extended charged double-lobed defects have been ubiquitous in STM studies of black phosphorus. Recent theoretical studies have shown that the early interpretations of the defects arising from single vacancies are not consistent with the expected behaviour of such vacancies[98, 82]. Single missing phosphorus atoms were found to be highly mobile within the material at room temperature as well as at liquid nitrogen temperatures, and on average encounter each other every 3 seconds. Upon meeting, it is energeti-
ically favourable for mono vacancies to combine to form stable, electronically neutral divacancies[82]. Consequently, we have sought to explore alternative origins to these defects.

Defects were found to be identifiable as left- or right-handed based on which lattice site they are located in, and could be observed up to 5 layers deep. Only defects in the surface layer reveal the atomic rearrangement at their centre. The distinction between left- and right-handed defects likely arises from a symmetry-breaking at the black phosphorus surface, manifesting as a slight protruding of one side of each atom row as compared to the other. The origin of this buckling is not yet clear.

Theory calculations performed by Alex Shluger’s group at UCL and Johannes Lischner’s group at Imperial College allowed for the modelling of likely tin and iodine defects, as these impurities were identified in our chemical analysis. Iodine defect were found to be unstable on the black phosphorus surface. The least unfavourable configuration is substitutional but still has a large energy cost of 1.4 eV. Amongst tin-related defects, substitutional tin was found to be most stable. Its formation energy is similar to the cohesion energy of P atoms, suggesting it is easily incorporated in the lattice during crystal growth.

DFT calculations show a tin substitution forms a negatively charged defect at almost all Fermi level positions. A transition to a positively charged state can be forced through strong tip-induced band bending, resulting in a sharp change in tunnelling contrast in STM. The potential well around the negative charge hosts a distorted hydrogen-like series of states, of which the 1s and 2py states are accessible in the bandgap and valence band edge respectively. Other hydrogen-like states sit deeper in the valence band and cannot be distinguished from normal valence band states.

A tunnelling model incorporating interactions with the 1s and 2py states, as well as tip-induced and defect-induced band bending, is able to explain the appearance of the double lobed defects observed in STM studies at all sample biases.

Our experimental findings are consistent with a substitutional tin impurity whose negative charge contributes to the p-doping of the material, and whose double lobed appearance arises from a set of hydrogen-like states originating in a potential well creating by the negative charge.
7 Phase changes and coexistence at high pressure in black and blue phosphorus

In the previous chapter, we discussed the role defects play in the doping and the electronic properties of black phosphorus. We found the negatively charged defects provided local conductive channels in the black phosphorus bandgap, and by accepting electrons into their dangling bonds contributed to making black phosphorus a p-type semiconductor. Other studies showed that in monolayers, the high carrier Hall mobility predicted for black phosphorus is only achieved in a Van der Waals heterostructure with a graphite back gate which screens charged impurities.[231] Metal dopants are predicted to induce a transition from semiconductor to semi-metal and highlight how intentional defect engineering can benefit the versatility of black phosphorus for heterostructure applications.[232]

Impurities are also known to affect the structural properties of materials. Boron nanotubes were found to change their cylindrical structure and consequently their band structure as a result of C, N, and P doping[233]. Phases changes are known to be delayed[234] or inhibited[235] due to defects.

Phases changes in black phosphorus are of particular interest to us in light of the next allotrope of phosphorus as pressure is increased, rhombohedral blue, which is predicted to be exfoliable similarly to black phosphorus, and would be suitable as a semi-metal layer in Van der Waals heterostructures[24][90][13]. At higher pressures still, black phosphorus goes on to exhibit a simple cubic and simple hexagonal phases, as well as superconductivity[236] and a transition to a semi-metallic band structure[237].

Theoretical studies predict blue P to be equally stable at ambient conditions as its black counterpart, with the further advantage of being air-stable when combined with h-BN[238] but experiments have yet to be successful in isolating it at ambient conditions. Recently, Blue Phosphorus monolayers were produced on a Au surface through in-vacuum epitaxial growth[31]. However, similarly to other 2D material as silicene or germanene, these layers have not been shown to be stable in isolation from a supporting substrate. Blue phosphorus has attracted attention in several theoretical studies highlighting its promise for 2D applications.[91][104] In particular, blue phosphorus could be used as precursor to produce nanoribbons and nanotubes with tunable semiconducting properties[95].

These studies highlight the potential of blue phosphorus for nanoscale applications and have incited several new studies into the high pressure phase changes black phosphorus undergoes[29, 239]. While these studies are typically able to identify the transi-
tions from black to blue to simple cubic phosphorus as pressure is increased, gaps in the literature as to the wide range of pressure transitions occur in, the coexistence of phases, and the pressure behaviour of Raman active modes remain.

The transition to rhombohedral and then simple cubic P has been observed in XRD experiments as well as in Raman spectroscopy studies, but the phase transitions are often quoted with a large range of different pressure values between experimental or theoretical studies[240, 241, 29]. Slow transitions with a large degree of co-existence of different phases are reported[241]. Studies at low temperature did not detect the blue phosphorus peaks at all[242]. As it has recently come to light that black phosphorus is prone to contain metal and halide impurities[40], which are known to affect phase transitions, further studies of black phosphorus at high pressure are warranted.

We sought to investigate several aspects of black phosphorus’ high pressure behaviour with this study. Firstly, to observe the transition between black and blue phosphorus in respect to the broad range of pressures this occurs at and how slow the process is. Secondly, to observe the hysteresis in the transition between the two phases, which is an indication of the stability of the blue phosphorus phase. As an extension to this, we sought to explore whether repeated cycling of the pressure range would increase the hysteresis to the point where the black and blue phases could coexist at ambient pressure. This is of particular interest as it is a first step towards isolating blue phosphorus at ambient conditions, which experimental studies have yet to achieve.

To investigate this, we have studied the high pressure behaviour of black phosphorus samples produced through the vapour transport method using Diamond Anvil Cells (DACs) to controllably vary pressure while maintaining an optical path to the sample. Two pressure dispersing media were used to ensure hydrostatic conditions throughout the observed pressure range. Samples were repeatedly cycled through a pressure range up to 27 GPa. Raman spectroscopy allowed us to identify the phase changes in the material and observe the respective vibrational modes.

7.1 Phase transitions in black phosphorus

Black phosphorus undergoes two phase changes with increasing pressure, first to a semiconducting rhombohedral state (designation A7, known as blue phosphorus), then to a simple cubic phase at higher pressure. While older studies[242, 241, 30, 240] focused on mapping out the phase diagram of phosphorus, often starting from white and varying temperature and pressure, recently, studies have focused specifically on the black and blue phases[29, 31, 239], and the transition mechanism between them[92].
Fig. 7.1 summarises a study by Clark et al of the different phases of phosphorus.[243] Starting from white phosphorus, they investigated changes in lattice constants and molar volume using synchrotron XRD. In particular, they identify the transition between the orthorhombic (black) and rhombohedral (blue) phases of phosphorus to occur at 4.8 GPa at room temperature. The rhombohedral to simple cubic transition was found at 10 GPa. No further phase changes were seen upon increasing pressure.

The transition mechanism between black and blue phosphorus as described by Bouffelfel et al[92] is shown in Fig. 7.2. As pressure is applied, the black buckled structure flattens into a corrugated hexagonal lattice. As the c-axis is compressed, the lone pairs on each phosphorus atom combine across the VdW gap and bond be-

Figure 7.1: Phase diagram of metastable phases of white phosphorus: phase changes were identified through XRD and calculations of molar volume. Dashed lines represent estimates. bP here designates black phosphorus. The rhombohedral bP phase has recently been dubbed blue P. Taken from Clark, 2010 [243]
Figure 7.2: Transition mechanism between black (A17) and blue (A7) phosphorus. Taken from Boulfelfel et al [92]

between atom sheets. This result in vertical bond chains throughout the material, inducing in-plane strain which breaks in-plane bonds and recreates lone pairs. This lone pair re-orientation results in a restacking of the material in a direction diagonal to the original.

The transition from an orthorhombic to rhombohedral to simple cubic lattice is visible in XRD [29], as well as Raman spectroscopy [239] where new sets of peaks can be seen to emerge and co-exist with the black phosphorus peaks. The complete transition occurs over a wide pressure range. The rhombohedral peaks are first identified at 5 GPa and only at 15 GPa is the spectrum exclusively simple cubic.

The literature is not as clear regarding the lattice vibrations of the different phases. Raman spectroscopy studies were performed at room temperature [244, 29, 30, 239] and at 15K [242]. While some studies identify the blue phase at room temperature [29, 30, 239] along with unidentified peaks, others were not able to resolve it, possibly due to a significant background spectrum [244] and the difficulty of distinguishing black and blue peak close to the transition pressure. The blue phase was not identified at 15 K [242]. Instead, the black peaks disappear at 7.7 GPa, after which no further Raman scattering was detected. Due to the weak overall scattering signal, it is not clear whether the blue phase was skipped and the simple cubic phase adopted, or whether the signal was too poor to resolve the blue peaks.
Figure 7.3: a. Survey Raman spectrum showing a strong peak at 1350 cm\(^{-1}\) from light reflected by the diamond. The three black phosphorus peaks are located at much lower wavenumbers and experience no other interferences. b. Zoom in on the black phosphorus peaks shows their location in good agreement with literature values (363, 440, 467 cm\(^{-1}\)) [245].

7.2 Sample preparation and ambient spectra

As described in the characterisation methods chapter, high pressure measurements use a diamond anvil cell within which hydrostaticity is ensured by a pressure-dispersing medium. Salt is a suitable medium for measurements at ambient and low pressures, but undergoes a phase change and loses its hydrostatic behaviour at 10 GPa. NaCl is ground into powder and then pressed into small chips, which are introduced below and on top of the sample in the chamber. Liquid nitrogen behaves hydrostatically up to 30 GPa but misses out on lower pressures. The DAC is filled with the sample and a small ruby crystal, which acts as pressure gauge, and closed but not tightened. The cell is then submerged in liquid \(N_2\), which floods the sample chamber. The cell is tightened while submerged, and heated back to room temperature. The \(N_2\) stays trapped in the sample chamber and disperses the applied pressure. Because the cell is already tightened, we cannot take measurements at ambient or low pressures.

Spectroscopy samples were prepared using the DACs described. 50 micron diameter flakes of black phosphorus were quickly transferred from an air tight container to the DAC to minimise sample oxidation. Separate cells, filled with salt and nitrogen, where prepared for low pressure and high pressure measurements. With the DAC closed, but not tightened, Raman spectra were taken as reference at ambient conditions, as shown in Fig. 7.3. Panel a. shows the mostly screened Rayleigh scattering at 0 cm\(^{-1}\), the three
Figure 7.4: Pressures extrapolated from PL shift of ruby R1 and R2 peaks. The peaks are clearly defined and fitted throughout the range relevant to our experiment.

BP peaks around 400 cm\(^{-1}\), and the first diamond peak at 1350 cm\(^{-1}\). A high-resolution Raman spectrum of the BP peaks identifies the \(A_{1g}\), \(B_{1g}\) and \(A_{2g}\) peaks at 363 cm\(^{-1}\) ± 2 cm\(^{-1}\), 441 cm\(^{-1}\) ± 2 cm\(^{-1}\), and 467 cm\(^{-1}\) ± 1 cm\(^{-1}\) respectively, in good agreement with literature values: 363, 442, and 467 cm\(^{-1}\) respectively[86].

Wavenumber values for Raman modes, and photoluminescence values for the ruby crystal are obtained by fitting Lorentzian curves in the OriginPro2015 graphing software. For ruby photoluminescence, the standard error of the fitted peaks depended strongly on whether the sample was correctly in focus under the microscope. The standard errors on photoluminescence peaks were found to be 0.0027 nm for measurement adequately in focus, and 0.012 nm for measurements that were taken out of focus. Converting to pressure, this results in errors of 0.008 GPa and 0.03 GPa respectively. While digital peak fitting offers small errors, our measurement accuracy is also limited by the spectrometer step size. In the case of ruby PL measurements, the step size was 0.26 nm. An accuracy of half that step size results in a pressure uncertainty of 0.3 GPa. In the following experiments, pressure measurements will be quoted with this uncertainty. Raman peak fitting had a standard error of 2.3 cm\(^{-1}\) for the \(A_{1g}\) and \(B_{1g}\) peaks, and a standard error of 1.0 cm\(^{-1}\) for the sharper \(A_{2g}\) at pressures up to 5 GPa. With increased pressure or repeated cycling, the intensity of Raman modes significantly decreased and
peaks became broader. The blue phosphorus peaks are observed to be very broad and weak, and their behaviour with pressure is only discussed qualitatively.

Fig. 7.4 shows the progression of ruby photoluminescence peaks used in our experiment. The first measurement, in black at the bottom of the graph, was taken with the cell very lightly tightened after assembly, the higher ruby peak is identified at 694.3 nm, corresponding to an applied pressure of 0.1 GPa \(\pm\) 0.3 GPa. As the DAC is tightened, we can see the R1 and R2 peaks in the ruby spectrum shift to higher wavelengths. Either peak can be used to determine the pressure but traditionally, measurements are based off the longer wavelength R1 peak. The peaks vary slightly in intensity at each measurement. The cell needs to be removed from the spectrometer between each measurement to be further tightened. When it is replaced, the stage needs to be realigned, resulting in a slightly different focus and rotation of the cell compared to the polarisation of the light. If the sample is slightly out of focus, the resulting peaks will be weaker, while a different alignment with the polarised light will affect the relative intensity of the two peaks. The tightening screws on the cell were typically turned 1/6 rotation at a time, corresponding broadly to 0.3-0.5 GPa. For clarity, not all measurements are shown in fig. 7.4

Fig. 7.5 shows the three lattice vibrations that are commonly identified in black phosphorus. The \(A_{1g}\) mode corresponds to an out of plane vibration, with atoms bonded in the zigzag direction moving in phase, resulting in a buckling and flattening of the phosphorus layers. The \(B_{2g}\) mode and \(A_{2g}\) mode are in-plane vibrations, and correspond to neighbouring atoms moving out of phase, in the zigzag and armchair directions respectively.
There are three more modes that can be detected in few-layer BP: $B_{1g}$, $B_{3g}^1$, and $B_{3g}^2$ [86]. These modes require a polarisation component in the out of plane direction. Because exfoliated BP flakes typically expose their surface normal to the out of plane direction, these are not detected for a laser incident perpendicular to the surface.
7.3 Pressure cycling

Fig. 7.6 a. shows Raman spectra of the black phosphorus sample as pressure is increased from ambient to 11.6 GPa. The variations in the relative intensity of the Raman peaks arise from slight rotations of the cell as it is removed and replaced into the Raman spectrometer as described by Wu et al [245], who showed that the black phosphorus Raman modes go through maxima and minima as the angle between the polarisation of the spectrometer light and the black phosphorus lattice directions is varied by 90°.

The three black phosphorus peaks, initially at 363 cm\(^{-1}\) ± 2 cm\(^{-1}\), 441 cm\(^{-1}\) ± 2 cm\(^{-1}\), and 467 cm\(^{-1}\) ± 1 cm\(^{-1}\), respectively, are linearly shifted to higher wavenumbers as pressure is increased. They are visible (black curves in Fig. 7.6 a.) up to 9.1 GPa ± 0.3 GPa, but their intensity is steadily reduced from 3.4 GPa ± 0.3 GPa onwards. From 7.8 GPa ± 0.3 GPa, two new broad peaks can be identified at 320 cm\(^{-1}\) and 360 cm\(^{-1}\). There is a small pressure range between 7.8 GPa ± 0.3 GPa and 9.1 GPa ± 0.3 GPa, marked in blue on Fig. 7.6 a., during which these new modes co-exist with the three black phosphorus modes. The new peaks are attributed the \(E_{1g}\) and \(A_{g}\) modes of blue phosphorus, as shown by Akahama [30], and more recently by Joseph [29] and Gupta [239]. These modes are discussed in more detail below.

At higher pressures, marked in red, the original black phosphorus peaks can no longer be identified, and the two blue phosphorus peaks shift to lower wavenumbers. The blue peaks persist at 11.6 GPa ± 0.3 GPa, but are overshadowed by a broad band
Figure 7.7: Raman spectra of N\textsubscript{2} pressure cell as pressure is increased on the second pressure cycle. At pressures beyond 16.6 GPa, the spectra lose any features.

at low wavenumbers, outside our scanning range. The broad band is seen in both the salt and nitrogen cells, and only at pressures above 8 GPa. A reduction in intensity of the blue phosphorus peaks is expected due to the transition to the simple cubic phase. However, the simple cubic structure is perfectly symmetric and hence does not exhibit changes in polarisability, and consequently no Raman modes. This leaves the question of the origin of the broad band at low wavenumbers. Joseph et al\cite{29} attribute this band to the breakdown of hydrostatic conditions inside their cell, while Gupta et al\cite{239} attribute these this to the persistence of the blue phosphorus peaks. Both of their studies are revisited in the chapter discussion.

The transition at 10 GPa can be more clearly seen as the sample is brought back down in pressure. The lowest curve in Fig. 7.6 b., corresponding to 11.7 GPa ± 0.3 GPa, shows no indication of black or blue phosphorus peaks, but only the tail of a larger peak at low wavenumber. Between 11.7 and 10.3 GPa, the $A_1^g$ mode of blue phosphorus becomes visible again, and shifts to higher wavenumbers with decreasing pressure. Due to our limited scanning range, the $E_{1g}$ mode of blue phosphorus is only visible from 3.9 GPa ± 0.3 GPa onwards, at approximately 310 cm\textsuperscript{-1}.

The blue phosphorus modes appear strongly suppressed up to 3.9 GPa ± 0.3 GPa, perhaps suggesting a progressive transition between simple cubic and blue, or a coex-
istence of the two phases. The black to blue transition shows a large hysteresis. While the two blue phosphorus modes did not appear until 7.8 GPa ± 0.3 GPa upon increasing pressure, they are still detectable at 1.5 GPa ± 0.3 GPa upon decreasing. From 2.4 GPa ± 0.3 GPa, the black phosphorus peaks reappear with very weak intensity. The two phases coexist up to 0.5 GPa ± 0.3 GPa, when only black phosphorus peaks are identified. When black P transforms to blue P, the crystal axes are rotated by almost 90°, so that the force is mostly applied in plane as opposed to normal to planes. although in a hydrostatic environment, pressure should be evenly applied from all directions.

The Raman spectra at 10.3 ± 0.3 GPa and 11.6 GPa ± 0.3 GPa weakly show the $E_{1g}$ and $A_{1g}$ modes of blue phosphorus along with a broad peak in the lower wavenumbers. To further explore this broad peak, spectra were taken at higher pressures on the second round of pressure cycling. Spectra from 10 GPa ± 0.3 GPa to 27.5 GPa ± 0.3 GPa are shown in Fig. 7.7. Due to the repeated cycling, parts of the sample are starting to become amorphous, resulting in a strong photoluminescence background at higher wavenumbers.

The blue phosphorus $A_{1g}$ peak can be seen to persist up to 12.3 GPa ± 0.3 GPa. From 14.2 GPa ± 0.3 GPa, the blue peaks are no longer detected, and the sample seems to have fully transitioned to the simple cubic phase. No further features are detected as pressure is increased to 27.5 GPa ± 0.3 GPa. Unfortunately, our experimental range was not wide enough to follow the evolution of the broad peak at 300 cm$^{-1}$, which, in this round of spectra, also appears weaker.

The strong hysteresis seen in the black-blue transition between increasing and decreasing pressure measurements led us to explore whether further pressure cycling would enhance this hysteresis. Of specific interest was the question whether the coexistence of the black and blue phases could be extended to ambient conditions to directly observe blue phosphorus.

After the previously discussed higher pressure measurements, the sample was cycled through a pressure range of 1 GPa to 15 GPa two more times, for a total of 4 pressure cycles. The screws on the DAC were tightened and loosened one third turn at a time, and only the ruby photoluminescence peaks were measured to determine the pressure. Near the end of the fourth pressure cycle, we resumed Raman measurements, followed the pressure down to atmosphere, and increased pressure one more time.

Fig. 7.8 shows Raman spectra from 3 GPa to ambient pressure for the sample at the end of the fourth pressure cycle. The purple spectrum at the top is included for reference and shows the sample at 2.6 GPa ± 0.3 GPa as pressure was decreased on
Figure 7.8: Raman spectra of sample that has been repeatedly pressure cycled. Even at low pressure, both black and blue P peaks are visible. The top curve is from the previous cycle and is included as reference for the blue P spectrum, taken at 2.6 GPa ± 0.3 GPa.

the previous cycle. The two blue P peaks can be identified without any other apparent features. The sample was then fully relaxed. As shown on the black curve at the bottom of the graph, the sample almost completely reverts back to black P but still very weakly shows blue P peaks. The black P peaks are clearly recognisable, albeit significantly weaker and broader than for pristine black P.

When the pressure is again increased, the coexistence of the black and blue phases is more obvious, with the blue $A_{1g}$ peak clearly distinguishable. The three black peaks and the blue $A_{1g}$ have similar intensities in the 0.6 GPa to 2.9 GPa pressure range, at which point the black and blue $A_{1g}$ peaks combine. With increasing pressure, the three black P modes move to higher wavenumbers while the blue P mode relaxes. This effect is also observed by Gupta et al[239] and further discussed in the chapter discussion.

Repeated cycling results in a broadening of all present modes and an increasing photoluminescence background increasing at higher wavenumbers, suggesting the sample becomes increasingly amorphous. The black and blue phase were observed together at very low pressures, down to 0.6 GPa ± 0.3 GPa, but the blue phase was not found to persist at atmospheric pressure.
Figure 7.9: Pressure dependence of Raman modes of black phosphorus. From low to high Raman shift: the $A_{1g}$, $B_{1g}$, and $A_{2g}$ modes from a salt-filled DAC (in blue) and a nitrogen-filled DAC (in red).

7.4 Comparison of peak shifts from different pressure media

Throughout the pressure ranges, both the nitrogen and salt cells were found to be in good agreement. Fig. 7.9 shows the pressure dependence of BP Raman modes in a salt DAC (blue) and a nitrogen DAC (red).

As mentioned previously, data on the nitrogen cell could not be collected at low pressures due to the requirement to securely close the cell while it is still submerged in liquid nitrogen.

Fig. 7.9 clearly highlights the linear shift on the three modes with pressure. The $A_{1g}$ peak has a significantly faster increase in energy (4.9 cm$^{-1}$ GPa$^{-1}$) than the $B_{1g}$ and $A_{2g}$ modes (1.5 and 1.0 cm$^{-1}$ GPa$^{-1}$ respectively). This can be explained by the relative ease of compressing the material in the z direction, orthogonal to the atom planes. The Van der Waals gap separating the planes can be easily reduced, resulting in higher energy out of plane vibrations. The in-plane covalent structure is much more resistant to compressive force.
7.5 Chapter discussion

Synchrotron XRD measurements of white[243] and black[29] phosphorus samples have provided a foundation for the exploration of phase changes between phosphorus allotropes. The black/blue transition was observed between $4.5 \pm 0.3$ GPa and $7.3$ GPa $\pm 0.3$ GPa, and the blue/simple cubic transition between $10 \pm 0.3$ GPa and $13.6$ GPa $\pm 0.3$ GPa. The layered nature of black phosphorus, as well as the anisotropy in the lattice, has further fuelled studies into the lattice vibrations and phase transition in the form of Raman studies at high pressure.

Amongst Raman studies, there continues to be disparities relating to the observation of the blue phosphorus phase, as well as unidentified phases at high pressure.

Our study using two different pressure-dispersing media shows good agreement between the two cells. Black phosphorus Raman modes are found at wavenumbers consistent with literature values and are shifted linearly with pressure.

An initial study without pressure medium showed more erratic behaviour. Raman shifts with pressure were consistent up to $6$ GPa $\pm 0.3$ GPa, at which point the in-plane Raman modes maintained the same energy while the out-of-plane mode rapidly increased in pressure. Black phosphorus peaks were still identified at $12$ GPa. This indicates strong differences in pressure across the cell, as well as an uneven distribution of pressure in terms of direction, leading to the rapid compression of the sample as pressure shifts across the cell.

In DACs with a pressure medium, as pressure is increased, the blue phosphorus peaks are identified from $7.2$ GPa $\pm 0.3$ GPa onwards, in agreement with XRD studies. The black and blue phases coexist for a pressure range of approximately $3$ GPa, corroborating the slow transition mechanism as described by Boulfelfel et al[92].

At $10$ GPa, the black phase is completely gone. The blue phase peaks are still detectable, but merge into a broad band at lower wavenumbers. Joseph et al[29] assign this broad band to a new unidentified phase, and see it persist up to $18$ GPa, which is the full range of their measurements. They attribute the phase to the breakdown of hydrostaticity in their pressure cell. As they use NaCl as pressure-dispersing medium, which is known not to behave hydrostatically beyond $10$ GPa, this seems plausible. However, as in our study the broad band is observed in both NaCl and N$_2$ cells, this explanation is incomplete for us.

Gupta et al[239] similarly observe the broad band from $9.3$ GPa onwards. They identify 2 components to it, and assign these to the $A_{1g}^1$ and $E_{1g}$ modes of blue phosphorus, with a third component at lower wavenumbers, near $250$ cm$^{-1}$. They find that
as pressure is increased from 9.3 GPa to 15.1 GPa, the intensity of the blue \( E_{1g} \) rapidly increases and dominates the spectrum, while the \( A^1_g \) component gradually weakens, only weakly contributes to the broad band at 15.1 GPa \( \pm 0.3 \) GPa, and disappears completely at 17.1 GPa \( \pm 0.3 \) GPa.

Their observations are in good agreement with our when considering our measurements were taken only down to 300 cm\(^{-1}\). Indeed, our measurements observed the \( A^1_g \) and \( E_{1g} \) strongly from 9.1 GPa onwards, with both components rapidly merging into a broad band. Due to our limited scanning range, we were unable to observe the full extent of the band, nor could we see the third component Gupta et al observed at lower wavenumbers.

Gupta et al find the broad band decrease in intensity from 15.1 GPa onwards but it is still present in their last measurement, at 23.5 GPa. Our own measurements in this pressure range see the band decrease in intensity from 12.3 GPa \( \pm 0.3 \) GPa and no longer observe it from 16.6 GPa \( \pm 0.3 \) GPa onwards. The discrepancies between our studies may be explained by the fact our measurements were taken on a cell that had already previously been cycled through the pressure range once. As a result, our sample was already degraded to some extent, and the broad band we observed was significantly weaker than the one observed by Gupta. The sample degradation may also explain the comparatively early onset of the broad band, as we have seen the phases in our samples to coexist and mix more as samples are pressure cycled.

Our pressure cycling results show that there is a strong hysteresis in both the black-blue and blue-sc phase changes. While the black phosphorus phase is detected up to 9 GPa upon increasing pressure, it is not retrieved until 1.5 GPa upon reducing pressure. While this demonstrates it is possible to create samples with mixed phases, these mixed samples were never stable at ambient pressure, at which point they reverted completely back to black phosphorus.

Repeatedly cycling pressure resulted in a broadening and reduced intensity of all Raman modes, suggesting the samples become increasingly amorphous. Through repeated cycling, the samples also acquired a strong photoluminescence background, which may have obscured the increasingly weak peaks. Cycling was stopped after 4 cycles when samples had degraded to such an extent that peaks were becoming very difficult to detect.

Through our experiments, we have shown that Raman spectroscopy can be used to identify phase changes in black phosphorus. The black to blue transition was identified through the observation of the \( A^1_g \) and \( E_{1g} \) modes of blue phosphorus, and the blue to simple cubic through the disappearing of all modes.
We have observed a large hysteresis in the black-blue transition, making the blue phase accessible at lower pressures, but this was paired with an overall degradation and amorphisation of the sample. The blue phase was not found to be stable at ambient pressures.
8 Conclusion

8.1 Summary

In this thesis, I have investigated the ambient degradation and intentional degradation of black phosphorus at the atomic scale. I have studied the electronic and topographic structures of point defects on the black phosphorus surface and explored the transitions and hysteresis between high-pressure phosphorus allotropes.

I have described technique for cleaving samples and exposing atomically flat and clean surfaces. The exposed surfaces were studied for their chemical composition and samples produced through the two common synthesis methods were found to contain tin and iodine impurities. Cleaved samples were studied at the atomic scale using STM and NC-AFM. We found crystallographic and electronic properties to be in good agreement with previous studies. Several different types of point defects could be identified using the two characterisation techniques.

The double lobed defects commonly observed in STM studies have been explored thoroughly, assessing their behaviour with depth and sample bias. Through DFT and Tight-binding calculations, a model was developed based on tin substitutions which successfully describes the appearance of the defects at all sample biases explored in experiments. The defects were found to hold a negative charge, accounting for the p-type behaviour of black phosphorus as a semiconductor. The charge on the defects can be switched to positive by applying a strong positive scanning bias. The neutral charge state was not observed in the defects and was calculated to always lie higher in energy than the positive or negative states. The double lobed appearance of the defects was found to arise from a potential well forming around the negative charge which hosts hydrogen-like s and p states. These states lie in the black phosphorus bandgap or just on the valence band edge, and thus can be clearly observed when contributions from valence band states are weak.

In ambient AFM, the degradation over time of the black phosphorus surface was observed and found to slow down significantly in a low-humidity environment, supporting theories that the degradation is a multi-step process involving both molecular oxygen and water vapour. The black phosphorus surface was found to become more hydrophilic when oxidised, catalysing the degradation reaction, which forms phosphoric acid and etches away at the surface.

Defects were induced in atomic scale STM by dosing with water and silicon. Several new defect types were observed. None of them presented a negative charge or the
double-lobed shape.

At high pressure, the transitions to rhombohedral and simple cubic phosphorus were found to be slow and exhibit a strong hysteresis. After repeated pressure cycling, the blue and black phases coexisted at pressures below 1 GPa, but the blue phase was not maintained at ambient pressure.

8.2 The study in context

Layered materials are fertile grounds for the study of exotic physics. The dimensional confinement of charge within 2D layers, and the potential for perfect atomic interfaces between different materials enables the study of superconductivity, charge density waves, and interactions such as spin orbit coupling and induced secondary Dirac points.

While black phosphorus was known to be Van der Waals layered since the 1960’s, the development of exfoliation techniques has renewed interest in the material, and prompted an extensive effort to better understand its surface chemistry, electronic behaviour, and purity. Proof of concept applications have appeared in rapid succession. Of particular interest are the adjustable direct bandgap, the predicted high charge carrier mobility, and the optical and electronic anisotropy of the material. Any application seeking to harness these strengths is reliant on the material being pure, with defects either passivated or absent.

In this thesis, I have presented a detailed study of the defects present in black phosphorus at the atomic scale. Through the combination of NC-AFM and STM, we have shown that while these defects have minimal structural impact on the black phosphorus surface, they result in extensive electronic defects, which likely act as charge carrier scattering centres, and limit applications relying on high charge mobility. Through chemical analysis and calculations, we have shown that substitutional tin defects are present in black phosphorus and could account for the observed defects. Identifying tin as an origin of these defects is a key step towards passifying or removing them, possibly by focusing efforts on avoiding contamination during the black phosphorus synthesis process.

Our study of the ubiquitous double lobed defects in black phosphorus further improves the understanding of the p-type character consistently found in black phosphorus. We find the defects to hold a negative charge and introduce electronic states in the bandgap near the valence band edge. Our insights here inform the design of electronic devices based on black phosphorus by defect engineering.
Finally, our studies contribute to the understanding of the high-pressure phase transitions in black phosphorus. In particular, we found that repeated pressure cycling promotes the coexistence of phosphorus phases, stabilising a mixed phase of black and blue phosphorus for a wide range of pressures. However, there is a trade off between the increased mixing of phases and the degradation of the material and even when accepting significant degradation, the blue phase was not observed at atmospheric pressure. As such, a high pressure transition from black phosphorus does not seem to be a promising pathway to isolate blue phosphorus under ambient conditions.

8.3 List of novel achievements

- **First NC-AFM images of the black phosphorus surface**: While visiting the university of Hamburg, I used their microscopy labs to obtain the first in-vacuum atomic-resolution NC-AFM images of cleaved black phosphorus. I adopted a post cleave method suitable for their experimental set-up and found titanium coated silicon AFM tips to be best suited on the surface. The lattice parameters of black phosphorus were confirmed and the symmetry-breaking within the atom rows was observed, confirming that the observed buckling arises from structural as opposed to electronic effects. Two main types of defects could be identified, which are consistent with divacancies and substitutional tin defects.

- **Confirmed tin and iodine impurities are found in black phosphorus produced through both synthesis methods**: Through chemical analysis using XPS and SIMS, I consistently observed small amounts of tin and iodine not only in vapour-grown samples, where these impurities are expected, but also in high-pressure grown samples. Correspondence with the supplier suggested they were unable to produce impurity free samples as advertised. This calls into question the assumed impurity of samples used in other studies and may call for a review and refinement of the synthesis process.

- **Developed a tunnelling model able to explain the appearance of double lobed defects at different scanning biases in STM**: The appearance of double lobed defects in STM topographies changes with the applied sample bias. I developed a model taking into consideration contributions to the tunnelling current from tunnelling into a hydrogen-like s and p state, as well as a further Coulomb contribution from both tip-induced and defect-induced band bending. The model successfully explains the appearance of the defects at all observed sample bi-
ases, corroborating the negatively charged nature of these defects.

- **Found the ubiquitous double-lobed defects observed in STM to be consistent with substitutional tin defects**: In collaboration with the groups of Alex Shluger in UCL Physics, and Johannes Lischner in Imperial College Materials, providing DFT and tight-binding calculations respectively, we were able to determine likely configurations of tin defects in black phosphorus. We found that substitutional defects are most stable and produce a set of bandgap states consistent with the experimental spectroscopy. The negative charge on the tin defect produces a potential well that hosts a distorted hydrogen-like series of states, of which the 1s and 2p_y lie in the bandgap and are readily identified in both topographic and spectroscopic data.

- **Produced defects on the black phosphorus surface arising from water and silicon impurities, and ruled them out as possible origins for double lobed defects**: Through dosing samples with water and silicon in vacuum, I was able to produce and catalogue defects related to each of these impurities. Water was chosen due to its role in the ambient degradation of black phosphorus, and silicon because it is isovalent with tin. Each impurity produced a range of defects, none of them resulting in the double-lobes apparently inherent to the material.

- **Observed that the ambient degradation of black phosphorus slows down in a low-humidity environment**: The formation of droplets on sample surfaces was found to be significantly slower in a low-humidity environment. The process was found to be non-linear: one fourth of the humidity led to less than one fourth of the droplet growth rate. It is understood from theoretical studies that without the presence of water, oxygen will cover the surface, but not etch away at or degrade it, and only in the combined presence of water and oxygen does the material wear away.

- **Found the originally hydrophobic black phosphorus surface to become hydrophilic as oxidation occurs**: Water droplets on samples surfaces were observed to adopt perfectly round shapes on unoxidised samples, but progressively took on more amorphous, spread out configurations, until they grew large enough to once again make a round shape favourable. This suggests the hydrophobicity of the samples changes with time. Indeed, theoretical studies have found that the oxidation process creates oxygen dangling bonds at the black phosphorus surface which provide easy anchors for water to bond to, making the surface more
• Found the phase changes between high pressure allotropes of phosphorus to show a strong hysteresis: While previous studies have observed the transitions between black, blue, and simple cubic phosphorus, our study of the repeated pressure cycling across a wide range is novel, and shows the large hysteresis in recovering the black phosphorus phase after the blue is reached. While the blue phase doesn’t start to form until 7 GPa, the black isn’t fully recovered until 0.5 GPa or less. It is possible that with the right dopants, the blue phase may be stabilised enough to be retained at atmospheric pressure.

8.4 Suggestions for further work

• Perform DFT calculations to identify the reconstructions corresponding to the water and silicon-related defects: STM topographies were able to identify several defect structures after dosing with water and later with silicon. DFT calculations of common reconstructions involving these impurities would allow the identification of these defects, and shine a light on the adsorption of water on the black phosphorus surface.

• Synthesise black phosphorus samples with higher, or zero, tin content: The synthesis of samples with varying concentrations of tin impurities would further strengthen the evidence supporting tin substitutions as the origin of the double lobed defects. If samples with increased tin content, or with zero tin content were made and resulted in STM topographies with increased, or zero concentration of double lobes, this would dispel any doubt remaining as to the origin of the defects.

• Test the effect of different impurity concentrations on the onset and hysteresis in phosphorus phase changes: The concentration and nature of impurities in black phosphorus is likely to result in different onset pressures of phase transitions, and increase the hysteresis between phase changes. A sufficiently high dopant concentration would likely be able to allow the black and blue phases to coexist for a broader pressure range and may be able to stabilise the blue phosphorus phase at ambient pressure.

• Explore the screening effect of thin layer graphene as substrate to thin layer black phosphorus: This black phosphorus transistors have so far struggled to realise the theoretical potential of BP devices in terms of on/off ratio and charge
carrier mobility. Recent studies suggested using graphene as a substrate to thin layer black phosphorus provides sufficient charge screening to dramatically enhance the charge mobility. The capability of stacking thin layer black phosphorus on other Van der Waals materials is also integral to the development of stable VdW heterostructures and would provide protection against ambient degradation for phosphorus flakes.

9 References


