Development of Automated Methods for Extraction of Structural Information from Microscopy Data

By

Oleg Sergeevich Ovchinnikov

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Physics

October 31, 2018

Nashville, Tennessee

Approved:

Sokrates Pantelides, Ph.D.

Sait Umar, Ph.D.

Paul Sheldon, Ph.D.

Sergei V. Kalinin, Ph.D.

Albina Y. Borisevich, Ph.D.

Copyright © 2018 by Oleg Sergeevich Ovchinnikov

All Rights Reserved

ACKNOWLEDGEMENTS

I would like to thank all the people who made this work possible. I would like to give a special thanks to Dr. Sokrates Pantelides for taking me as his student and for all his advice through the years. I would also like to thank Dr. Albina Y. Borisevich for all her advice and guidance in the field of Scanning Transmission Electron Microscopy (STEM). I would like to thank Dr. Sergei V. Kalinin for all his support and his wealth of knowledge in atomically resolved imaging. I am also grateful to the other members of my PhD committee Dr. Sait Umar and Dr. Paul Sheldon for their guidance. I would like to thank Dr. Yevgeniy Puzyrev and Dr. Andrew O'hara for their guidance with Density Functional theory. Finally, I would like to thank Oak Ridge National Laboratory (ORNL) Center for Nanophase Material Science (CNMS) for allowing me use of the facilities and computational resources. This work was supported by a Department of Energy grant and Oak Ridge National Laboratory, Laboratory Directed Research & Development.

CONTENTS

Page

ACKNOWLEDGEMENTS iii

List of figures v

Chapter

1. Introduction 1

1.1. Density Functional Theory (DFT) 1
1.2. Scanning Transmission Electron Microcopy (STEM) 7
1.3. Scanning Tunneling Microcopy (STM) 11
1.4. Defect Identification. 13
1.5. Atomic column isolation and localization 19
1.6. Image filtering 22
1.7. Thesis overview 28

2. Hough atomic-column finder 30

2.1. Introduction 30
2.2. Methods 33
2.3. Results 35
2.4. Discussion 41
2.5. Conclusions 43

3. Atomic-column recovery and defect identification aided by DFT 45

3.1. Introduction 45
3.2. Methods and materials 47
3.3. Results and discussion 53
3.4. Outlook 61
3.5. Conclusions 63

4. Defect detection and grouping 64

4.1. Introduction 64
4.2. Methods and materials 66
4.3. Results 71
4.4. Conclusions 75

References 76

LIST OF FIGURES

Figure 1-1: Diagram showing the process to calculate the total energy using DFT with a plane-wave basis set. 5

Figure 1-2: As the probe interacts with the sample, the electrons are scattered at different angles (based on the atoms the beam interacts with). The scattered elections are then collected by the detectors. Electrons scattered at low angles can be collected by either a detector or a spectrometer. 7

Figure 1-3: Dark field image of Re doped MoS_2, this image contain atomic columns of three different intensities. Atomic columns containing S_2 having the lowest followed by Mo and finally Re having the highest intensity........ 9

Figure 1-4: Simplified schematic of an aberration-corrected STEM; examples of electron trajectories are indicated. In the (quadrupole-octupole) aberration corrector, the different trajectories in the XZ and YZ planes are indicated with dotted and solid lines, respectively. The electrons are accelerated from the gun; condenser lenses are used to adjust the beam current and beam coherence and to couple to the aberration corrector. The objective lens focuses the probe, which is scanned across the sample by the scan coils. The high-angle detector collects electrons scattered to high angles. Removable bright field and Ronchigram detectors are shown as well as the important components of the spectrometer........ 10

Figure 1-5: Modified adatom model. The underlying top layer atom positions are shown by dots, and the rest atoms with unsatisfied dangling bonds carry circles, whose thickness indicates the depth measured as discussed in the text. The adatoms are represented by large dots with corresponding bonding arms. The empty potential adatom position is indicated by an empty circle in the triangle of adjacent rest atoms. The grid indicates the 7x 7 unit cells. 12

Figure 1-6 : STEM Image of Silicon (Z =14). Atomic columns with much higher intensity contain Bismuth (Z = 83) dopants. The image also demonstrates that surface contamination can alter the intensity of atomic columns that do not contain any defects........ 14

Figure 1-7 STEM image of $LaMnO_3$ and simulations of visibility and atomic column displacements. (A) Relaxed crystal structure of $LaMnO_3$ by DFT. (B) Experimental HAADF STEM image on a 6.5 nm thick specimen along [100], with matching simulated image (white box) and projected $LaMnO_3$ unit cell. Displacements of the La-O columns due to a single La vacancy in the center are shown schematically by white arrows. (C) Simulated visibility of a single La vacancy in a 10 nm (solid) and 18 nm (dash) thick $LaMnO_3$ specimen. (D) Simulated ΔS caused by a single La vacancy in a 10 nm thick specimen. Definitions of S1 and S2 are shown in the inserted HAADF image. (E) Simulated intensity of the electron probe as it propagates along a La-O [100] column. Yellow is higher intensity. 15

Figure 1-8: (a) First eigenvector, in the upper left, and a corresponding loading map. (b) Second eigenvector, in the upper left, and a corresponding loading map. (c) Third eigenvector, in the upper left, and a corresponding loading map. (d) Fourth eigenvector, in the upper left, and a corresponding loading map. (e) Fifth eigenvector, in the upper left, and a corresponding loading map. (f) Sixth eigenvector, in the upper left, and a corresponding loading map. 17

Figure 1-9: (a, b) Dynamic process of the atomic defects created at S sites. (a) Edge-atom loss of S and (b) vacancy creation. (c) Simple W monovacancy and (d) W vacancy with a large distortion in the surrounding lattice. The open blue arrows indicate the loss of S atoms and the red open arrows indicate W vacancies. Scale bar, 0.2 nm.. 18

Figure 1-10: Schematic illustrating the fundamentals of the singular value decomposition (SVD)-based image denoising technique and the pattern matching-based techniques for identifying atoms in images. (a) The denoising process starts with sliding a small window across the given image column-by-column and then row-by-row. (b) A stack of (N − m) 2 windows, each with m × m pixels, is built by copying the contents of the window at each location. (c) This 3D stack of windows is fattened to a 2D matrix by fattening the m × m pixel windows to 1D arrays with m2 elements. SVD is performed on this 2D matrix to decompose the data into the most correlated and least uncorrelated (noise) components. The image is denoised by reconstructing the 2D matrix in c with only the most correlated SVD components and reversing the steps from c to a. (d) K-means clustering on the SVD results groups pixels exhibiting similar trends together in a cluster label map. Representative examples of repeating patterns or motifs in the label map are selected for pattern matching. (e) Each motif is compared to every section in the label map to generate a pattern matching scores' map. Each continuous-valued scores map is thresholded to generate binary maps with segments and the centroids of these segments provide the coordinates of the repeating patterns such as atoms.. 21

Figure 1-11: (a) 7 points are placed along the circumference of a circle with radius R in real space. (b) In Hough space, these 7 points from real space become 7 circles with RH < R. (c) the Hough-space circles of the 7 points from real space all overlap at the center of the real space circle when RH = R 26

Figure 2-1: (a) Circle in real space with radius R contenting 7 points in real space. (b) Hough space circles of the 7 points from real space with a RH < R . (c) Hough space circles of the 7 points from real space with a RH = R 32

Figure 2-2: Examples of test data with different levels of noise ranging from 0% to 250% of the max amplitude of the atomic column intensity... 36

Figure 2-3 : Performance of atom finder on test data with different levels of noise, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice........................ 37

Figure 2-4: Examples of test data with different atomic column width, where N is the ratio of largest width divided by smallest width in test image. 38

Figure 2-5: Performance of atom finder on test data with different atomic column width, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice. 39

Figure 2-6: Examples of test data with different atomic column intensity, where N is 1 minus the ratio of strongest intensity atomic column divided by weakest intensity atomic column in the test image.. 40

Figure 2-7: Performance of atom finder on test data with different atomic column intensity, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice. 41

Figure 2-8: (a) STEM Image of Mo–V–M–O and (b) binary image of atomic columns and background generated by the algorithm using Gaussian PSF. 42

Figure 3-1: (a) ADF STEM image of a graphene bilayer with relative rotation between the layers (mismatched or twisted graphene bilayer), as acquired. (b) The same image with a small Gaussian blur more clearly shows the presence of two types of moiré nodes in the moiré pattern (primary – higher intensity, secondary – lower intensity). (c-d) The same image with a heavy Gaussian blur used to indicate the deviations from a perfect moiré pattern due to strain. (c) Distances between secondary and primary nodes (in nm: black) and angle between secondary nodes around a central primary node (in degrees: red). (d) Distances between secondary nodes (in nm: black). 49

Figure 3-2: Example of local area extraction of a 11x11 area (red and yellow areas) around a center pixel (yellow area) from a 29x29 pixel image.. 51

Figure 3-3: (a) Atomistic model showing the optimized positions of all carbon atoms with the two layers in different colors (b) PCA filtered simulated image of the post-DFT-optimized locations. (c) PCA filtered patch from the original image. (d) Map of local 2d correlation coefficients between the filtered simulated image and the filtered experimental image. 54

Figure 3-4: (a) 2D cross-correlation map with area of interest highlighted and separated into (b) layer 1 with atom of interest highlighted and (c) layer 2 with atom of interest highlighted....... 55

Figure 3-5: (a) Plot of the effect on the local 2D cross correlation when atoms are substituted, (b) zoomed in area of cross-correlation map for pristine bi layer graphene, (c) Carbon in layer 2 is replaced by oxygen, (d) Carbon in layer 2 is replaced by nitrogen (e) Carbon in layer 1 is replaced by nitrogen, (f) Carbon in layer 2 is replace by boron, and (g) Carbon in layer 1 is replaced by boron.. 56

Figure 3-6. Displacement of the atoms perpendicular to the imaging-plane from the optimized positions produced by DFT for (a) the bottom layer, and (b) the top layer. (c) Atomistic model of the rippling relative to the layer separation. 57

Figure 3-7:. (a) Layer-separation between the two layers from the optimized positions. (b) Layer-separation (black solid line) and *z* rippling for the top layer (red dotted line) across the line profile shown in (a). 60

Figure 4-1: (a) Rejected cycle due to cycle lines crossing each other (b) rejected cycle due to an atom enclosed in the cycle (c) rejected cycle due to cycle is not the shortest path 68

Figure 4-2:: The centers of atomic columns for MoS_2 colored based on the number of points in an the cycles associated with them for (a) searching all cycles (b) using Delaunay triangulation (c) and using Delaunay triangulation plus pre-filtering. 69

Figure 4-3:: (a) Original HAADF image of Mo–V–M–O (b) atom column location colored based on number of atoms in associated cycles (c) atomic columns whose cycles are not the part of the perfect crystal (d) no perfect crystal atomic columns grouped into defects and colored based on the number of atoms in a defect (single missing column/ yellow , 2 adjoining missing atomic columns/purple, large staking fault /black) 74

Figure 4-4: (a) STEM image of Re doped MoS_2 (b) atomic columns identified to be near defect using number of points in cycles (c) area of the cycles smaller areas are due to S vacancies 71

Figure 4-5(a) raw HAADF STEM image of Mo–V–M oxide (b) interpolated map of local intensity around every atomic column in figure d with the location of atomic columns shown (c) interpolated map of size of cycle associated with every atomic column in figure d with the location of atomic columns shown (d) raw HAADF STEM image of Bi doped Si (e) location of every atomic column in figure a colored based on the local intensity around every atomic column (f) location of every atomic column in figure a colored based size of cycle associated every atomic column 72

CHAPTER 1

1. INTRODUCTION

1.1. DENSITY FUNCTIONAL THEORY (DFT)

The properties of materials are ultimately determined by the arrangement of their constituent atoms. Crystallography provides information that identifies the lattice of a crystalline material, but one relies on atomic-resolution microscopies to provide structural information for complex nanostructures, e.g., the presence of interfaces, dislocations, even point defects and point-defect clusters. Density functional theory (DFT) is a great complement to these microscopies because, by minimizing the total energy with respect to atomic positions, it can provide candidate structures and candidate defects for comparison. Once the structure is established, DFT calculations can provide predictions for electronic, optical, and magnetic properties, the properties of interfaces and defects, etc.

DFT addresses the question of solving Schrödinger's equation for a many-electron system, atoms, molecules, and solids. The first step in the process is to adopt the Born-Oppenheimer approximation, which allows for the separation of the motion of the electrons and the atomic nuclei. This separation is made possible by the difference in the masses of the two being several orders of magnitude. The Born-Oppenheimer approximation decouples the electronic and atomic motion of the particle, allowing each to be calculated separately, thus greatly reducing the number of interactions that need to be calculated.

By allowing Schrödinger's equation to be solved using a fixed spatial configuration of the atomic nuclei, the Born-Oppenheimer approximation also lays the ground work for the iterative solving

of Schrödinger's equation. Even using the Born-Oppenheimer approximation, it is still not practical to solve directly the many-electron equation, which has the form

$$\widehat{\boldsymbol{H}} = -\sum_{I=1}^{N} \frac{\hbar^2}{2\boldsymbol{m}} \nabla^2 + \frac{\mathbf{1}}{\boldsymbol{\pi\epsilon_0}} \sum_{i=1}^{N_e} \sum_{j\neq i}^{N_e} \frac{\boldsymbol{e}^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_{i=1}^{N_e} \sum_{\alpha=1}^{N_i} \frac{\boldsymbol{eZ_\alpha}}{|\boldsymbol{r}_i - \boldsymbol{R}_\alpha|} + \sum_{\beta=1}^{N_i} \sum_{\beta=\alpha+1}^{N_i} \frac{\boldsymbol{Z_\alpha Z_\beta}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_\beta|} \quad \text{(1-1)}$$

Where r are the electron coordinates and R are the ion coordinates. Historically, the next important simplification was made by Hartree, who replaced the many-electron wave function by a product of single-electron functions:

$$\boldsymbol{\Psi}(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_2) = \boldsymbol{\psi}_1(\boldsymbol{r}_1)\boldsymbol{\psi}_2(\boldsymbol{r}_2)\cdots, \boldsymbol{\psi}_N(\boldsymbol{r}_N), \quad \textbf{(1-2)}$$

accompanied by a normalization constraint. The potential that appears in the resulting one-electron Schrödinger equation depends on the one-electron wave functions themselves, calling for a self-consistent solution, i.e., the equation is solved by feeding a solution back into the original equation until the resulting solution does not change. The Hartree method was found to accurately represent the atomic orbitals but the energies were highly inaccurate. The difficulty arises because the Hartree method does not satisfy the Pauli exclusion principle. This difficult was eliminated by John C. Slater and Vladimir Fock in what is now called the Hartree-Fock method by approximating the wave function as a single Slater determinant of one-particle orthonormal

orbitals:

$$\boldsymbol{\Psi}(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \begin{bmatrix} \boldsymbol{\Psi}_1(\boldsymbol{r}_1) & \cdots & \boldsymbol{\Psi}_1(\boldsymbol{r}_N) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_N(\boldsymbol{r}_1) & \cdots & \boldsymbol{\Psi}_N(\boldsymbol{r}_N) \end{bmatrix} \quad \textbf{(1-3)}$$

This method still produces energy-level results that are relatively inaccurate, although more accurate than the Hartree method.

In 1930 Paul Dirac presaged density functional theory (DFT) in a paper titled "Note to exchange phenomena in a Thomas Atom"[1] as follows: "Each three-dimensional wave function will give rise to a certain electric density. This electric density is really a matrix, like all dynamical variables in the quantum theory. By adding the electric densities from all the wave functions, we can obtain the total electric density for the atom. If we adopt the equations of the self-consistent field as amended for exchange, then this total electric density (the matrix) has one important property, namely, if the value of the total electric density at any time is given, then its value at any later time is determined by the equations of motion. This means that the whole state of the atom is completely determined by this electric density; it is not necessary to specify the individual three-dimensional wave functions that make up the total electric density. Thus one can deal with any number of electrons by working with just one matrix density function". The key point is that "it is not necessary to specify the individual three-dimensional wave functions that make up the total electric density", but he makes this assertion without proof.

In 1964, Hohenberg and Kohn[2] proved the necessary ground-breaking theorem, showing that the ground-state total energy of a many-electron system is uniquely determined by

$$\mathbf{E}[\boldsymbol{n}(\boldsymbol{r})] = \boldsymbol{T_s}[\boldsymbol{r}] + \frac{e^2}{2}\int \frac{\boldsymbol{n}(\boldsymbol{r})\boldsymbol{n}(\boldsymbol{r}')}{|\boldsymbol{n}(\boldsymbol{r})} \boldsymbol{drdr}' + \mathbf{E}_{xc}[\boldsymbol{n}(\boldsymbol{r})] + \int \boldsymbol{v}(\boldsymbol{r})\boldsymbol{n}(\boldsymbol{r})\boldsymbol{dr}\,, \qquad \textbf{(1-4)}$$

the ground-state electron density. The total energy can then be written in the form

where

$$T_s[r] = \sum_i \left\langle \psi_i \middle| -\frac{\hbar^2}{2m_e}\nabla^2 \middle| \psi_i \right\rangle \tag{1-5}$$

is the kinetic energy. The second term is known as the Hartree energy, E_H. E_{xc} is referred to as the exchange-correlation energy functional. The electron density $n(r)$ is then constructed from one-electron wave functions that satisfy a Schrödinger equation in which the effective one-electron potential has a Hartree and an exchange-correlation term. DFT has since been established as the method of choice for electronic structure calculations and optimizations of atomic structures in solids and complex nanostructures. Walter Kohn was awarded a share of the Nobel Prize in Chemistry in 1998 for the development of DFT. In most cases, DFT is used in conjunction with pseudopotentials, whose purpose is to remove the core electrons from the calculation by replacing their effect by an effective potential. In this thesis we will use DFT calculations as a complement to algorithms that are developed for automated analysis of atomic-microscopy images of materials, including defect identification.

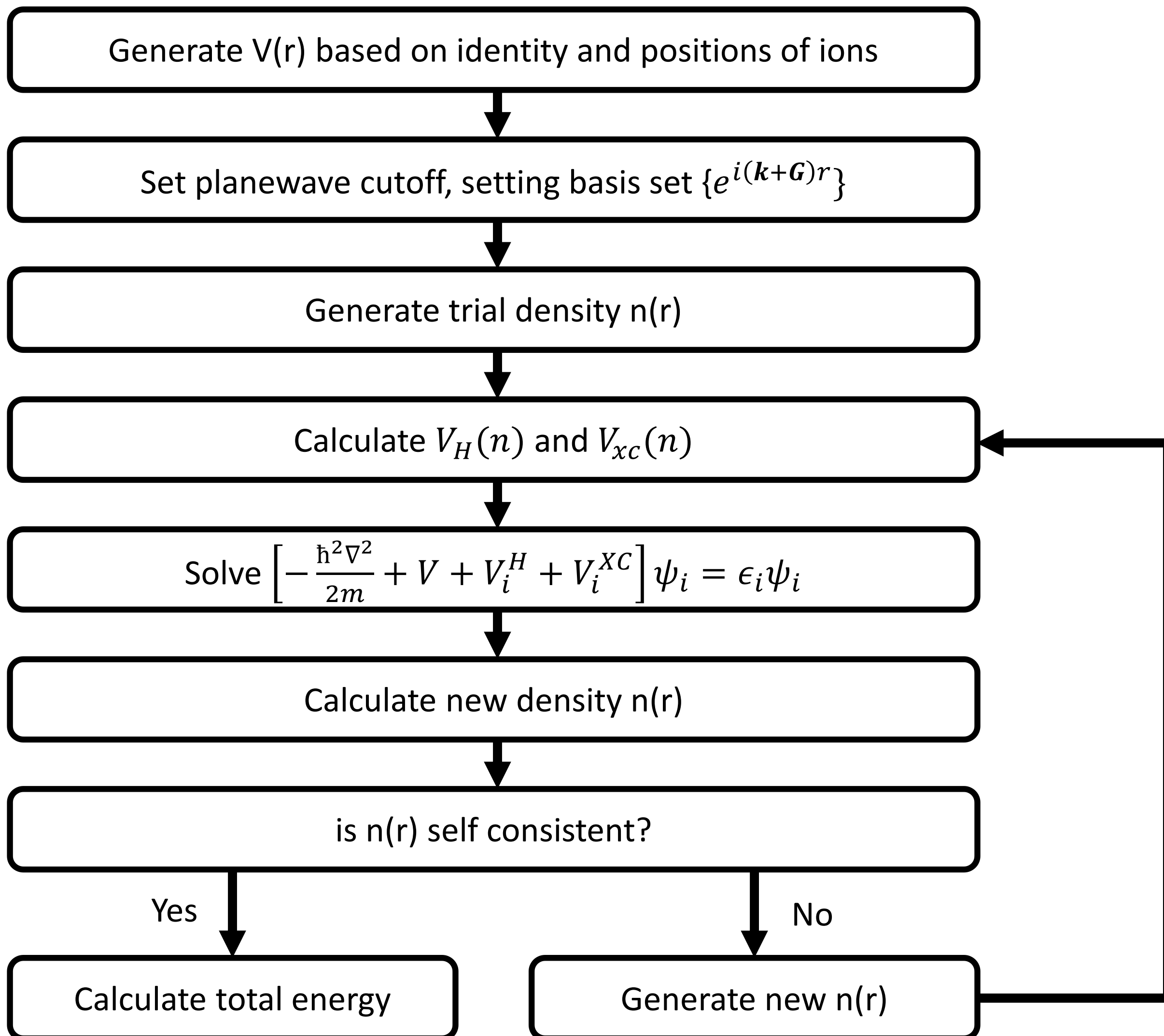

Figure 1-1: Diagram showing the process to calculate the total energy using DFT with a plane-wave basis set.

Figure 1-1 shows how to solve the DFT Kohn and Sham equations using a plane wave basis set. If E_{xc} was known exactly, then it would be possible to calculate the exact many-body total energies. However the exact form of E_{xc} is not known and instead approximations based on the exchange-correlation energy of a homogeneous electron gas as a function of density are used. The simplest

approximation is the Local-Density Approximation (LDA)[3], which adopts the expression for E_{xc} for a homogeneous electron gas and evaluates it for the materials *n*(*r*) at each position *r*. A more elaborate approximation is the Generalized-Gradient Approximation (GGA) which depends not only on the local density but also on its local gradient [4], [5].

DFT captures all interactions that lead to chemical bonding via overlapping wave functions. Atoms can, however, interact even with negligible wave function orbitals by inducing dipoles in each other, which then lead to dipole interactions. These are known as van der Waals (vdW) interactions. They are important in layered materials, e.g., graphite or MoS_2, as they are responsible for holding the layers together. They are usually included as a semiempirical correction to the total energy

$$E_T = E_{DFT} + E_{vdW} \tag{1-6}$$

where E_T is the total energy, E_{DFT} is the DFT total energy as in Eq. (1-4), and E_{vdW} is the correction to the total energy. In one of the commonly used forms[6], E_{vdW} is defined by

$$E_{vdW} = -S_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}\left(R_{ij}\right). \tag{1-7}$$

Here S_6 is a scaling factor, N_{at} is the number of atoms, C_6^{ij} is the dispersion coefficient for a pair of atoms, R_{ij} is the interatomic distance between at a pair of atoms and f_{dmp} is a dampening force defined by

$$f_{dmp}(R) = \frac{1}{1 + e^{-i\propto(R/R_0-1)}}. \tag{1-8}$$

where R_0 is the sum of atomic van der Waals radii. The reason this damping force is chosen is because for small R it approaches zero fast enough that non-van-der-Waals bonds are not effected significantly.

1.2. Scanning Transmission Electron Microcopy (STEM)

Scanning Transmission Electron Microcopy (STEM) is an imaging technique that is capable of imaging a crystalline material's atomic structure based on an intensity that is roughly proportional to the square of the total atomic number (Z) in each atomic column[7]. STEM has been used to provide sub-angstrom resolution of individual atoms [8]–[10]. STEM works by focusing an electron beam into a tiny probe and then scanning it across the material. As the electron beam passes through the sample, it interacts with atoms in the material, causing

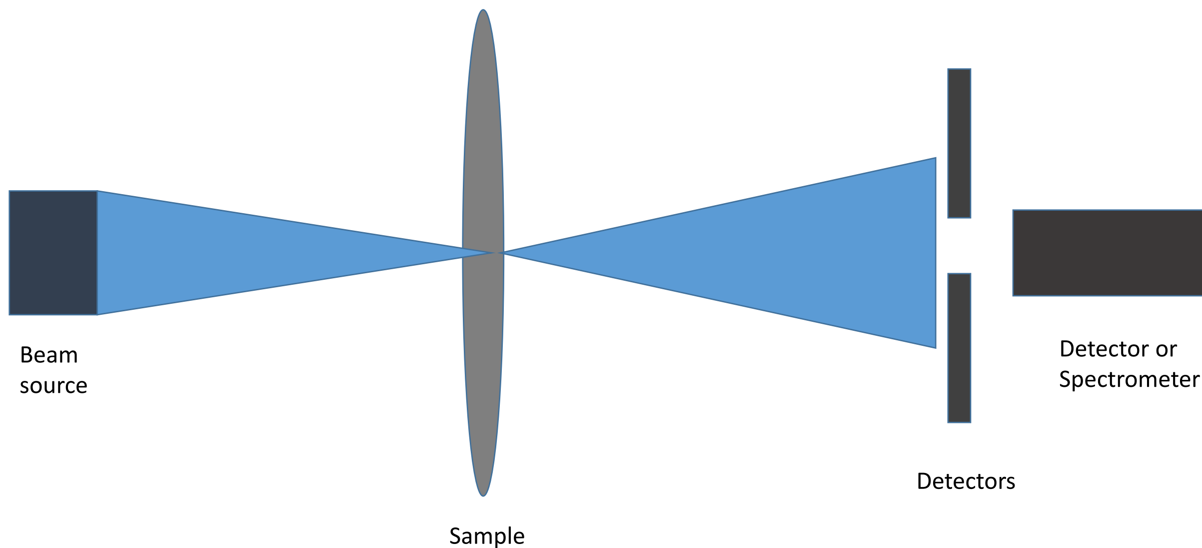

Figure 1-2: As the probe interacts with the sample, the electrons are scattered at different angles (based on the atoms the beam interacts with). The scattered elections are then collected by the detectors. Electrons scattered at low angles can be collected by either a detector or a spectrometer.

electrons to be scattered. These scattered electrons are collected by a series of detectors or a spectrometer and this information is then used to construct an image of the material (Figure 1-2). When then beam interacts with atomic columns it scatters at high angles and the electrons are collected by detectors. Images reconstructed with information from these detectors are referred to as annular dark field (ADF) and high annular dark field (HADF) images . The electrons that pass through the sample without interacting with atoms can also be collected by a separate removable detector. Images reconstructed with information from low angle scatter are referred to as bright field (BF) images. Instead of using the removable detector to capture the low-angle electrons, a spectrometer can be used, which registers electron-energy loss spectra (EELS).

Bright field images tend to be hard to interpret as they depend on the sample thickness and the amount of defocus. Atoms can appear either darker or brighter than the background. The images formed from the dark field images are much easier to interpret, with bright atoms on a dark background. The intensity of the atoms is proportional to Z^2, where Z is the total atomic number in the column of the atom (Figure 1-3). The proportionality to Z^2 is due to electrons passing close to the nucleus of an atom to undergo Rutherford scattering. As lower angles of scatter are added into the construction of the image, the intensity proportionality starts to move from Z^2 towards Z. The ease with which atoms or atomic columns can be identified and atomic composition compared makes dark field images the preferred imaging technique in many applications.

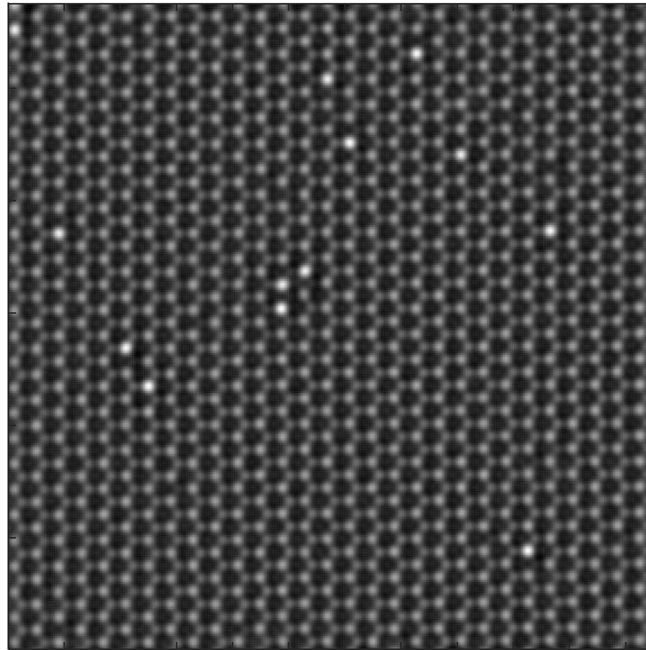

Figure 1-3: Dark field image of Re doped MoS_2, this image contain atomic columns of three different intensities. Atomic columns containing S_2 having the lowest followed by Mo and finally Re having the highest intensity

Image courtesy of Shize Yang, Wu Zhou, and Matthew F. Chisholm

The electron beam begins with the acceleration of electrons from a source. The faster electrons are traveling, the smaller their wavelength is, according to the de Broglie equation, resulting in higher resolution images. Due to this effect, very high energy sources tend to be used and, in the days before aberration corrections, these could get into the megavolt. However, high-energy electrons can damage the sample and cause the sample to undergo changes in the structure. Most modern STEMs operate between 60-300 kV. As mentioned, aberration correction has allowed for microscopes with higher resolution without requiring megavolt sources. Modern aberration-corrected STEMs (Figure 1-4) can achieve sub Angstrom resolution.

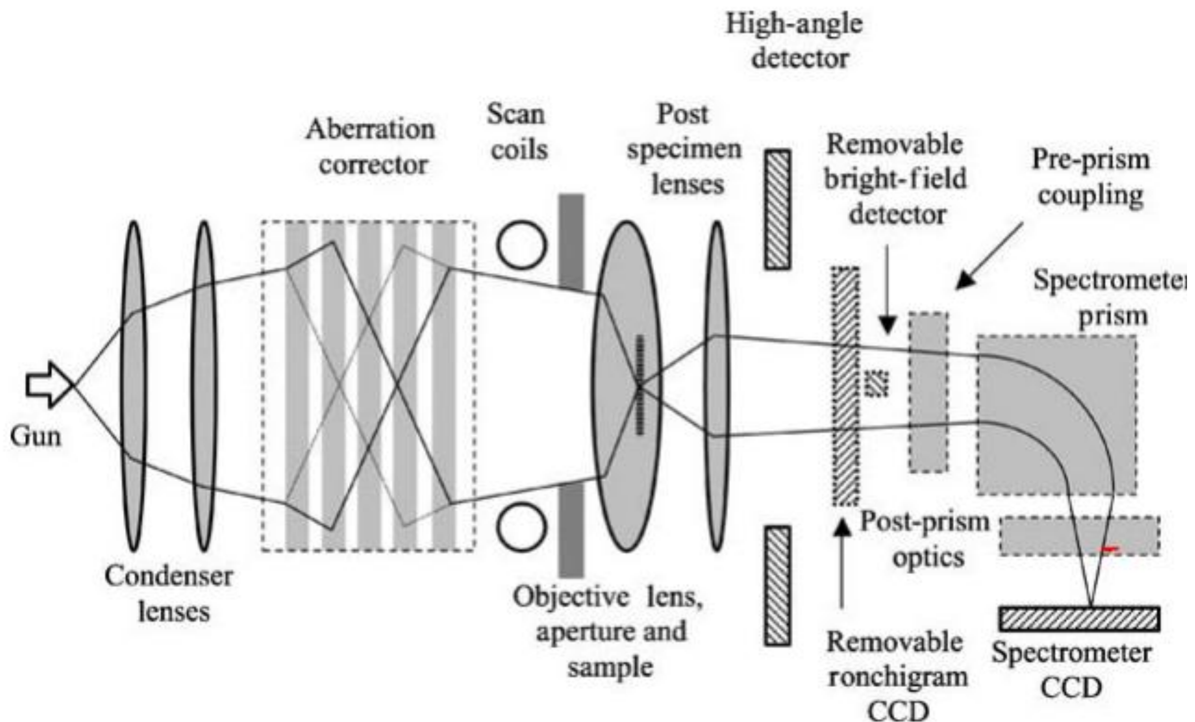

Figure and caption taken from: Varela, M., et al. "Materials characterization in the aberration-corrected scanning transmission electron microscope." Annu. Rev. Mater. Res. 35 (2005): 539-569.

Figure 1-4: Simplified schematic of an aberration-corrected STEM; examples of electron trajectories are indicated. In the (quadrupole-octupole) aberration corrector, the different trajectories in the XZ and YZ planes are indicated with dotted and solid lines, respectively. The electrons are accelerated from the gun; condenser lenses are used to adjust the beam current and beam coherence and to couple to the aberration corrector. The objective lens focuses the probe, which is scanned across the sample by the scan coils. The high-angle detector collects electrons scattered to high angles. Removable bright field and Ronchigram detectors are shown as well as the important components of the spectrometer.

1.3. Scanning Tunneling Microcopy (STM)

Scanning Tunneling Microcopy (STM) is a surface imaging technique that works by bringing a small metal tip close enough to the surface of a material that the vacuum tunneling resistance between surface and tip becomes measurable. The tip is then moved along the surface maintaining a uniform tunneling resistance by adjusting the height of the tip. By mapping the height of the tip across the surface a detailed map of the surface can be produced. This was first done in 1982 by Gerd Binnig and Heinrich Rohrer at IBM's Zurich Research Laboratory. They successfully mapped the surface of both gold and $CaIrSn_4$. The importance of this accomplishment was recognized in 1986 when the Nobel Prize in Physics was awarded to Binnig and Rohrer for the scanning tunneling microscope.

One of the first major discoveries achieved using a STM was the determination of the structure of silicon (111) surfaces[11] (Figure 1-5). Images produced by Binnig and Rohrer showed that Si(111) has a rhombohedral unit cell with sixfold rotational symmetry. Until 2001, STM was performed under ultrahigh vacuum. In 2001, Lars Österlund created an STM that operated at up to 1 bar of pressure. This feature allowed the STM to study the surfaces of materials undergoing a chemical reaction due to gas being inserted into the chamber. Österland used this new STEM to study how a copper surface reacts with hydrogen gas.

While the material of the tip seems to have little impact on the technique, with W, Mo, and stainless steel tips producing similar results. However the shape of the tip can greatly affect the image[12]. The shape of the tip was shown to not only affect the resolution but can also cause different shapes to be seen for the same object. This is important as the shape of the tip

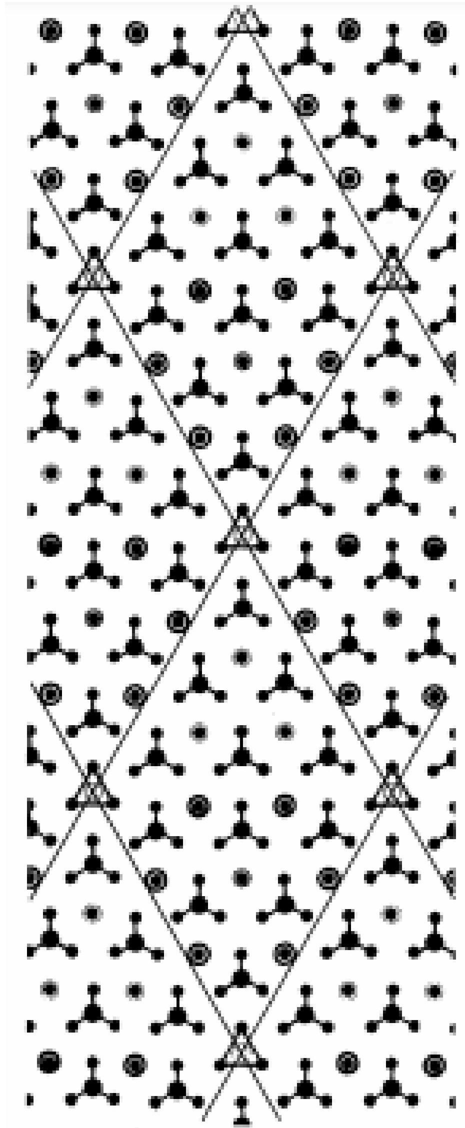

Figure and caption taken from : Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. 7 × 7 Reconstruction on Si(111) Resolved in Real Space. Phys. Rev. Lett. 1983, 50, 120–123.

Figure 1-5: Modified adatom model. The underlying top layer atom positions are shown by dots, and the rest atoms with unsatisfied dangling bonds carry circles, whose thickness indicates the depth measured as discussed in the text. The adatoms are represented by large dots with corresponding bonding arms. The empty potential adatom position is indicated by an empty circle in the triangle of adjacent rest atoms. The grid indicates the 7x 7 unit cells.

can be changed by its interaction with the surface or if the force applied to the tip is too high. For materials with many ledges that can potentially damage the tip, a different method of imaging is used. Instead of directly sweeping the tip across the surface, the tip oscillates up and down on a cantilever. By measuring the resonant frequency of the tip at every point along the surface, it is possible to reconstruct an image of the surface.

While it was designed to study the surface of materials, the STM has proven to be a very versatile tool. It has been used for many applications, including using the tip to arrange atoms into patterns on the surface of a material, to measure the lifetimes of surface electrons, and to image the electronic orbitals of molecules[13]. The STM has also been used to move atoms on a surface and form artificial structures. In theory, an STM can even be used to build a material atom by atom [14].

1.4. Defect Identification.

In STEM images, defects within atomic columns change the intensity of the atomic column. This change is because the intensity of the atomic columns is roughly proportional to the square of the sum of the atomic numbers Z of all of the atoms in the atomic column. Any change in the type or number of atoms in the column changes the Z-squared and hence the intensity (Figure 1-6). This change can be detected by measuring the intensity at the center of all of the atomic columns. Any atomic columns with an intensity that differs from the rest potentially contains a defect. When an atomic column has multiple vacancies or dopants with a greatly different Z, the change in intensity is easy to detect. This method suffers from the problem that

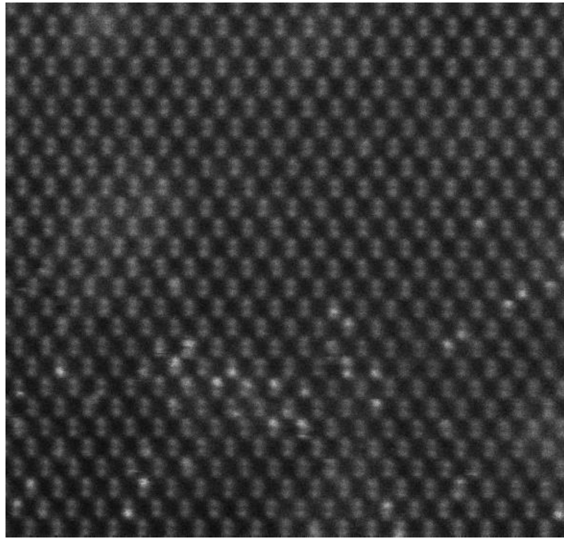

Figure 1-6 : STEM Image of Silicon (Z =14). Atomic columns with much higher intensity contain Bismuth (Z = 83) dopants. The image also demonstrates that surface contamination can alter the intensity of atomic columns that do not contain any defects.

Image courtesy of Andrew R. Lupini and Bethany M Hudak

the intensity change from surface contamination and noise can be much greater than the intensity change caused by defects. For a single defect, this intensity change might not be detectable due to the small percent change in the intensity of the atomic column.

Honggyu Kim and Jie Feng [15], [16] showed that in an atomic column, defects cannot only be detected by the change in the intensity of the atomic column but also by the effect they cause on neighboring atomic columns (Figure 1-7). By very carefully measuring the distance between neighboring atomic columns in the lattice, it is possible to detect the presence of a single defect in an atomic column. Along with being able to detect the presence of the defect, it is possible to use the magnitude of the movement of neighboring atomic columns to determine the

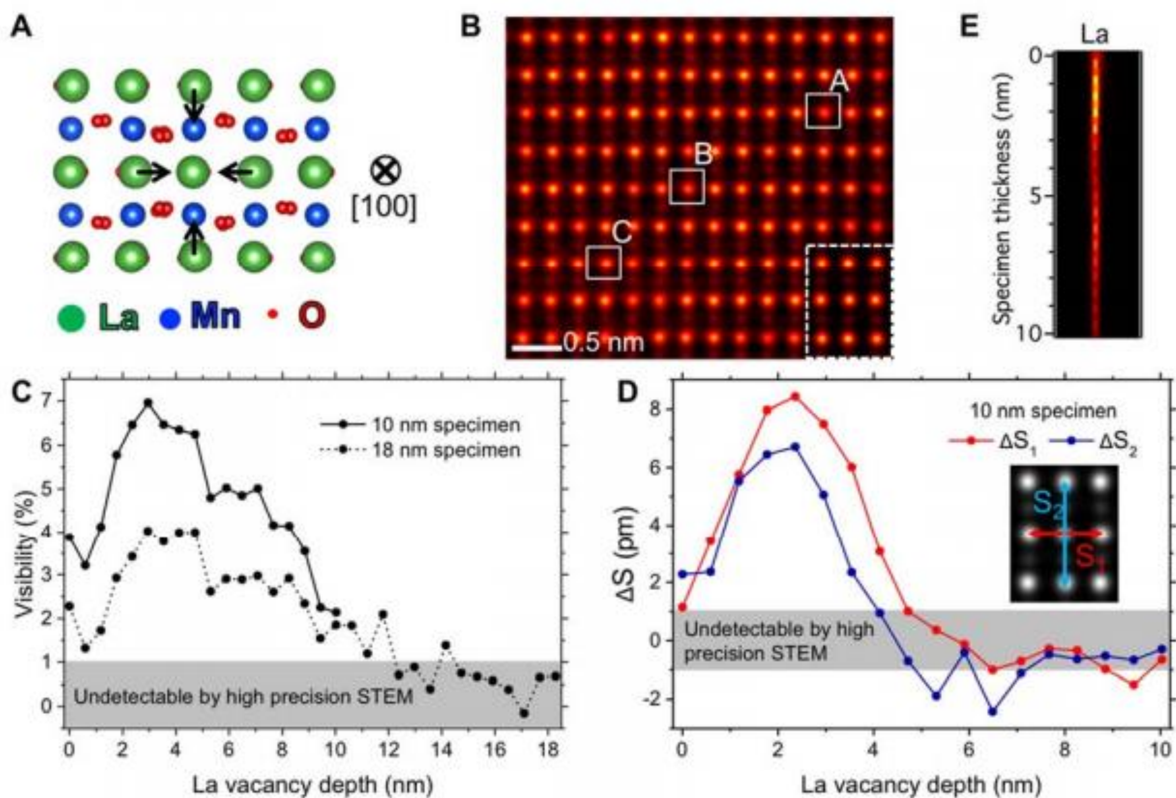

Figure and caption taken from: Jie Feng, Alexander V. Kvit, Chenyu Zhang, Jason Hoffman, Anand Bhattacharya, Dane Morgan, P. M. V. Imaging of Single La Vacancies in $LaMnO_3$.

Figure 1-7 STEM image of $LaMnO_3$ and simulations of visibility and atomic column displacements. (A) Relaxed crystal structure of $LaMnO_3$ by DFT. (B) Experimental HAADF STEM image on a 6.5 nm thick specimen along [100], with matching simulated image (white box) and projected $LaMnO_3$ unit cell. Displacements of the La-O columns due to a single La vacancy in the center are shown schematically by white arrows. (C) Simulated visibility of a single La vacancy in a 10 nm (solid) and 18 nm (dash) thick $LaMnO_3$ specimen. (D) Simulated ΔS caused by a single La vacancy in a 10 nm thick specimen. Definitions of S1 and S2 are shown in the inserted HAADF image. (E) Simulated intensity of the electron probe as it propagates along a La-O [100] column. Yellow is higher intensity.

depth of the defect within the atomic column. These deviations from the ideal lattice can be easily measured and are not affected by surface contamination and noise. The largest downside of this method is that it can only detect defects relatively close to the surface of the imaging plane, in contrast to using the change in the intensity of the atomic column.

Alex Belianinov [17] showed that by mapping the location of the neighboring atomic columns it is possible to detect defects. Unlike Honggyu Kim and Jie Feng, who measured the average distance between neighbors and looked at deviations from this to find defects, Belianinov took the relative positions of the neighboring atoms and then applied Principal Component Analysis (PCA) to map the ideal lattice and statistical deviations from the lattice. This method visualizes even tiny changes in the atomic column positions caused by defects, grain boundaries, and other perturbations that affect the lattice. This approach requires interpretations of the outputs to find and determine the source defect. This method requires the user to manually assign a number of neighbors beforehand (Figure 1-8). On top of this, multiple defects can be represented in the same output map, potentially hiding them.

Defects that can be detected with the human eye, such as defects in 2D materials, tend to be identified using a human user. As the human eye and mind are very well suited to this task, it is often easiest to rely on the user to identity and classify the defect in question. Some methods exist to automate this procedure. These include overlaying the ideal lattice onto the image. This ideal lattice can be manually generated, as was explained in atomic-column identification, or automatically by using the Fourier transform of the positions to find the lattice vectors. Another method is through tracing of the lattice vector as shown by Zheng Liu[18]. An initial atomic

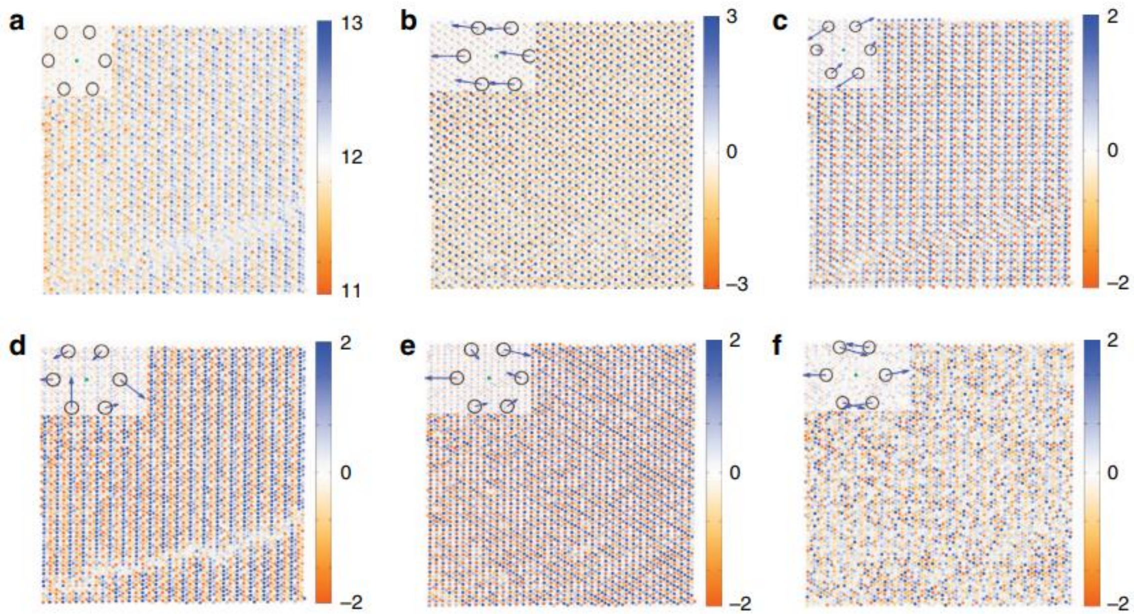

Figure and caption taken from : Belianinov, A. et al. Identification of phases, symmetries and defects through local crystallography. Nat. Commun. 6, (2015).

Figure 1-8: (a) First eigenvector, in the upper left, and a corresponding loading map. (b) Second eigenvector, in the upper left, and a corresponding loading map. (c) Third eigenvector, in the upper left, and a corresponding loading map. (d) Fourth eigenvector, in the upper left, and a corresponding loading map. (e) Fifth eigenvector, in the upper left, and a corresponding loading map. (f) Sixth eigenvector, in the upper left, and a corresponding loading map.

column is chosen and a lattice vector is given. From the location of the first atomic column, the lattice vector is followed and an atomic column is searched for near the predicted location. If an atomic column is found at this location then no defect is thought to exist (Figure 1-9). The lattice vector is updated to be the path between these two atomic columns, and the process is repeated.

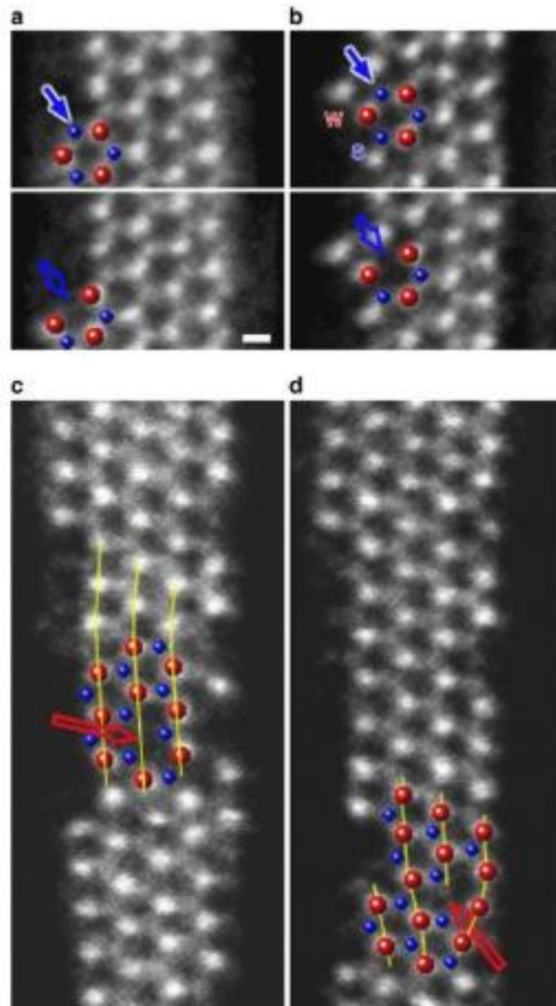

Figure and caption taken from: Liu, Z. et al. Identification of active atomic defects in a monolayer tungsten disulphide nanoribbon. Nat. Commun. 2, (2011).

Figure 1-9: (a, b) Dynamic process of the atomic defects created at S sites. (a) Edge-atom loss of S and (b) vacancy creation. (c) Simple W monovacancy and (d) W vacancy with a large distortion in the surrounding lattice. The open blue arrows indicate the loss of S atoms and the red open arrows indicate W vacancies. Scale bar, 0.2 nm

1.5. ATOMIC COLUMN ISOLATION AND LOCALIZATION

The simplest method, and most likely the first method used for the identification of atomic columns, is through the use of a human to manually identify and mark the positions of every atom in the image. This method relies on the ability of the human user to be able to identify and accurately mark the center of an atomic column. The problem with this method is that it is extremely slow, non-reproducible, and not particularly accurate in determining the centers of the atomic columns.

The next simplest method to determine the centers of atomic columns is to overlay a grid onto the image and then refine the positions in the grid. This procedure works by selecting the edge atoms of the grid and figuring out how many atoms are along each edge of the lattice. A lattice is constructed and overlaid on to the image using this information. In general, the atomic column positions in this lattice are not perfectly set on top of the centers of the atomic columns and require some form of position refinement. This refinement process can run into difficulties if the center of the atomic column is greatly different than what the lattice predicts.

A more sophisticated method to identify atomic columns is through the use of cross-correlation with a 2D Gaussian, developed by James LeBeau[19]. This scheme utilizes a sliding window, which removes a small section of the image with a set width given by[20], [21]

$$C = A_{i:i+w,j:j+w}, \qquad \textbf{(1-9)}$$

where A is the original image matrix and C is the submatrix of pixel values with a width *w* being extracted in relation to i, j from A . This window is then run through a 2D cross-correlation

$$\boldsymbol{B_{i,j}} = \frac{\sum_m \sum_n (C_{mn} - \bar{C})(G_{mn} - \bar{G})}{\sqrt{(\sum_m \sum_n (C_{mn} - \bar{C})^2)(\sum_m \sum_n (G_{mn} - G)^2)}}, \tag{1-10}$$

where G is a 2D Gaussian. This procedure is repeated for every possible pixel in the image. It is combined with filtering techniques that remove noise such as the Wiener filter[22] to produce a de-noised image. This image is then used to find the atoms through methods such as intensity thresholding of the image [23].

A further method for the identification of atomic columns in STEM images can be seen in the work of Suhas Somnath[24]. This again works through a sliding window as shown in equation **(1-9)**. However, these sliding windows are collected together and then run through a Singular-Value Decomposition (SVD) [25]. The results of the SVD are then run through the K-means clustering algorithm [26]. The results of K-means clustering are then again clustered together to find the locations of the atomic columns (Figure 1-10).

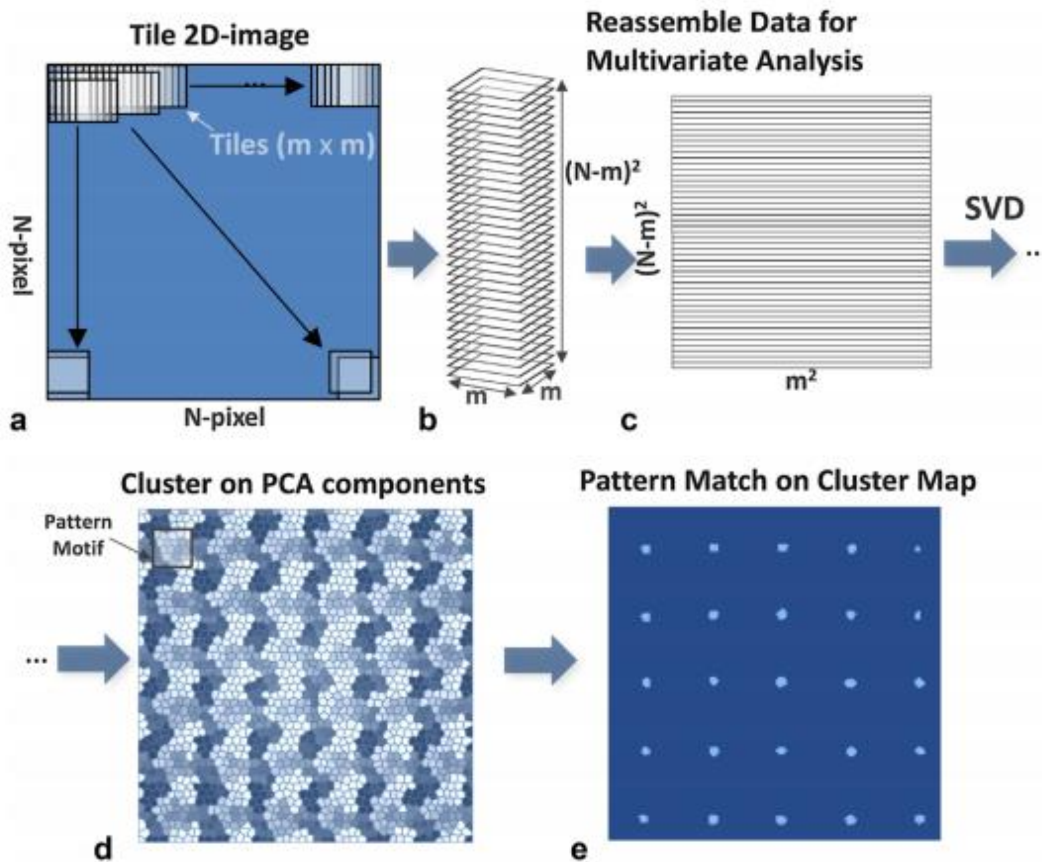

Figure and caption taken from: S. Somnath et al., "Feature extraction via similarity search: application to atom finding and denoising in electron and scanning probe microscopy imaging," Adv. Struct. Chem. Imaging, vol. 4, no. 1, p. 3, 2018.

Figure 1-10: Schematic illustrating the fundamentals of the singular value decomposition (SVD)-based image denoising technique and the pattern matching-based techniques for identifying atoms in images. (a) The denoising process starts with sliding a small window across the given image column-by-column and then row-by-row. (b) A stack of (N − m) 2 windows, each with m × m pixels, is built by copying the contents of the window at each location. (c) This 3D stack of windows is fattened to a 2D matrix by fattening the m × m pixel windows to 1D arrays with m2 elements. SVD is performed on this 2D matrix to decompose the data into the most correlated and least uncorrelated (noise) components. The image is

denoised by reconstructing the 2D matrix in c with only the most correlated SVD components and reversing the steps from c to a. (d) K-means clustering on the SVD results groups pixels exhibiting similar trends together in a cluster label map. Representative examples of repeating patterns or motifs in the label map are selected for pattern matching. (e) Each motif is compared to every section in the label map to generate a pattern matching scores' map. Each continuous-valued scores map is thresholded to generate binary maps with segments and the centroids of these segments provide the coordinates of the repeating patterns such as atoms

1.6. Image filtering

1.6.1. Gaussian filtering

The most common form of image filtering used in STEM imaging is Gaussian filtering. In STEM, Gaussian filtering refers to a matrix convolution between the image and a 2D Gaussian defined by

$$\boldsymbol{G}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2\pi\sigma^2}\boldsymbol{e}^{-\frac{x^2+y^2}{2\sigma^2}}. \qquad \textbf{(1-11)}$$

The Gaussian matrix tends to be a square matrix and is centered at 0 and extended to ±S with S typically being a multiple of σ. A matrix convolution is the process of multiplying two matrices of same dimensionality but of vastly different size. Matrix convolution works by aligning the smaller matrix in the corner of the larger matrix then sweeping the smaller matrix in each dimension,

multiplying at each possible step. For the purposes of Gaussian filtering, this operation is defined by

$$\boldsymbol{B}(\boldsymbol{M}',\boldsymbol{N}')=\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}}\sum_{\boldsymbol{j}=1}^{\boldsymbol{n}}\boldsymbol{A}(\boldsymbol{M}'+\boldsymbol{i},\boldsymbol{N}'+\boldsymbol{j})\boldsymbol{G}(\boldsymbol{i},\boldsymbol{j}) \tag{1-12}$$

where A is the original-image matrix and B is the filtered-image matrix. Due to the nature of the matrix convolution, the size of B is smaller than A by 2S in each dimension.

1.6.2. Density-based clustering

Density-based clustering uses a simple iteration between neighbors or "handshake" to cluster pixels together. In this method, a seed pixel (any pixel of value one) is chosen randomly and this pixel then "shakes hands" with all pixels of value one within its reach and they in turn attempt to "shake hands" with any pixels of value one within their reach; this process continues until no more "handshakes" are possible. At this point, all pixels that are "shaking hands" are grouped together to form an atomic column and are removed from the image, after which a new seed pixel is chosen. This procedure is then combined with a simple criterion that any cluster smaller than a few pixels is assumed to be a false positive. The use of density-based clustering and a minimum requirement for the number of pixels that would be considered an atomic column allows for the exclusion of noise and false positive pixels.

1.6.3. Hough transform

The Hough transform is a useful algorithm that was originally introduced to help identify lines in an image and has since been extended to circles, ellipses, etc. The Hough transform was first used to identify lines in images and works by defining any straight line as

$$\boldsymbol{y} = \left(-\frac{\cos\boldsymbol{\theta}}{\sin\boldsymbol{\theta}}\right)\cdot\boldsymbol{x} + \left(\frac{\boldsymbol{r}}{\sin\boldsymbol{\theta}}\right). \tag{1-13}$$

This equation defines any possible line using a unique (r,θ). The Hough transform works by creating a parameter space that spans all r and θ (Hough space). It then fills this parameter space by using the inverse of the function of a line. To find the inverse of (1-13), we simply rearrange the equation to find r as a function of θ at fixed values of x and y

$$\boldsymbol{r_H} = \boldsymbol{x_i}\cdot\cos\boldsymbol{\theta_H} + \boldsymbol{y_i}\cdot\sin\boldsymbol{\theta_H}. \tag{1-14}$$

Equation (1-14) transforms any point $(\boldsymbol{x_i}, \boldsymbol{y_i})$ in real space into a curve *r(θ)* in Hough space. The curve represents all possible lines that could contain the point $(\boldsymbol{x_i}, \boldsymbol{y_i})$ in real space. Two curves overlap in Hough space at the values of (r,θ) for the line that connects them. If three or more curves overlap at a single point this means that all these points in real space lie on the same line. To find a line, we simply take all the points in the image and overlap their curves into Hough space looking for areas of overlap. These areas of overlap in Hough space indicate that a line is present in the image with the (r,θ) where the overlap happens. Equations (1-13) and (1-14) both assume a binary image, i.e., all pixels have a value of 1 or 0. To extend these equations to work on images with continuous intensity, we simply multiply the binary line/curve by the intensity. Now instead of looking for areas where many curves overlap we are looking for areas with the highest intensity. The advantages of this method for line detection include its speed, simplicity, and repeatability. Although initially developed for a line, the Hough transform has been adapted to be able to identify other geometric shapes such as circles[27]. Many commercial circle finders such as MATLAB utilize the Hough transform to detect features in the image.

To make the Hough transform work for circles we simply define a circle centered at a,b with a radius R using

$$\left(\frac{x-a}{R}\right)^2 + \left(\frac{y-b}{R}\right)^2 = 1. \qquad (1\text{-}15)$$

Taking the inverse of equation (1-15) produces

$$\left(\frac{x_H - x_i}{R_H}\right)^2 + \left(\frac{y_H - y_i}{R_H}\right)^2 = 1, \qquad (1\text{-}16)$$

which is a circle but centered at xi, yi (points along the circle in real space) instead of a, b. In Hough space, a circle is transformed into multiple circles that overlap at the center of the original circle when the correct radius is chosen R_H =R, as illustrated in Fig. 1-11 (again, equations (1-15) and (1-16) assume a binary image, but can be similarly extended for non-binary images by multiplying the circles by the intensity in question).

A large difference between the line Hough and the circle Hough is the increase of dimensionality in Hough space from 2 to 3. This increase in dimensionality is accompanied by an increase of computational time and an even larger increase in memory usage. It is possible to extend the Hough transform for a circle to a cylinder by taking the integral of multiple circles,

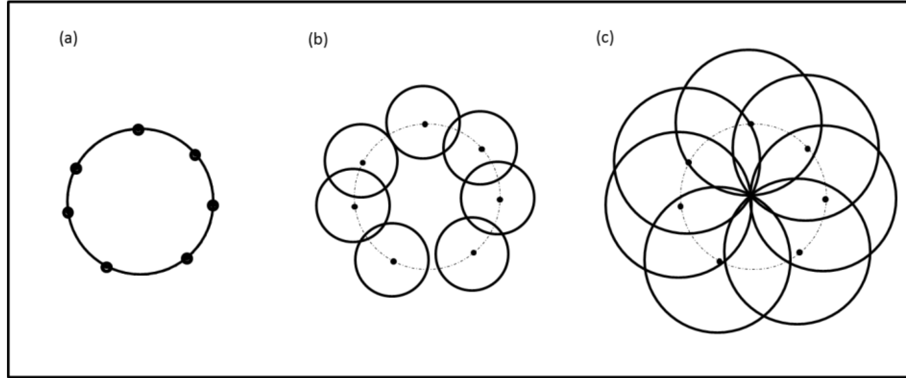

Figure 1-11: (a) 7 points are placed along the circumference of a circle with radius R in real space. (b) In Hough space, these 7 points from real space become 7 circles with RH < R. (c) the Hough-space circles of the 7 points from real space all overlap at the center of the real space circle when RH = R .

sum in finite element systems, over the range 0 to Rmax . This finite sum transforms equation (1-15) into

$$\sum_{r=0}^{R}\left(\frac{x-a}{r}\right)^2+\left(\frac{y-b}{r}\right)^2=1 \qquad \text{(1-17)}$$

and equation(1-16) into

$$\sum_{r_H=0}^{R_H}\left(\frac{x_H-x_i}{r_H}\right)^2+\left(\frac{y_H-y_i}{r_H}\right)^2=1. \qquad \text{(1-18)}$$

Since calculating Hough space for a circle stores the previous parts of the sum, not much more computation is needed to generate the Hough space of a cylinder than that of a simple circle. However, since a cylinder contains smaller solid circles inside of it, the Hough transform does not return a single point for the center but rather a large central area of the cylinder. The solid circles

method can be modified to account for any radially symmetric Point Spread Function (PSF) by adding a weighting factor to each circle of different radius, i.e., equation (1-18) becomes

$$\sum_{r_H=0}^{R_H} \boldsymbol{W(r_H)} \left(\left(\frac{x_H - x_i}{r_H} \right)^2 + \left(\frac{y_H - y_i}{r_H} \right)^2 \right) = \mathbf{1}. \tag{1-19}$$

The Gaussian Hough is approximated by

$$\sum_{r_H=0}^{R_H} \boldsymbol{R_H \sqrt{2\pi} e^{\frac{r_H^2}{2R_H^2}}} \left(\left(\frac{x_H - x_i}{r_H} \right)^2 + \left(\frac{y_H - y_i}{r_H} \right)^2 \right) = \mathbf{1}. \tag{1-20}$$

This is accomplished through using a weighting solid circle sum in which the width of the Gaussian is RH. While not a true Gaussian Hough, it is a simple approximation that meets the need to demonstrate the potential of different PSF's while not requiring the use of a higher-dimensional Hough space.

1.6.4. Principal component analysis

Principal Component Analysis (PCA) is a type of a data analysis operation that is used to decompose noisy data into orthogonal components in the order of decreasing statistical significance. This is a subtype of an orthogonalizing transformation, the purpose of which is to take a data matrix A and return a matrix of uncorrelated elements B,

$$\mathbf{TA} = \mathbf{B}. \tag{1-21}$$

The transformation matrix is usually denoted as V. These V matrices are not unique and can be found by a variety of methods. The approach specific to PCA involves finding V by setting the first component of V to be the one with the highest variance, second as the one with next highest variance and so on. Due to using variance, the first few components contain most of the

information of interest while later components contain mostly noise. The form of the transform is

$$\mathbf{AV} = \boldsymbol{U}\mathbf{E} \Rightarrow \boldsymbol{A} = \boldsymbol{U}\boldsymbol{E}\boldsymbol{V}' \Rightarrow \boldsymbol{U}'\boldsymbol{A} = (\boldsymbol{E}\boldsymbol{V}') \ , \tag{1-22}$$

where U is the basis transform while EV' represents the information in the new basis. PCA separates data into 3 parts, eigenvectors (U), loadings (V), and eigenvalues (E). By using the first few loadings that correspond to the eigenvectors of greatest importance (and thus contain the majority of data), the parameter space is transformed into {x, y, l} with l being only loadings of significantly high variance and as such much smaller dimensionally, offering a path to statistically-weighed data compression and denoising. The number of loading values needed is determined by looking at the eigenvalues as a function of the component number (so-called "scree plot") to choose the upper cutoff above which the rest of the data is most likely noise.

1.7. Thesis overview

Chapter 2 describes the development of a method for the detection of atomic columns based on Hough transforms combined with PCA and K-means clustering. This method was developed as a near-real-time method for the detection of atomic columns in STEM images. The strength of the method lies in that is requires no training and has low computational cost.

Chapter 3 demonstrates how information can be deduced by combining limited information from a STEM image with DFT. In this chapter a STEM image of bi-layer graphene is analyzed and a limited amount of atomic columns could be isolated. Combining these atomic columns with prior knowledge about the nature of graphene allow for the construction of an

idealized lattice. This lattice is then optimized using DFT and the relaxed structured used to simulate an ideal STEM image that is compared to the original. Any discrepancies in the comparison can be isolated and then possible reasons tested to identify defects in the lattice.

Chapter 4 describes the development of a method for the detection and grouping of defects in STEM images using previously found positions of atomic-column centers. This method works through the use of graph theory and requires no prior information about the material. The method is tested on both 2D and bulk materials.

2. HOUGH ATOMIC-COLUMN FINDER [1]

2.1. Introduction

The implementation of aberration correction for Scanning-Transmission-Electron-Microscopy (STEM) in the last two decades has transformed the method from a qualitative imaging tool into a quantitative structural tool capable of locating atomic columns with picometer-level precision. This development complements the intrinsic ability of STEM to collect large quantities of data in the form of images, time-resolved movies, and focal series. In the last decade, the value of such data in revealing the local atomic structure, which in turn can determine the properties of thin films, superlattices, and nanoparticles, has been repeatedly demonstrated[28]–[32].

Most of the reported experimental STEM studies of complex structures (e.g. references 1-5 above), involve extensive analysis of just a few hand-picked images, locating atomic columns with subsequent refinement of their positions, intensities, and shapes. Automated refinement processes based on methods such as center-of-mass and function fitting are routinely used[28], achieving sub-pixel localization of the center of an atomic column much more accurately and consistently than any human can achieve. On the other hand, no generic, fully-automated method exists to reliably detect atomic columns in complex structures. Currently, many of the available automated methods rely on some user input, e.g. manually overlaying a lattice onto an image[33]–[35] These methods are viable options for analysis of small batches of images, but are

[1] This chapter is based on a paper under review: O. Ovchinnikov, Y. Puzyrev, S. Jesse, A. Borisevich, S.T. Pantelides and S.V. Kalinin, 2018. Fast, fully automated detection of atomic columns in atomic-resolution microscopy images (under review at Microscopy and Microanalysis)

not capable of dealing with the entirety of the data generated by a modern STEM system. Fourier-based methods offer a fully-automated alternative[36], but they are more complex and implicitly rely on the presence of unbroken lattice periodicity, excluding multiphase systems, heterostructures, or systems with grain boundaries or other structural defects.

An alternative approach is offered by local-analysis methods, the simplest of which rely solely on the intensity of the individual pixel and neighboring pixels to detect atomic columns, allowing for easy full automation. These methods include the direct application of K-means [37], Markov chains[38], and neural networks[39]. To further improve the performance of these methods, pre-filtering can be used, which helps eliminate white noise and impose a uniform background. Such a process usually requires multiple steps, often along with asking for additional user input and approval of each step, rendering the approaches no longer fully-automated. In order to overcome this limitation, more complex methods have been developed to detect atomic columns in STEM images. Examples include the use of 2D correlation with the a reference atomic column; an image is then created where high correlation represents an atomic column and low correlation represents background; clustering of the resulting intensities completes the atom finding process[17], [33], [36].

Current automated methods focus on accuracy and precision over speed for the detection of atomic column locations. However, even the slowest program is much faster than any human user could do manually. To achieve higher degrees of accuracy and precision many of these methods[17], [33] choose to rely on user input. Even without user input these current methods[17], [33] do not come close to the speed of data collection and are, therefore, relegated

to post-collection analysis of data. This limitation of methods that rely on local area operations such as 2D correlation is the necessary consequence for such a complex operation to be performed for each individual pixel[17], resulting in a large computational cost. As a result, a bottleneck is created, causing the need to select by hand which data to analyze. This need for human interaction causes many unintended consequences, including the introduction of inherent bias into the process of data analysis along with the accumulation of large amounts of un-analyzed data, which highlights the need for an acquisition-speed, fully-automated solution.

In this Chapter we describe the development of a method that eschews the computationally expensive pixel-level operations, replacing them with a quick, whole-image Hough transform[40], [41] coupled with K-means clustering[26]. The Hough transform projects an image into a parameter space that reveals the prevalence of a certain shape in the image (Figure 2-1). K-means clustering treats inputs as positions in space and minimizes the average variation in distance within an individual cluster while maximizing the difference between

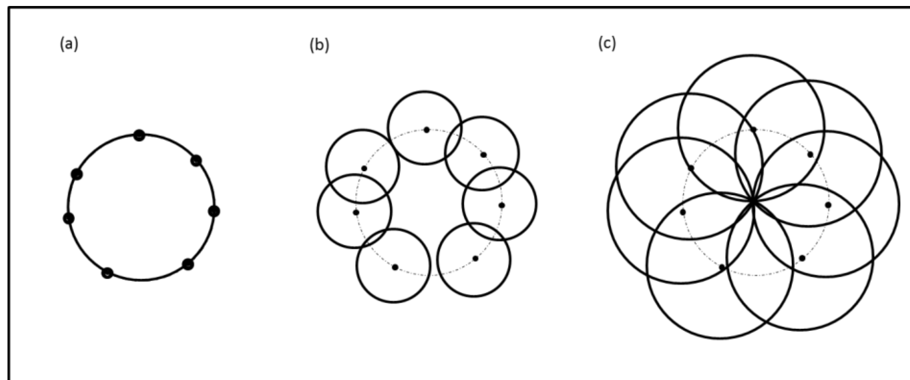

Figure 2-1: (a) Circle in real space with radius R contenting 7 points in real space. (b) Hough space circles of the 7 points from real space with a RH < R . (c) Hough space circles of the 7 points from real space with a RH = R .

clusters. The two algorithms have great synergy, with the Hough transform mapping the areas of high function prevalence, allowing K-means to cluster pixels into objects within the image space. The resulting method proves to be highly tunable for the detection of atomic columns in STEM images: By adjusting the function chosen for the Hough transform, it is possible to tune the method for different types of materials, allowing for the detection of atomic columns under any conditions. The new method is computationally inexpensive and has the potential to keep pace with image collection.

2.2. METHODS

Many current atomic column detection methods use some type of pixel-level enhancement operation that has the ability to produce excellent contrast at the price of computational cost. These methods often tend to trade speed for precision and accuracy, making them ideal for post-collection analysis. To allow for on-instrument, real-time atomic-column detection, a whole-image transform will be used here, leading to low computational cost, but higher memory usage. The Hough transform[40], [41] is a whole-image transform that has been extensively used for the detection of straight lines, but it can be extended to any functionalizable object.The Hough transform is widely implemented in commercial software such as Matlab, which utilizes the Hough transform for circle detection.

In STEM images of zone-axis axis-aligned crystals, atomic columns can be represented as delta functions convoluted with a Point Spread Function (PSF) of the electron beam. It is this PSF that we will be trying to encapsulate in the Hough transform approach. Here, we approximate the PSF with a parameterized function that captures the essential features of the PSF, i.e.

centered at an atomic column and well-localized. As a simple approximation of the true atomic column PSF, we chose to use a cylinder/ Gaussian in the Hough transform. The Hough transform changes the parameter space of the image from {x, y} to {x, y, r}, where r represents the radius of the PSF. The r component of this space can be referred to as the Hough response and is a representation of how well the PSF of that radius matches the surrounding space. The intensity and shape of the Hough response distinguishes a pixel that is part of an atomic column from one that is part of the background. One effective method of improving atomic-column detection is through pre-filtering. We found that multiplying the initial image by the local gradient is a computationally inexpensive pre-filtering method that improved the atom finding results via Hough transform (see Appendix B).

To untangle the information stored in the components and reduce the number of parameters that K-means clustering is performed on, Principal Component Analysis (PCA) is applied to the output of the Hough transform. PCA remaps the components to a new basis set, allowing for the un-mixing of information and greatly enhancing the effectiveness of K-means clustering (see appendix B). By using only the first few loadings corresponding to the eigenvectors of greatest variance (and thus containing the majority of the information), the parameter space is greatly reduced. K-means clustering groups the chosen loadings into a specified number of clusters, returning the {x, y, r} parameter space back to the original parameter space of {x, y}. However, now the values at {x(i), y(j)} are cluster assignments for each pixel.

The clusters with the highest average intensity are chosen to represent atomic columns while the rest are assumed to contain mainly background. This choice allows for the creation of a binary image where pixels are either part of an atomic column and have the value of 1 or are background and have a value of 0, representing a conversion of the original STEM image into a binary image of atomic columns and background. Density-based clustering[42] is then used to ensure that a critical minimum size of an atomic column is achieved before the group of pixels is deemed to represent an atomic column This step eliminates columns containing only a single pixel or small numbers of pixels (which may be the result of noise) from being labelled as an atomic column.

The above method can be further optimized for the detection of atomic columns in any image via the modification of the PSF. In this paper we have only used a simple zeroth- and first-order approximation of the PSF, but by better approximating the PSF, it may be possible to further improve the whole image transform to distinguish between atomic columns and background. By finding a PSF that more closely resembles the atomic columns, the performance would improve.

2.3. Results

In order to test the method, a synthetic image was generated containing 9 by 9 atomic columns modelled by symmetric Gaussian distributions (Figure 2-2a). The size of 9 by 9 was selected to have enough atomic columns to be statistically significant, while still being small enough to allow for the method to be run many times at the same conditions for reproducibility testing. To establish the performance of the algorithm, we explored the effects of random noise,

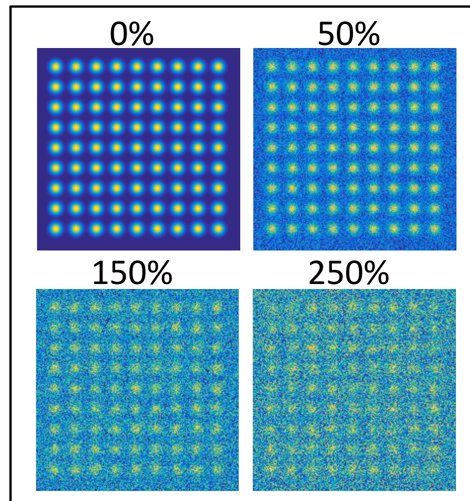

Figure 2-2: Examples of test data with different levels of noise ranging from 0% to 250% of the max amplitude of the atomic column intensity.

variation in radius, deviation from a perfect lattice, and more These tests were performed using both a circle and a Gaussian as the PSF. In all cases, the performance of the algorithm was ascertained by defining the number of atomic columns detected correctly, atomic columns not found, and false positives. Once the performance was tested on a model test case. It was then run on real data to show its ability to detect atomic columns in a real world environment.

2.3.1. How the method handles image noise:

To test the effect of the image noise on the atom finding, various degrees of random (Gaussian?) noise were added to the test image, in the range of 0 to 300% of the maximum value in the original image, which is the amplitude at the center of a Gaussian (Figure 2-2). Using a circle as the Hough transform kernel allowed us to detect all atomic columns at up to 100% noise. With the Gaussian PSF, performed better at noise greater than 150% but worse between 100 and

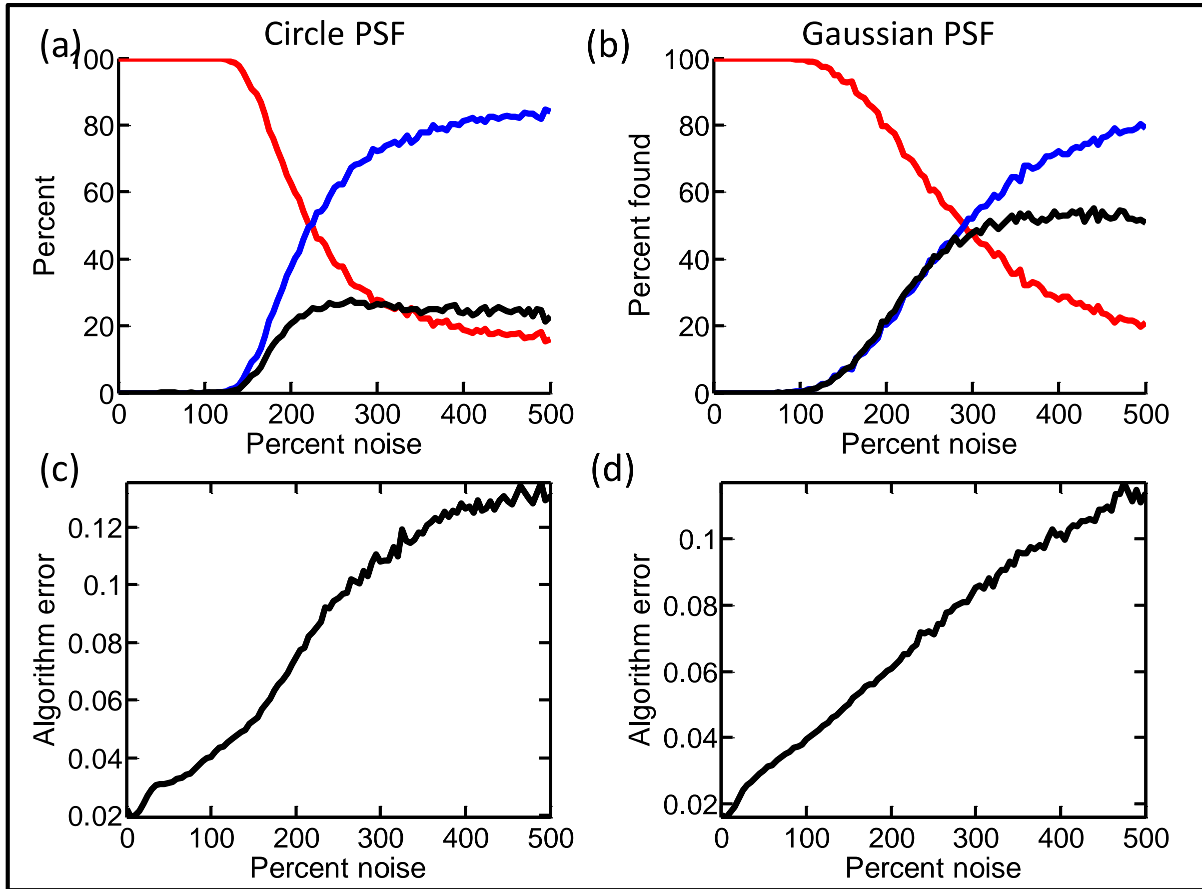

Figure 2-3 : Performance of atom finder on test data with different levels of noise, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice.

150%. The corresponding images are shown in Figure 2-3, suggesting that the algorithm performs at a level comparable to human perception sensitivity.

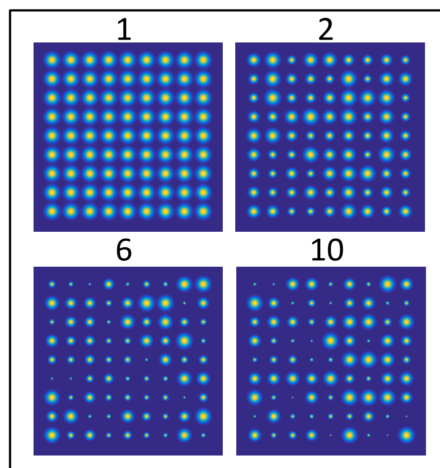

Figure 2-4: Examples of test data with different atomic column width, where N is the ratio of largest width divided by smallest width in test image.

2.3.2. How the method handles atomic columns of different size:

To test the effect of different atom sizes, the minimum radius was adjusted with respect to the maximum (Figure 2-4). In this test, it was seen that the method performed perfectly up to a difference of five times for the Gaussian PSF (Figure 2-5). This amount of size difference is much larger than anything that would be seen in most real imaging and is always larger than any variation in a single type of atomic column. As such, these methods can be thought of as able to handle any real data when it comes to difference in size of atomic columns.

2.3.3. How the method handles atomic columns of different intensities:

To test how well this algorithm handles variations in amplitude of the atomic columns, the amplitude was adjusted randomly from a minimum of 1 to a maximum of 1+x (Figure 2-6). The circle PSF method showed a decrease in accuracy starting from x=.5, while the Gaussian PSF

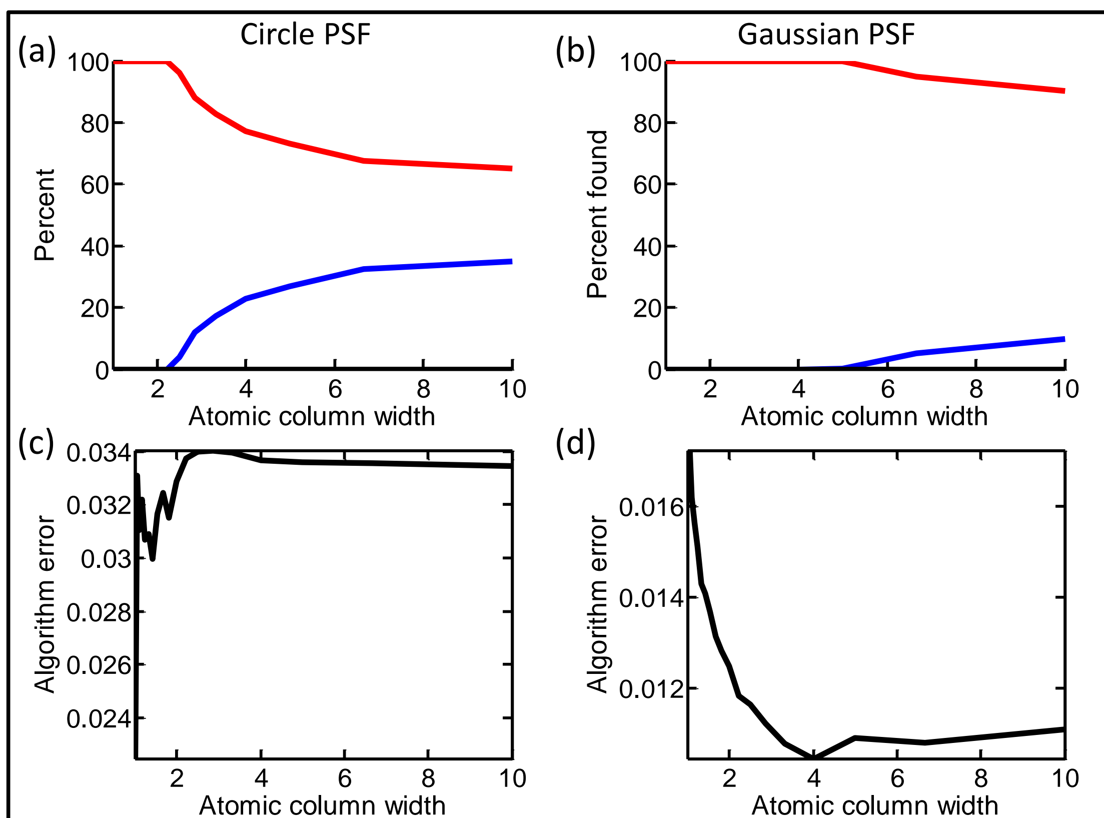

Figure 2-5: Performance of atom finder on test data with different atomic column width, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice.

showed a decrease in accuracy starting from x=1.7 (Figure 2-7). The use of a Gaussian Hough kernel usually yields better results, with the ability to perfectly detect atomic columns with half the amplitude as it better matches the shape of the atomic columns. This method would be very capable of handling multiple atom types in an image. It should be noted that intensity variation as implemented in this test approximates multiple atom types, namely abrupt changes in

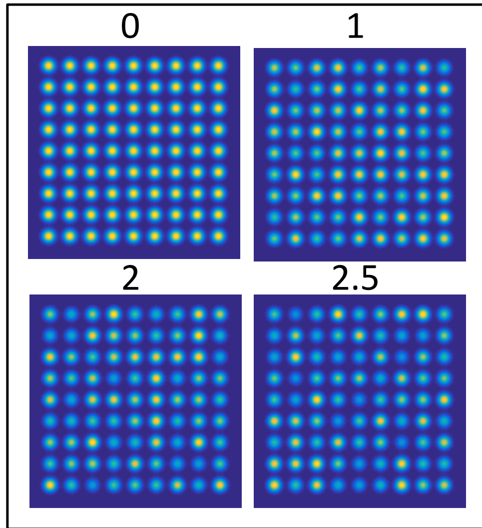

Figure 2-6: Examples of test data with different atomic column intensity, where N is 1 minus the ratio of strongest intensity atomic column divided by weakest intensity atomic column in the test image.

intensity from column to column. The detection performance might be different if the change in intensity is more gradual as related to e.g. thickness variation.

2.3.4. Testing the method for STEM images

We demonstrate this method on Mo–V–M–O (M=Nb, Ta, Te and/or Sb), a 3D material which has been studied as a promising catalyst for industrially reactions[17], [43]. This material was chosen for having atomic columns with good uniformity in intensity and size, while having many large-scale defects, which makes atomic column-detection through simple methods such as overlaying a grid all but impossible. A small section of a few STEM images was used which highlights these properties and allows for human check of the method performance. The method was able to detect every atomic column in this set of images and had no problem with defects as it does not rely on detecting any pattern in the configuration of atomic columns (Figure 2-8).

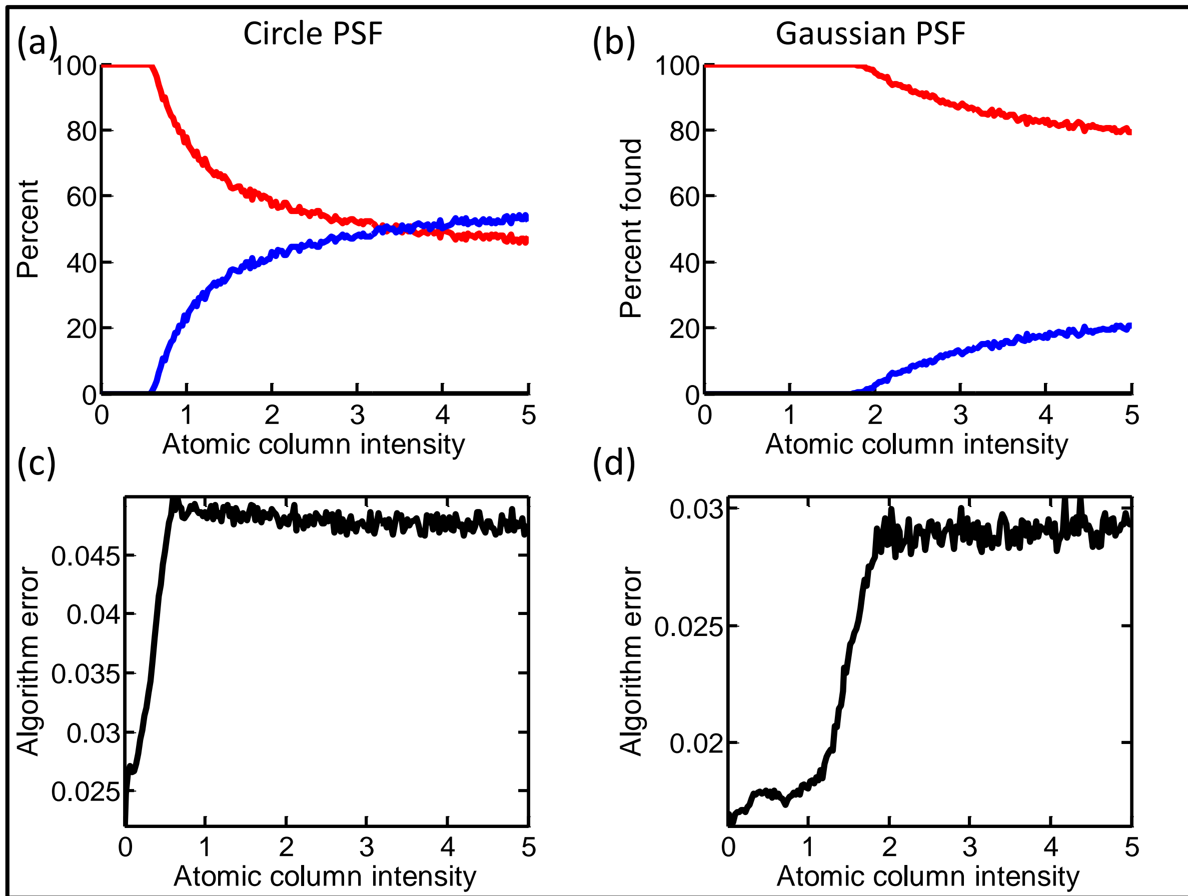

Figure 2-7: Performance of atom finder on test data with different atomic column intensity, (a)(b) graphs show the percentage of atomic columns identified (red line) and missed (blue line) along with false positives (black line); (c)(d) graphs show the accuracy of the found atomic column positions with algorithm error defined as the average deviation from the perfect lattice.

2.4. Discussion

The method presented in this chapter represents a step in the direction of fully-automated image segmentation and feature identification towards data acquisition speeds. This is accomplished by moving away from costly operations such as 2D correlation to more

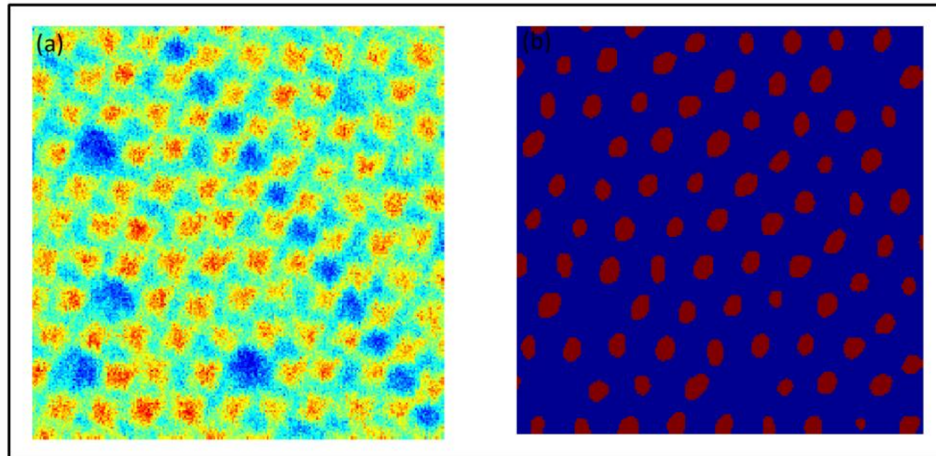

Figure 2-8: (a) STEM Image of Mo–V–M–O and (b) binary image of atomic columns and background generated by the algorithm using Gaussian PSF.

computationally efficient whole-image transforms. Matching analysis and acquisition speeds in the finding of atomic columns is the first step in STEM imaging for further, more complex analysis methods such as image segmentation and feature identification. It has been shown that this method is capable of detecting atomic columns under a wide range of conditions in a short amount of time.

While the performance of the algorithms developed here remains to be evaluated on a wider variety of real data, the results with respect to noise and intensity variance are encouraging. The detection accuracy can additionally be boosted by adding a pre-filtering step and tailoring the PSF for the atom type. The strength of the algorithm lies in the ability of the Hough transform to create a highly distinguishing characteristic between pixels, combined with the ability of PCA to transform this into just a few orthogonal parameters for fast clustering

through K-means. Due to rapid operation and high accuracy, this method can be deployed to analyze big batches of data without human input.

Testing has shown that this method features an improvement in speed of ~5 times over 2D correlation using a Gaussian as the PSF with the same radius. When compared to the speed of data acquisition, it still lags behind. However, through optimizations, it might be able to approach the speed of data collection. On top of optimization, it is also possible to pre-generate eigenvectors, saving a large percentage of the time by replacing the PCA step with a simple matrix multiplication. Eigenvectors can be reused from the first image in a collection or taken from a library of pre-generated Eigenvectors. With a combination of optimization and substitution of the PCA with a matrix multiplication step, it is likely to be possible to approach the speed of data collection.

2.5. Conclusions

The detection of atomic columns in STEM is an important task that is needed to extract physical information from data. Many scientists employ methods that require a large amount of human interaction, which creates a bottleneck and introduces human bias into the process. While many automated methods exist, none are capable of accomplishing the task under all conditions. The present algorithm is a flexible method for the detection of atomic columns in STEM images. While the algorithm does have its limitations, it also has the benefit of being relatively quick and, with suitable PSF, it is possible to tune the method to analyze almost any set of data. By having the potential to allow for real-time feedback to microscopists, the approach

can help optimize how and in what areas the data is collected, therefore changing the data collection from serendipitous to directed approach and greatly improving efficiency.

3. ATOMIC-COLUMN RECOVERY AND DEFECT IDENTIFICATION AIDED BY DFT[2]

3.1. Introduction

Graphene has been hailed and studied for its wide range of potential electronic and mechanical applications. Rippling has been seen at a variety of length scales in graphene systems [44]–[49], changes their electronic, magnetic, and other properties [44] and can induce a pseudo–magnetic field [45]. Bilayer graphene has also been extensively studied for a multitude of applications such as diodes, transistors, optoelectronics, and superconductivity [50]–[57]. For some of these applications, the tunability of its band gap is a valuable feature. Angle-mismatch between the layers of bilayer materials is common and leads to the formation of a moiré pattern. Moiré patterns in two-dimensional (2D) materials often obscure vital information when imaged, such as the underlying layer in Scanning Tunneling Microscopy (STM) and the localization of individual atomic “columns” in Scanning Transmission Electron Microscopy (STEM). Yet, obscured information can play a major role in determining the properties of the material, especially the presence of defects, interstitial dopants between the layers, and rippling, which affect both electronic and magnetic properties [44], [45].

Aberration-corrected STEM is a quantitative tool that is capable of locating atomic columns in crystals with picometer-level precision. The ability to achieve sub-pixel precision in the location of the center of an atom or atomic column in STEM images has been demonstrated through the use of center-of-mass and 2D function fitting, which in turn has been used to help

[2] This chapter is based on a published paper: O. Ovchinnikov, A. O'Hara, R.J. Nicholl, J.A. Hachtel, K. Bolotin, A. Lupini, S. Jesse, A.P. Baddorf, S.V. Kalinin, A. Borisevich and S.T. Pantelides,. “Theory-assisted determination of nano-rippling and impurities in atomic resolution images of angle-mismatched bilayer graphene”, *2D Materials* vol. 5, no. 4, p. 041008, (2018).

determine the properties of complex structures such as thin films, superlattices, and nanoparticles [29]–[32], [58]. While STEM possesses picometer-scale resolution in the in-image-plane (xy plane), the resolution perpendicular to the image-plane (z direction) is approximately a nanometer, which limits the instrument's ability to study rippling in 2D materials. In many important cases, precise determination of the xy coordinates of all atoms or atomic "columns" is not possible, e.g., in bilayers in which an angle mismatch between the layers exists and leads to moiré patterns. In such cases, due to apparent overlaps of atoms in the projected image, it is impossible to determine where one atom ends and another begins. Nevertheless, even in these materials systems, isolated areas exist that have clear and discernible atomic columns where it is possible to locate a subset of atomic coordinates with high precision.

In this chapter, we focus on nanoscale rippling in angle-mismatched bilayer graphene (the two layers are rotated with respect to each other) and demonstrate a reconstructive approach that allows us to determine the xyz coordinates of all atoms by using the xy coordinates of only the limited number of discernible atoms in the bilayer image and performing suitable density-functional-theory (DFT) calculations. By combining the information obtained from the distances between moiré nodes (from here on referred to simply as nodes) and those atomic positions that can be accurately extracted from the STEM image, a defect-free reference patch is constructed. DFT is then used to optimize the atomic coordinates in the reference patch under the constraint of the known xy atomic coordinates. By simulating the STEM image using the optimized reference patch, a local-area correlation is done with the experimental image. Areas of low correlation are then attributed to the presence of defects. The atom(s) in areas of low correlation can then be replaced by likely defect candidates and the local-area correlation is repeated until the defect

causing the low correlation is identified. In the present case, only a single atomic site exhibits low correlation and we are able to identify it as substitutional nitrogen. The z information for each layer can further be used to quantify the rippling in the sample. In the case at hand, the predominant source of rippling is determined to be caused by strain in the sample. Furthermore, the distance between the two layers can be used to determine the moiré-pattern-induced undulations, which are in agreement with those previously predicted by theory [59]. The reconstructive analysis of bilayers, which is accomplished by combining STEM with DFT to determine atomic coordinates in three dimensions, while potentially computationally expensive, allows for a complete characterization of nano-rippling in a way that is impossible from imaging alone.

3.2. Methods and materials

3.2.1. Sample preparation and data collection

High-quality graphene was grown on copper foils using atmospheric-pressure chemical vapor deposition (CVD) [60]. The graphene was subsequently transferred to TEM grids [61] following a direct transfer procedure [56]. A TEM grid was placed upside down directly on top of a copper foil coated with CVD graphene forming a stack. A drop of isopropyl alcohol (IPA) was applied to the stack. Upon evaporation of the IPA, graphene was pulled from the copper surface and adhered strongly to the TEM grid. Excess copper was removed with CE-100 copper etchant. The remaining TEM grid with graphene was rinsed in three baths of deionized water and IPA. This procedure resulted in large coverage of residue-free monolayer graphene and, critically, some regions of bilayer graphene originating from monolayer graphene that folded during transfer.

The bilayer graphene was experimentally analyzed using annular dark field (ADF) imaging on a Nion aberration-corrected UltraSTEM 100, operated at an accelerating voltage of 60 kV [62]. Figure 3-1(a) shows an atomic-resolution ADF image (1024 x 1024 pixels) of angle-mismatched bilayer graphene. To reduce noise in the image, Figure 3-1(a) was filtered with a Gaussian blur [63] with a standard deviation of two pixels, producing Figure 3-1(b). Figure 3-1(b) shows both the atomic columns of individual carbon atoms and the moiré pattern caused by the incommensurate rotation angle.

3.2.2. Determining the angle of relative rotation

To determine the angle of relative rotation between the two layers of graphene (figure 1a-1b), a Gaussian blur with a sigma of 10 pixels was applied to the original image (figure 1c-d). From this image, it is clear that there are two different types of moiré nodes in the moiré pattern: brighter nodes that correspond to areas of highest lattice correlation (primary nodes, AA stacking) and dimmer nodes that correspond to areas of highest anti-correlation (secondary nodes, AB stacking) of the two lattices. In order to determine the angle of relative rotation between the graphene layers, the distance between the moiré nodes must be determined. The moiré nodes were isolated by first using an image threshold of 75% of the maximum intensity followed by density-based clustering [42]. Once the moiré nodes where isolated, a center-of-mass algorithm was used to find the center of each node while the average intensity of the node was used to determine if a node is primary or secondary. Using the center point of each moiré node, the distance between a primary node and its surrounding secondary nodes (Figure 3-1c),

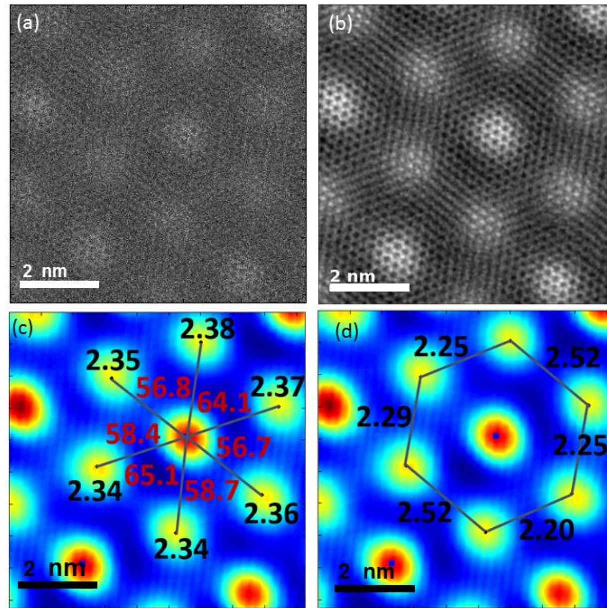

Figure 3-1: (a) ADF STEM image of a graphene bilayer with relative rotation between the layers (mismatched or twisted graphene bilayer), as acquired. (b) The same image with a small Gaussian blur more clearly shows the presence of two types of moiré nodes in the moiré pattern (primary – higher intensity, secondary – lower intensity). (c-d) The same image with a heavy Gaussian blur used to indicate the deviations from a perfect moiré pattern due to strain. (c) Distances between secondary and primary nodes (in nm: black) and angle between secondary nodes around a central primary node (in degrees: red). (d) Distances between secondary nodes (in nm: black).

the distance between the secondary nodes around the primary node (Figure 3-1d), and the angle separating the secondary nodes around the primary node (Figure 3-1c) were calculated. These distances and angles are not uniform, implying the presence of strain in the sample.

The angle $\boldsymbol{\theta}$ of relative rotation between the two layers can then be determined from the primary-to-primary moiré-node distance via [64]:

$$\boldsymbol{L(\theta) = \frac{ab}{\sqrt{a^2 + b^2 - 2ab\cos\theta}}} \tag{3-1}$$

where a and b represent the lattice constants of the two layers. For bilayer graphene, both are 2.46Å. By substituting these values into Eq. (3-1, we obtain:

$$\boldsymbol{L(\theta) = \frac{2.46}{2\sin{}^{\theta}/_{2}}} \tag{3-2}$$

Due to the variance in the moiré node distances, an average node-to-node distance was used to find the angle of relative rotation. To improve the average, not only were the primary-to-primary moiré node distances ($\boldsymbol{L_{pp}}$) used, but secondary-to-primary moiré-node distances ($\boldsymbol{L_{ps}}$) where converted into related primary-to-primary moiré-node distances using simple trigonometry:

$$\boldsymbol{L_{pp} = 2\,cos(30°)\,L_{ps} = \sqrt{3}L_{ps}}. \tag{3-3}$$

Using the average node-to-node distance, the angle of relative rotation was found to be 4.45° in the present case.

3.2.3. Filtering image and identification of individual atomic columns

In order to identify individual atomic positions, the original image was filtered to remove unwanted image collection phenomena such as noise and surface contamination. Due to the nature of the moiré interference, atomic columns can only be distinguished around the center of

a moiré node, whereby it is only necessary to filter these areas. Using the center of each moiré node, a box 201x201 pixels was removed from the image for filtering. A similar box area extraction (Figure 3-2) was used throughout this chapter and is referred to as "local area extraction".

The extracted areas around moiré nodes were filtered using a previously reported filtering technique that utilizes principal component analysis (PCA) to remove noise and surface contamination from the image[24]. This filtering technique requires slicing the image into smaller sub-images, each of which is then filtered separately using statistical correlations to remove noise and surface contamination. The full image is then reconstructed from these filtered sub-images.

For the secondary moiré nodes, this level of filtering is adequate. However, for primary moiré nodes where distinguishable atomic columns are closer together, an additional filtering

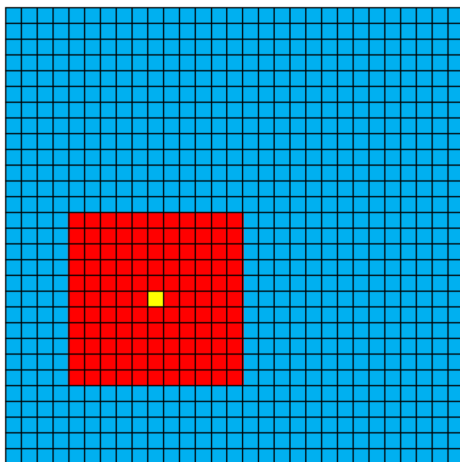

Figure 3-2: Example of local area extraction of a 11x11 area (red and yellow areas) around a center pixel (yellow area) from a 29x29 pixel image.

step was necessary. This filtering step consists of a 2D correlation with an ideal atomic column (a

2D Gaussian). At each pixel, a local area (Figure 3-2) of 21x21 pixels was extracted. These areas where passed through a 2D cross-correlation [65] defined by

$$\boldsymbol{r}(\boldsymbol{A},\boldsymbol{B}) = \frac{\sum_m \sum_n (\boldsymbol{A}_{mn} - \bar{\boldsymbol{A}})(\boldsymbol{B}_{mn} - \bar{\boldsymbol{B}})}{\sqrt{\left(\sum_m \sum_n (\boldsymbol{A}_{mn} - \bar{\boldsymbol{A}})^2\right)\left(\sum_m \sum_n (\boldsymbol{B}_{mn} - \boldsymbol{B})^2\right)}}, \tag{3-4}$$

with A being the extracted area and B a 2D Gaussian with a sigma of 13 pixels. In each node, the atomic columns were identified using two common image analysis techniques, intensity thresholding followed by density-based clustering [42].

3.2.4. Creating the initial unit cell and DFT calculations

DFT simulations of the entire image is computationally prohibitive due to the number of atoms required to make a periodic image with a small relative rotation angle. Instead, a patch was used. We considered an area of the image that encompasses two primary and two secondary nodes and constructed a corresponding idealized patch by taking two pristine graphene lattices overlaid in an AA-stacking configuration and a separation of 3.5 Å. The top lattice was then rotated to the desired angle of relative rotation (4.45°) using the center of a hexagon as the rotation point. The atomic coordinates in this idealized bilayer graphene patch are then replaced by the coordinates of the atomic columns that where determined from the experimental image. Protruding atoms that contain only a single carbon-carbon bond were removed from the edge of the patch. Edge atoms with fewer than three neighbors were passivated with hydrogen atoms.

The constructed bilayer patch consists of 899 carbon atoms and 111 hydrogen atoms, with the patch spanning 45.64 x 39.31x 3.5 Å. The patch was placed into a unit cell spanning 55x55x25 Å. DFT calculations were performed using the Vienna ab initio Simulation Package

(VASP) [66] with the Brillouin zone sampled at the Γ point only. The exchange-correlation functional was described via the Perdew-Burke-Ernzerhof parametrization [67] of the generalized gradient approximation. A van der Waals dispersion correction was included via Grimme's DFT-D3 method [68] and the DFT-D3 plus Becke-Jonson damping method [69]. The DFT-D3(BJ) correction produced results in better agreement with the experimental data as in a previous study of bilayer graphene by Lebedeva et al. [70]. The projector-augmented wave (PAW) method [71], [72] was used to describe the interaction between core and valence electrons and the plane-wave basis-set cutoff energy was set as 400 eV. During structural optimization, the x- and y-coordinates of atoms that could be determined from the experimental data were held fixed while the x- and y-coordinates of the rest of the atoms and the z-coordinates of all atoms were allowed to move until all residual forces were less than 0.01 eV/Å.

3.3. Results and discussion

To test how well theory recovered the atomic coordinates of all the atoms, the patch described above was removed from the original image and filtered in its entirety using the PCA filtering as in Section 2.3. The DFT-optimized patch (Figure 3-3a) was used to simulate the STEM image using the QSTEM simulation software [73] (Figure 3-3b). The beam parameters for the simulation were taken from those used during the experimental image acquisition, with the focus varied until the simulation was visually similar to that in the filtered patch (Figure 3-4c). To compare the simulated and experimental images, a local-area 2D cross-correlation (Eq. (3-4)) was employed between the filtered and simulated STEM images. This procedure necessitated the interpolation of the simulated image on to the same coordinate system as the experimental image. Once the two images were in the same coordinate system, a local area (Figure 3-2) of

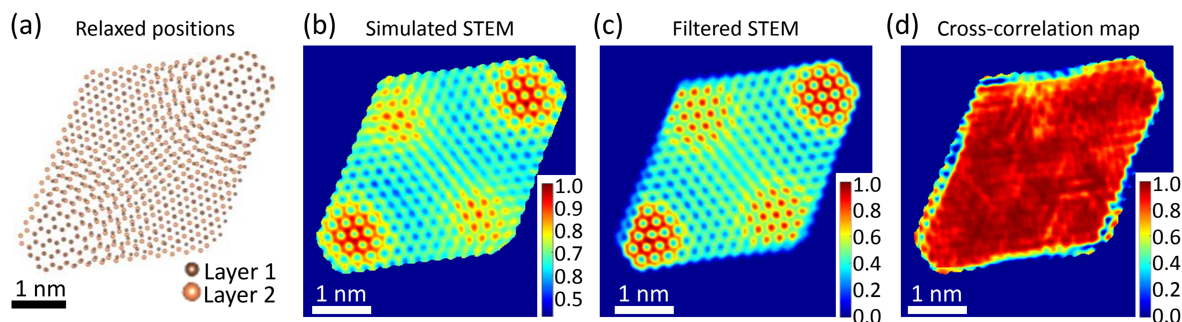

Figure 3-3: (a) Atomistic model showing the optimized positions of all carbon atoms with the two layers in different colors (b) PCA filtered simulated image of the post-DFT-optimized locations. (c) PCA filtered patch from the original image. (d) Map of local 2d correlation coefficients between the filtered simulated image and the filtered experimental image.

27x27 pixels was extracted from both images at each pixel. The correlation between simulation and experiment is very good, over 0.8 in most places, with 1.0 indicating perfect correlation (Figure 3-3d), providing confidence that the deduced atomic positions are correct or extremely close to the actual positions. Any defects in the experimental lattice should appear as areas of low correlation.

While edge effects cause areas of low correlation to occur near the edge of the patch, one significant area of low correlation just above the center of the patch stands out (Figure 3-4a). In order to determine if this area of low correlation is caused by a defect or is a limitation of the present reconstructive-analysis method, the most probable defects can be inserted into the optimized DFT patch and the patch re-optimized. First, however, the coordinates of the atom(s) that cause the low correlation must be determined by separating the 2D cross-correlation map into a cross-correlation map for each layer (Figure 3-4 b-c). This separation is done by extracting

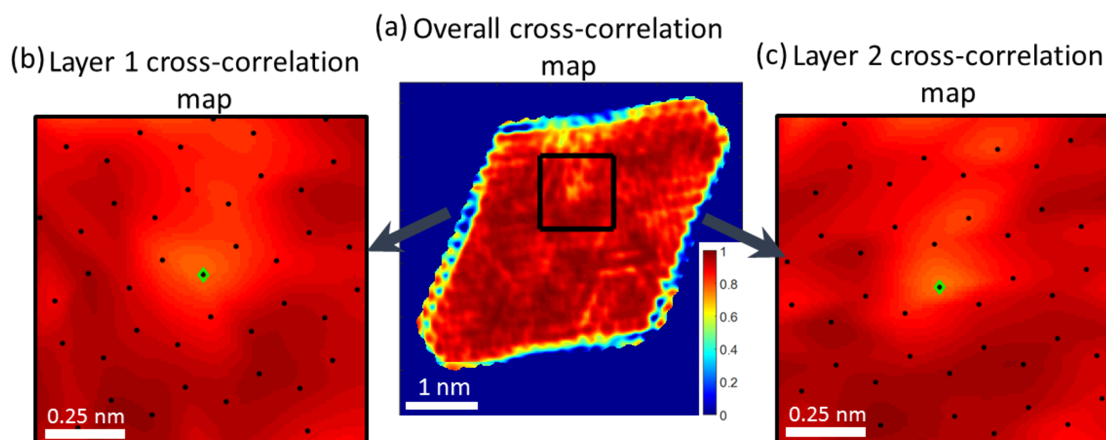

Figure 3-4: (a) 2D cross-correlation map with area of interest highlighted and separated into (b) layer 1 with atom of interest highlighted and (c) layer 2 with atom of interest highlighted.

local areas (Figure 3-2) from the 2D cross-correlation image at every atom in the lattice using a box 21x21 pixels. These local areas are then averaged and the results are interpolated onto the coordinate system of the 2D cross-correlation image. We find that the observed area of low correlation is centered in each layer on a single atom. The xy coordinates of these atoms are extremely close together, making it impossible to determine which layer contains the defect. Although it is impossible to determine which layer the potential defect is in, it is possible to narrow it to a single atom in each layer (Figure 3-4 b-c)

With the possible location of the defect causing the low correlation narrowed down to two potential atomic sites determination of the defect type was performed. A set of the most likely defects was analyzed including a carbon monovacancy and substitional boron, nitrogen, and oxygen atoms. The previously optimized DFT patch was modified and re-optimized, with each defect type tested at each of the two candidate sites. A simulated STEM image was then

generated for each patch and new 2D cross-correlation maps were calculated to determine if any of these defects improved the overall correlation (Figure 3-5a). These new 2D cross-correlation maps were compared with the original defect-free scenario visually and statistically using the percent change in cross-correlation around the area of low correlation (Figure 3-5). Based on both the visual and statistical comparisons, the low correlation is most likely caused by a substitutional nitrogen defect in layer 2.

The z information from the optimized DFT positions was interpolated onto a square lattice. The optimized z displacements show two clear ripples in both layers (Figure 3-6) in the primary-primary node direction. These ripples have a height of ~0.75 Å (using the distance of highest to lowest atom in each layer's graphene lattice) and the peak-to-peak distance of the two

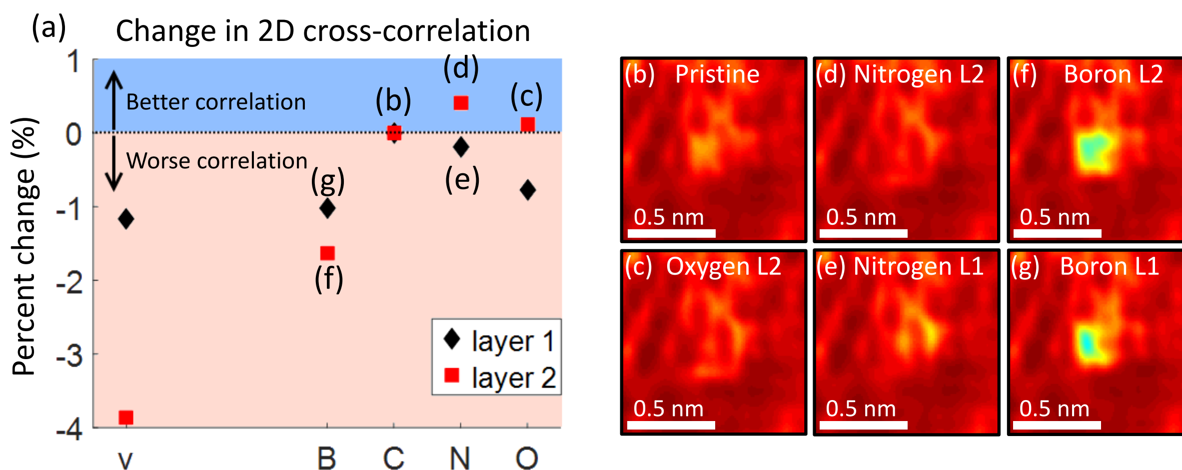

Figure 3-5: (a) Plot of the effect on the local 2D cross correlation when atoms are substituted, (b) zoomed in area of cross-correlation map for pristine bi layer graphene, (c) Carbon in layer 2 is replaced by oxygen, (d) Carbon in layer 2 is replaced by nitrogen (e) Carbon in layer 1 is replaced by nitrogen, (f) Carbon in layer 2 is replace by boron, and (g) Carbon in layer 1 is replaced by boron.

ripples is ~1.5 nm. When comparing the rippling predicted by the optimized reference patch with the experimental image, a correlation between the rippling and the background intensity can be visually observed. To determine the exact wavelength and amplitude of the rippling, the positions of the atoms were treated as a grid of points with coordinates (x,y,z). The atomic positions were then interpolated onto a square grid with a tight spacing between points, ~0.1 Å (Figure 6a-6b). Once the two layers were interpolated onto a square grid (Figure 3-6 a-b), line profiles were taken across the images. The line profiles were taken left to right along the x axis through the center of mass of the image in Figure 3-6 a and b, with 100 line profiles spread along an area between ±5 Å from the center of mass. The images were then rotated by increments of 2 degrees around the center of mass of all the image was then rotated and the line profiles again extracted until the

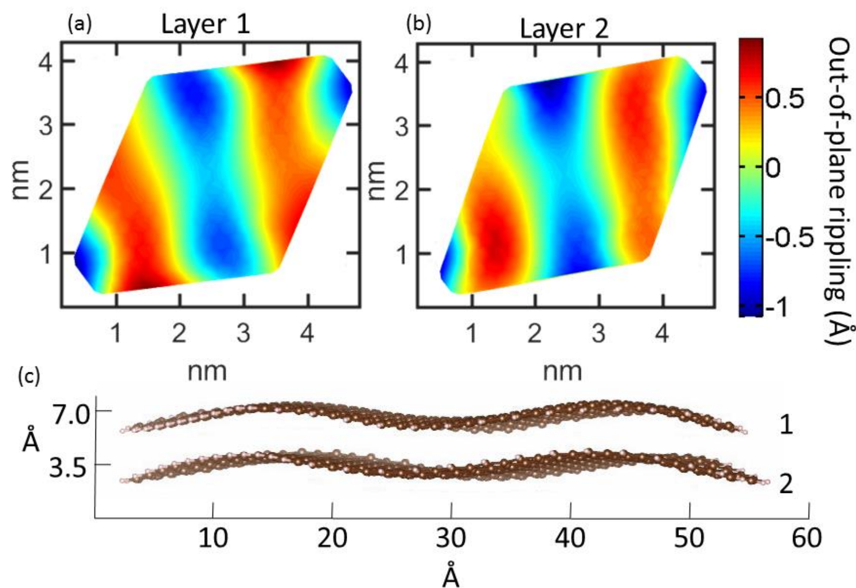

Figure 3-6. Displacement of the atoms perpendicular to the imaging-plane from the optimized positions produced by DFT for (a) the bottom layer, and (b) the top layer. (c) Atomistic model of the rippling relative to the layer separation.

average wavelength extracted from the top and bottom layer matched. Using this method the wavelength and amplitude for the layers where found to be $\boldsymbol{\lambda_1} = \mathbf{14.64} \pm 0.\mathbf{61}$Å, $\boldsymbol{A_1} = \mathbf{0.73} \pm \mathbf{0.08}$ Å, and $\boldsymbol{\lambda_2} = \mathbf{14.64} \pm \mathbf{0.54}$ Å, $\boldsymbol{A_2} = \mathbf{0.77} \pm \mathbf{0.06}$Å, giving an average wavelength and amplitude of $\boldsymbol{\lambda} = \mathbf{14.64} \pm \mathbf{0.57}$ Å and $\boldsymbol{A} = \mathbf{0.75} \pm \mathbf{0.07}$ Å.

To estimate the amount of strain, the primary-to-primary node distance and the secondary-to-secondary node distance in the patch were compared to the ideal distance based on the average angle of relative rotation in the whole image. The patch was chosen so that the primary-to-primary node distance and the secondary-to-secondary node distance are at approximately 90° to each other in the patch, allowing the strain in each direction to be directly related to the compressive and tensile strain:

$$\boldsymbol{\varepsilon} = \frac{\Delta \boldsymbol{L}}{\boldsymbol{L}} \tag{3-5}$$

$$\boldsymbol{\nu} = \left| \frac{\boldsymbol{\varepsilon_y}}{\boldsymbol{\varepsilon_x}} \right| \tag{3-6}$$

In Eq. 5 [74], L is the predicted node-to-node distance based on the average angle of relative rotation (4.45°) and ΔL is the difference between L and the measured node-to-node distance. Using Eq. (3-5), a tensile strain of $\boldsymbol{\varepsilon_x} = \mathbf{6.78}\%$ and a compressive strain of $\boldsymbol{\varepsilon_y}$ =1.23% were calculated. Then, using Eq. (3-6), an estimate of the in-plane Poisson's ratio ($\boldsymbol{\nu}$) can be determined as 0.18 for the patch. Using well-converged DFT calculations, we find that for both AA- and AB-stacked bilayers, as well as monolayer graphene, the in-plane Poisson ratio ($\boldsymbol{\nu}$) is 0.19, in agreement with experimental measurements on graphene and graphite [75], indicating that

there is practically no sensitivity to the number of layers for the in-plane Poisson's ratio in both theory and experiment.

Given the presence of strain-induced rippling in our sample, the question arises whether the observed rippling obeys continuum mechanics. For example, monolayer graphene appears to satisfy [47], [48], [76] or violate continuum mechanics [77], [78], depending on the particular geometric setup conditions [79]. The continuum mechanics model of strained elastic sheets can be written either in terms of an applied tensile strain or in terms of a compressive strain. When uniaxial tensile strain is applied to an elastic sheet of length $\boldsymbol{L}$ and thickness $\boldsymbol{t}$, the wavelength ($\boldsymbol{\lambda}$) and amplitude ($\boldsymbol{A}$) of rippling are given by [80]:

$$\boldsymbol{\lambda = (2\pi L t)^{1/2}[3(1-\nu^2)\varepsilon_x]^{-1/4}} \tag{3-7}$$

$$\boldsymbol{A = (\nu L t)^{1/2}\left[16\varepsilon_x/\left(3\pi^2(1-\nu^2)\right)\right]^{1/4}}, \tag{3-8}$$

where $\boldsymbol{\nu}$ is the in-plane Poisson's ratio and $\boldsymbol{\varepsilon}$ is the applied strain along the length. Likewise, for compressive strain of magnitude $\boldsymbol{\varepsilon_y}$, the relevant equations are [76]:

$$\boldsymbol{\lambda = (2\pi L t)^{1/2}\nu^{1/4}\left[3(1-\nu^2)\varepsilon_y\right]^{-1/4}} \tag{3-9}$$

$$\boldsymbol{A = (\nu L t)^{1/2}\left[16\varepsilon_y/\left(3\pi^2(1-\nu^2)\right)\right]^{1/4}}. \tag{3-10}$$

Equations (3-7), (3-8) and (3-9), (3-10) are fundamentally related through the in-plane Poisson's ratio for the magnitude of tensile and compressive strains ($\boldsymbol{\varepsilon_y = \nu\varepsilon_x}$). In the current experimental setup, it is difficult to concretely define $\boldsymbol{L}$. Furthermore, there is an ambiguity in

how to define the thickness of atomically thin materials. Therefore, we can use Eqs. (3-7) and (3-8) to derive a relationship between $\boldsymbol{\lambda}$ and $\boldsymbol{A}$ in terms of $\boldsymbol{\varepsilon_x}$:

$$\boldsymbol{A = \frac{\lambda}{\pi}(2\nu\varepsilon_x)^{1/2}}, \tag{3-11}$$

or Eqs. (3-9) and (3-10) to derive such a relationship in terms of $\boldsymbol{\varepsilon_y}$:

$$\boldsymbol{A = \frac{\lambda}{\pi}\left(2\varepsilon_y\right)^{1/2}}. \tag{3-12}$$

Using the average extracted wavelength of $\boldsymbol{\lambda = 14.64 \pm 0.57}$ Å with a Poisson's ratio of 0.19 in Eq. (3-12) gives an expected amplitude of $\boldsymbol{A = 0.73 \pm 0.03}$ Å. This expected amplitude and our extracted amplitude of $\boldsymbol{A = 0.75 \pm 0.07}$ Å agree within uncertainty, implying that the observed rippling in bilayer graphene is consistent with the continuum theory of elasticity. Previous works have sparked the controversy as to whether rippling in graphene does obey [81]

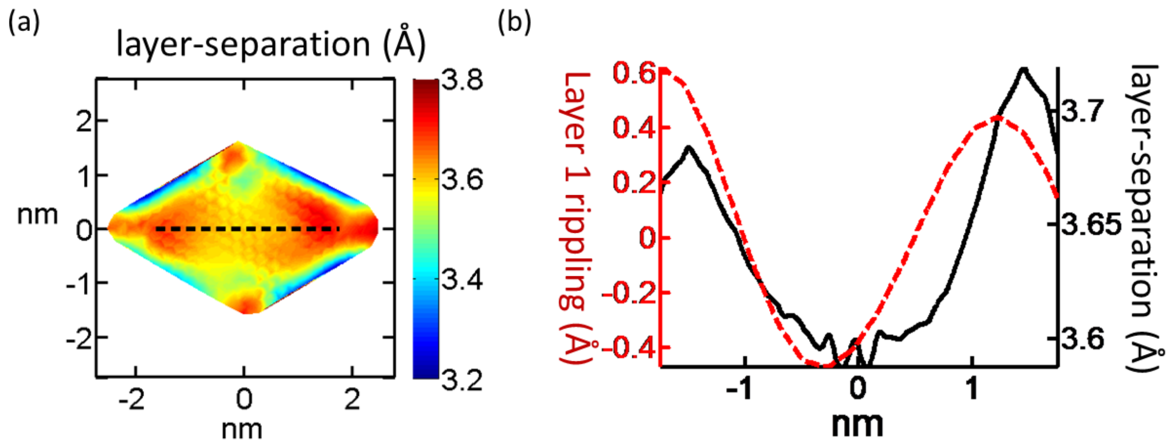

Figure 3-7:. (a) Layer-separation between the two layers from the optimized positions. (b) Layer-separation (black solid line) and *z* rippling for the top layer (red dotted line) across the line profile shown in (a).

or does not obey [77] the continuum theory model; while in the present case for bilayer graphene, the rippling does obey the continuum theory.

Undulations in the z direction in bilayer graphene with a low angle of relative rotation result in a change in the separation between the layers that coincides with the intensity change of the moiré pattern [82]. To study the z undulations caused by the angle of relative rotation between the two layers of bilayer graphene, the distance between the layers (Figure 3-6 a-b) is used (Figure 3-7 a). To measure these undulations in the current data, a series of line profiles were taken along a path between the centers of the primary moiré nodes in the patch. These line profiles show undulations with a wavelength that matches the primary-to-primary node distance and an amplitude of 0.07Å (Figure 3-7 b). Semiclassical theory [82] predicts these undulations to have an amplitude of ~0.1 Å, which is slightly larger than what is observed in our optimized unit cell. An interesting note is that the strain-based rippling along the same line profiles (Figure 7b) appears to have the same wavelength and orientation as the moiré rippling; however, the strain rippling amplitude is an order of magnitude larger. Due to the moiré undulations and the strain rippling being aligned, the moiré-induced undulations would be undetectable without the aid of DFT. This orientation match between the moiré-induced undulations and strain-based rippling might simply be a coincidence or an actual correlation.

3.4. Outlook

We have introduced a reconstructive analysis methodology to deduce the coordinates in all three spatial dimensions of the atoms that cannot resolved from one another in an experimental STEM image. The approach is not only able to deduce the xy information that is lost

in the STEM image, but also provides information in the z direction for which STEM provides little information. The z information clearly shows rippling in the direction of strain as would be expected and in line with the continuum model. The method relies on the assumption of a defect-free lattice, but, using local-area correlation, it is possible to identify the location of potential defect areas. By inserting defects into the lattice computation and then re-optimizing the lattice, it is possible to identify the defect by optimizing the local correlation map in a self-consistent manner.

While we have demonstrated success using a STEM image, the same approach can be performed using other atomic-resolution imaging techniques. Scanning tunneling microscopy (STM) could be an ideal candidate. STM can resolve individual atoms, but, like STEM, has a “blind spot” preventing full three-dimensional measurement of positions in bilayer or multi-layer systems. STM has very high resolution in the z-direction normal to the surface and good resolution in the xy plane of the surface, but is predominantly sensitive only to the outermost layer of atoms. STM provides little information about the lower layer in a bi-layer system. A number of atomically-resolved STM studies have identified rippling in both mono- and multi-layer graphene [83]–[85]. The details of the subsurface structure influences the surface layer at a level that can be observed by STM, i.e. changes in the vertical positions of atoms located above or between atoms below. The present reconstructive-analysis approach, combining data analysis of measured atomic positions with DFT modeling should identify the missing subsurface atomic structures and could resolve outstanding questions on friction and the origins of rippling.

3.5. Conclusions

We have demonstrated an approach for the recovery of atomic positions in three dimensions that cannot be determined from the raw data alone. Using a STEM image of bilayer graphene as the test case, the xy positions of all atoms were recovered and the z positions deduced. The z positions were validated through the 2D cross-correlation of the simulated STEM image with a filtered raw-data image. Defects in the bi-layer graphene sample were detected by the presence of areas of low correlation in the 2D cross-correlation map. It is then possible to use the correlation maps to identify the type of defects in a self-constant manner. In the present case, a defect was found and identified as a substitutional nitrogen. The z positions show the presence of nanometer-scale rippling consistent with the continuum elastic model. Furthermore, we find that the angle of relative rotation in the bilayer causes additional moiré-induced undulations in the atomic positions. It should be possible to further extend this methodology to other types of atomically-resolved microscopy, such as STM.

4. DEFECT DETECTION AND GROUPING [3]

4.1. Introduction

Defects can vastly alter the performance of materials so that control of defect distribution and density is an important tool in engineering materials with novel functionality. Even small concentrations of defects can often change the properties of materials so that it is important to quantify the type and concentration of defects[86]–[88]. Over the last two decades, aberration-corrected Scanning-Transmission-Electron-Microscopy (STEM) has become a quantitative structural tool capable of locating atomic columns with picometer-level precision. The ability to achieve sub-pixel precision on the location of the center of an atomic column in STEM images has been demonstrated through image analysis techniques such as finding the center of mass and 2D function fitting with a Gaussian, allowing for accurate, consistent, and repeatable determination of the centers of atomic columns in STEM images[28]–[30], [89], [90]. Within a STEM image it is possible to visually identify many defects such as impurities, interstitials, stacking faults, and a plethora of other complex defects.

Several methods exist to detect and identify defects in STEM images, each having unique benefits and limitations. Defects within atomic columns can be detected by examining deviations in the contrast , looking for deviations in the local lattice[15], [16], overlaying an ideal lattice on the image[17], and by using vector tracing[18]. These methods include measuring the distance between neighboring atoms in the lattice and then using statistics and modeling to detect the

[3] This chapter is based on a to-be-submitted paper: O. Ovchinnikov, A. O'Hara, S. Jesse, B. Hudak, S. Yang, A. Lupini, M. Chisholm, Zhou, W., A. Borisevich, S.V. Kalinin, and S.T. Pantelides, 2018. Automated defect detection in images of materials using graph theory (in preparation)

presence and depth of a single defect in atomic columns[15], [16]; measuring the relative positions of neighboring atoms and then applying Principal Component Analysis (PCA) followed by K-means clustering to map the idea lattice and statistical deviations from the lattice[17]; using the Fourier transform of the image to determine the lattice parameter and then overlaying the lattice on the atomic coordinates. Using cross-correlation between the STEM image and a simulated STEM image based on coordinates obtained by relaxing a model structure by density-functional-theory (DFT) calculations and then detecting defects through areas of low correlation[91]. These methods have achieved detection of defects that would not be possible or would be extremely time-intensive with the human eye.

In this Chapter, we describe the development of a method that applies graph theory to the positions of atomic-column centers and is capable of detecting a wide range of defects in STEM images with no prior knowledge of the material. A cycle in graph theory is a path between points that connects a point back to itself. Multiple types of cycles exist such as the simple-walk cycle that does not allow any point or connection to be repeated. For this work, a particular type of cycle is created with the following conditions: no vertices may be repeated, no connecting line may intersect another connecting line, the cycle must enclose a reference atomic column, the cycle must not enclose any other atomic columns, and, finally, the cycle must be the shortest path connecting the vertices. For every atomic column in an image, a single cycle is found to represent it. Based on the number of vertices and the area of the cycles, it is possible to detect and categorize defects in the STEM image. The approach is applied to STEM images of both 2D and bulk materials. In 2D MoS_2 doped with Re, sulfur vacancies are detected using two different

cycle metrics. In bulk silicon doped with bismuth we demonstrate the ability of cycles to detect the Bi dopants in the atomic columns and compare with Z contrast.

4.2. Methods and materials

4.2.1. Image filtering, finding centers of atomic columns

All the raw STEM data are first processed to identify the centers of atomic columns as follows. At each pixel, a subimage is defined, centered at the pixel and encompassing an area roughly equal to the area per atomic column. These subimages are filtered using Principal Component Analysis (PCA) to remove noise and surface contamination[24]. The subimages are then passed through a 2D correlation[65] with an ideal atomic column (a 2D Gaussian) defined by

$$\boldsymbol{r}(\boldsymbol{A},\boldsymbol{B}) = \frac{\sum_{\boldsymbol{m}}\sum_{\boldsymbol{n}}(A_{mn}-\bar{A})(B_{mn}-\bar{B})}{\sqrt{(\sum_m\sum_n(A_{mn}-\bar{A})^2)(\sum_m\sum_n(B_{mn}-B)^2)}} \qquad (4\text{-}1)$$

which returns a single normalized intensity. From the filtered image, the centers of atomic columns are then found using a simple intensity threshold followed by density-based clustering [42]. Any clusters that do not meet a minimum size requirement are rejected. The center of mass of each cluster is treated as the center of an atomic column. Further refinement of the positions of the atomic-column centers is preformed using nonlinear least-squares curve fitting between the raw data and a 2D Gaussian. The center of the fitted Gaussian is then treated as the refined center of the atomic column.

4.2.2. *Finding cycles for each atomic column*

A cycle is a path that connects atomic-column centers or "points" in such a way that it forms a closed loop to the original point. For our purposes, paths that define cycles are restricted further as follows. To find the cycles around an atomic column we remove the m nearest points and operate on them separately to save time. At each point, each of the m nearest points is connected to its n nearest neighbors. Typically, m=40 and n=10. Then, using the nearest neighbor (or one of equidistant nearest neighbors) as a starting point and all possible connections are followed to neighboring points without repeating a point. This process is repeated until we return to the starting point. If such a return is not possible, we conclude that no more cycles can be defined.

4.2.3. Filtering cycles

Once the cycles are found as described above, they must be filtered to find a single cycle to represent each atomic column. This filtering is done by checking every cycle to see if it meets a series of rules. These rules are that no point may be repeated in the cycle, no connecting line may intersect another connecting line, the cycle must enclose the reference atomic column, the cycle must not enclose any other atomic column, the cycle must have no smaller angle then x degrees (x was set at 45°), and finally the cycle must be the shortest path connecting the points within a small fudge factor (within 1% of shortest path) (Figure 4-1). Once all the cycles that do not meet these criteria are removed, the cycles with the largest number of points are selected.

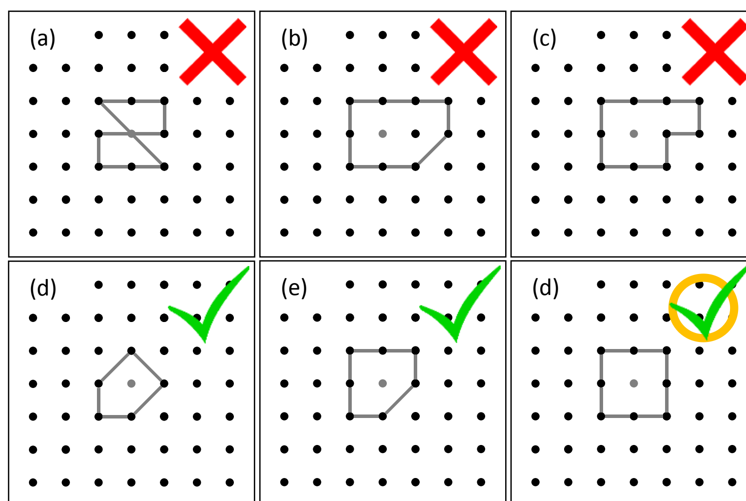

Figure 4-1: (a) Rejected cycle due to cycle lines crossing each other (b) rejected cycle due to an atom enclosed in the cycle (c) rejected cycle due to cycle is not the shortest path

From these cycles the one with the largest area is chosen as the cycle to represent an atomic column. The reason the largest cycle is used for repeatability. To insure that we choose the same cycle each time, it must have a unique feature. Since the smallest cycle would always be a triangle and provide little information, the largest cycle is used instead.

4.2.4. Constructing cycles using Delaunay triangulation

Finding all the possible cycles and checking them is a time consuming process. A faster way of finding a good guess of the best cycle was found using Delaunay triangulation [92]. For an atomic column, the positions of nearest m (typically m=40) neighbors are put into a Delaunay triangulation algorithm. Using the triangle that encloses the atomic column as the starting cycle, triangles from the Delaunay triangulation are combined with the cycle, testing at each step to make sure that the cycle meets the selection criteria, until no further triangles can be added that meet the criteria. This method of finding the cycle for an atomic column was found to be over an

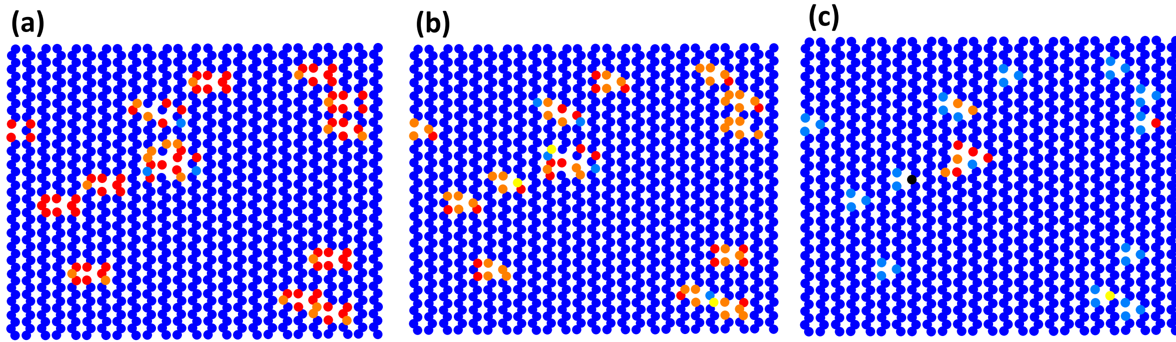

Figure 4-2:: The centers of atomic columns for MoS_2 colored based on the number of points in an the cycles associated with them for (a) searching all cycles (b) using Delaunay triangulation (c) and using Delaunay triangulation plus pre-filtering.

order of magnitude faster than searching all possible cycles. However, it does not always find the true correct cycle, though it gets close. Even so, it works well because it fails only around defects that change the lattice, which is what we are looking for. (Figure 4-2 a-b).

4.2.5. Pre-filtering of cycles

To further improve the speed of the algorithm, pre-filtering of cycles was tested. As all crystals are constructed from a combination of limited types of lattices, it is often possible to know what the correct cycle is before searching. To take advantage of this, we created a small library of possible cycles. We can overlay a library cycle onto the points by aligning it to the reference point and its nearest neighbor (or one of equidistant nearest neighbors), scaling it to fit the distance between the two points. Starting with the largest cycles in the library, the cycles are overlayed onto the points. If every point in the cycle coincides, within some uncertainty, with a point in the image, it is selected as the correct cycle. This process sometimes yields too small a cycle, but it has no effect on defect detection (Figure 4-2c). The speed improvement of pre-

filtering is based on the size of the library to test and the percent of defect atomic columns, as this is an additional operation that must be performed on defect atomic columns.

4.2.6. Clustering cycles into defects using number of points in cycle.

The first method of detecting defects is by looking at deviations in the number of points in the cycles. In 2D materials, atomic columns near defects that cause changes in the lattice such as vacancies, interstitials, or stacking faults have cycles that contain a different number of points than in the perfect crystal. To detect a defect, we mark as acceptable any atomic column that has a cycle with the same number of points in it as a cycle in the perfect crystal. The remaining cycles are clustered together using a density-based clustering algorithm[93]. This algorithm randomly selects a point as the start of a cluster and then adds every point that is with a specified radius into the cluster. This procedure is repeated until no point can be added to the cluster. These clusters are then grouped based on the number of atomic columns in the cluster. This procedure allows for the automatic detection and defect clustering in STEM data.

4.2.7. Using cycle area to find defects

Another method of using cycles to find defects is by looking at changes in the cycle's area. Using the cycle's area to look for defects allows for the detection of defects that do not change the lattice, such as interstitials and vacancies in bulk materials. Any cycle area that is much larger or smaller than the average cycle area represents the presence of a defect near that cycle. This works on a similar idea to previous works where single cation vacancies where detected by measuring changes in the distance to nearest neighbor atomic columns[16], [94].

4.3. Results

Here we demonstrate how the above methods work for the detection of defects in 2D and 3D materials.

4.3.1. 2D materials

For 2D materials, we have selected Re doped MoS_2 (Figure 4-3 a). This material was selected due to the nature of available defects, namely Re dopants at Mo sites along with single and double S vacancies. We filtered this image and found all the atomic columns using the procedure described in the methods sections 4.2.1. Using the centers of the atomic columns, the cycles are found for each atomic column more than 2.5 times the average nearest neighbor distance from the edge. The number of points in each cycle are analyzed and the defects are

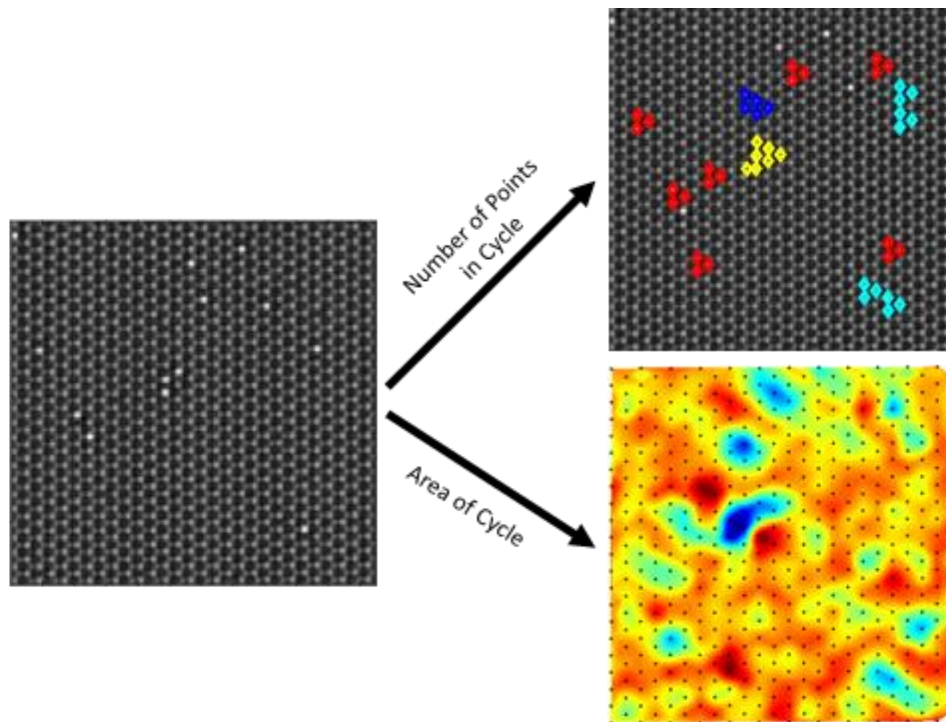

Figure 4-3: (a) STEM image of Re doped MoS_2 (b) atomic columns identified to be near defect using number of points in cycles (c) area of the cycles smaller areas are due to S vacancies

categorized (Figure 4-3 b). Using the number of points in the cycle, all the missing S columns were detected and categorized into single missing columns, two adjacent missing columns and three adjacent missing columns. To find the single S vacancies, the areas of the cycles were used (Figure 4-3 c). S vacancies cause noticeable decrease in the area of cycles. No method was found to identify the Re dopants using cycles because the presence of S vacancies near the Re makes

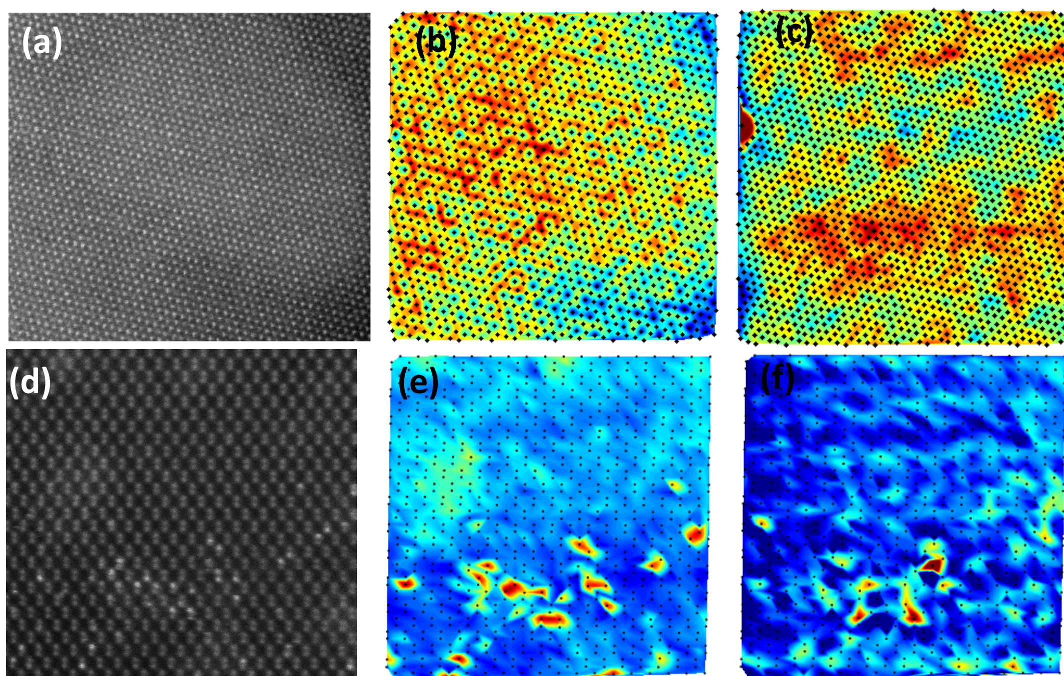

Figure 4-4: (a) raw HAADF STEM image of Mo–V–M oxide (b) interpolated map of local intensity around every atomic column in figure d with the location of atomic columns shown (c) interpolated map of size of cycle associated with every atomic column in figure d with the location of atomic columns shown (d) raw HAADF STEM image of Bi doped Si (e) location of every atomic column in figure a colored based on the local intensity around every atomic column (f) location of every atomic column in figure a colored based size of cycle associated every atomic column

the use of cycle area unreliable. A possible explanation for the presence of S vacancies near Re could be due to the nature of the defects. Re dopants are N-type defect[95], [96] while sulfur vacancies act as deep acceptor sites[97], [98]. S vacancies act as electron traps for the excess electron of the Re atom making them complementary defects.

4.3.2. 3D materials

This ability to look at defects under the surface is demonstrated in Mo–V–M–O, where the M can be one of a number of atoms[43], with Te the most likely in this sample. Mo–V–M–O is a material that has been studied as a potential catalyst and can display a variety of interesting phases and defects. In Mo–V–M–O, we can see the pooling of vacancies or Nb atoms under the surface of the material (Figure 4-4 c). The potential for large-scale vacancy clusters in this material and large stacking faults can be seen in Figure 4-5 which show an area with whole atomic columns missing. These areas are not visible in the Z-contrast image due to the presence of surface contamination on the sample masking the slight changes that might be present in the intensity of the atomic columns.

The ability of using cycle area to detect defects within an atomic column was tested using Bi-doped Si (Figure 4-4 d). This material was used for demonstration of the principle because the Z-squared difference between Bi and Si makes identifying the Bi locations very simple. The ability of cycle size to identify Bi within an atomic column was found to be worse than the reference of Z squared intensity. Using cycle size, it was only possible to identify the approximate location of roughly 80% of the Bi dopants (figure 4-5 e-f). In areas with more than one Bi dopant in close proximity, it is difficult to identify the number and exact location of the dopant. However, for

isolated Bi, it is much easier. The 80% identification is due to the depth of the defects in the material, as intensity can help identity defects at a greater depth than using the distortion in the local lattice that cycle area relies on. This result is in line with previous works that have used distortions in the local lattice to identify defects.

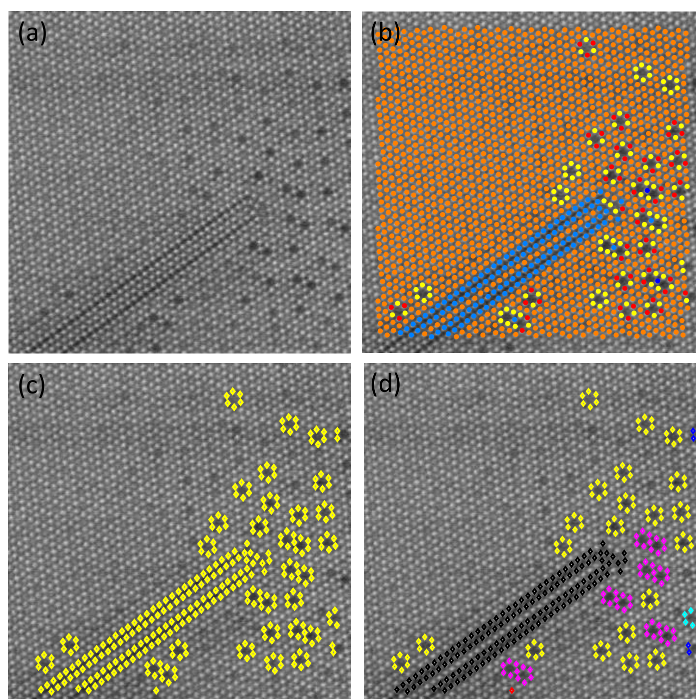

Figure 4-5: (a) Original HAADF image of Mo–V–M–O (b) atom column location colored based on number of atoms in associated cycles (c) atomic columns whose cycles are not the part of the perfect crystal (d) no perfect crystal atomic columns grouped into defects and colored based on the number of atoms in a defect (single missing column/ yellow , 2 adjoining missing atomic columns/purple, large staking fault /black)

4.4. Conclusions

We test the method on STEM data for 2D and 3D materials. The method is best suited to finding defects in 2D materials. Nevetheless, it can supply useful information about the presence of defects in 3D materials as well. In 2D materials we demonstrated the ability of the method to distinguish a number of common defects including interstitials and vacancies.

REFERENCES

[1] P. A. M. Dirac, “Note on Exchange Phenomena in the Thomas Atom,” *Math. Proc. Cambridge Philos. Soc.*, vol. 26, no. 03, p. 376, Jul. 1930.

[2] P. Hohenberg and W. Kohn, “Inhomogeneous Electron Gas,” *Phys. Rev.*, vol. 136, no. 3B, pp. B864–B871, Nov. 1964.

[3] W. Kohn and L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects,” *Phys. Rev.*, vol. 140, no. 4A, pp. A1133–A1138, Nov. 1965.

[4] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, “Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation,” *Phys. Rev. Lett.*, vol. 82, no. 12, pp. 2544–2547, 1999.

[5] M. Ernzerhof and G. E. Scuseria, “Kinetic energy density dependent approximations to the exchange energy,” *J. Chem. Phys.*, vol. 111, no. 3, pp. 911–915, 1999.

[6] S. Grimme, “Accurate description of van der Waals complexes by density functional theory including empirical corrections,” *J. Comput. Chem.*, vol. 25, no. 12, pp. 1463–1473, 2004.

[7] S. J. Pennycook and L. A. Boatner, “Chemically sensitive structure-imaging with a scanning transmission electron microscope,” *Nature*, vol. 336, no. 6199, pp. 565–567, 1988.

[8] P. M. Voyles, D. A. Muller, J. L. Grazul, P. H. Citrin, and H. J. L. Gossmann, “Atomic-scale imaging of individual dopant atoms and clusters in highly n-type bulk Si,” *Nature*, vol. 416, no. 6883, pp. 826–829, 2002.

[9] A. Singhal, J. C. Yang, and J. M. Gibson, “STEM-based mass spectroscopy of supported Re clusters,” *Ultramicroscopy*, vol. 67, no. 1–4, pp. 191–206, 1997.

[10] P. D. Nellist *et al.*, “Direct sub-angstrom imaging of a crystal lattice,” *Science (80-.).*, vol. 305, no. 5691, p. 1741, 2004.

[11] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, “7 × 7 Reconstruction on Si(111) Resolved in Real Space,” *Phys. Rev. Lett.*, vol. 50, no. 2, pp. 120–123, 1983.

[12] F. F. Abraham, I. P. Batra, and S. Ciraci, “Effect of tip profile on atomic-force microscope images: A model study,” *Phys. Rev. Lett.*, vol. 60, no. 13, pp. 1314–1317, 1988.

[13] D. M. Eigler and E. K. Schweizer, “Positioning single atoms with a scanning tunnelling microscope,” *Nature*, vol. 344, no. 6266, pp. 524–526, 1990.

[14] S. V. Kalinin, A. Borisevich, and S. Jesse, “Fire up the atom forge,” *Nature*, vol. 539, no. 7630. pp. 485–487, 2016.

[15] P. M. V. Jie Feng, Alexander V. Kvit, Chenyu Zhang, Jason Hoffman, Anand Bhattacharya, Dane Morgan, “Imaging of Single La Vacancies in LaMnO3.”

[16] H. Kim, J. Y. Zhang, S. Raghavan, and S. Stemmer, “Direct observation of Sr vacancies in SrTiO3 by quantitative scanning transmission electron microscopy,” *Phys. Rev. X*, vol. 6, no. 4, p. 041063, Dec. 2016.

[17] A. Belianinov, Q. He, M. Kravchenko, S. Jesse, A. Borisevich, and S. V. Kalinin, “Identification of phases, symmetries and defects through local crystallography,” *Nat. Commun.*, vol. 6, 2015.

[18] Z. Liu, K. Suenaga, Z. Wang, Z. Shi, E. Okunishi, and S. Iijima, “Identification of active atomic defects in a monolayered tungsten disulphide nanoribbon,” *Nat. Commun.*, vol. 2, no. 1, 2011.

[19] J. M. Lebeau, S. D. Findlay, L. J. Allen, and S. Stemmer, “Standardless atom counting in

scanning transmission electron microscopy," *Nano Lett.*, vol. 10, no. 11, pp. 4405–4408, 2010.

[20] S. Jukna, *Extremal combinatorics: with applications in computer science*. 2011.

[21] G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, pp. 1615–1619.

[22] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. 1964.

[23] J. M. Cowley, "Image contrast in a transmission scanning electron microscope," *Appl. Phys. Lett.*, vol. 15, no. 2, pp. 58–59, Jul. 1969.

[24] S. Somnath *et al.*, "Feature extraction via similarity search: application to atom finding and denoising in electron and scanning probe microscopy imaging," *Adv. Struct. Chem. Imaging*, vol. 4, no. 1, p. 3, 2018.

[25] R. Bro and A. K. Smilde, "Principal component analysis," *R. Soc. Chem.*, vol. 2, no. 4, pp. 433–459, 2014.

[26] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, 2002.

[27] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, 1981.

[28] A. Y. Borisevich *et al.*, "Suppression of octahedral tilts and associated changes in electronic properties at epitaxial oxide heterostructure interfaces," *Phys. Rev. Lett.*, vol. 105, no. 8, 2010.

[29] R. Ishikawa, A. R. Lupini, Y. Hinuma, and S. J. Pennycook, “Large-angle illumination STEM: Toward three-dimensional atom-by-atom imaging,” *Ultramicroscopy*, vol. 151, pp. 122–129, 2015.

[30] O. L. Krivanek *et al.*, “Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy,” *Nature*, vol. 464, no. 7288, pp. 571–574, 2010.

[31] X. Shen *et al.*, “Interlaced crystals having a perfect Bravais lattice and complex chemical order revealed by real-space crystallography,” *Nat. Commun.*, vol. 5, p. 5431, 2014.

[32] W. Zhou, J. Lee, J. Nanda, S. T. Pantelides, S. J. Pennycook, and J.-C. Idrobo, “Atomically localized plasmon enhancement in monolayer graphene,” *Nat. Nanotechnol.*, vol. 7, no. 3, pp. 161–165, 2012.

[33] B. Berkels *et al.*, “High Precision STEM Imaging by Non-Rigid Alignment and Averaging of a Series of Short Exposures,” *Microsc. Microanal.*, vol. 18, pp. 300–301, 2012.

[34] M. Ziatdinov *et al.*, “Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations,” *ACS Nano*, vol. 11, no. 12, pp. 12742–12752, 2017.

[35] A. De wael, A. De Backer, L. Jones, P. D. Nellist, and S. Van Aert, “Hybrid statistics-simulations based method for atom-counting from ADF STEM images,” *Ultramicroscopy*, vol. 177, pp. 69–77, 2017.

[36] X. Sang, A. A. Oni, and J. M. LeBeau, “Atom column indexing: atomic resolution image analysis through a matrix representation.,” *Microsc. Microanal.*, vol. 20, no. 6, pp. 1764–71, 2014.

[37] J. McQuenn, “Some methods for classification and analysis of multivariate observations,”

Computers and Chemistry, vol. 4. pp. 257–272, 1967.

[38] J. Vrbik and P. Vrbik, “Finite Markov Chains,” 2nd ed., 2013, pp. 5–38.

[39] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities.,” *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, 1982.

[40] P. V. C. Hough, “Method and means for recognizing complex patterns,” *US Pat. 3,069,654*, vol. 21, pp. 225–231, 1962.

[41] J. Illingworth and J. Kittler, “The Adaptive Hough Transform,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 690–698, 1987.

[42] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proc. 2nd Int. Conf. Knowl. Discov. Data Min.*, pp. 226–231, 1996.

[43] R. Shiju, V. V. Guliants, and N. R. Shiju, “Recent developments in catalysis using nanostructured materials,” *Appl. Catal. A Gen.*, vol. 356, no. 1, pp. 1–17, 2009.

[44] K. Xu, P. G. Cao, and J. R. Heath, “Scanning Tunneling Microscopy Characterization of the Electrical Properties of Wrinkles in Exfoliated Graphene Monolayers,” *Nano Lett.*, vol. 9, no. 12, pp. 4446–4451, 2009.

[45] N. Levy *et al.*, “Strain-Induced Pseudo-Magnetic Fields Greater Than 300 Tesla in Graphene Nanobubbles,” *Science (80-.).*, vol. 329, no. 5991, pp. 544–547, 2010.

[46] R. J. T. Nicholl *et al.*, “The effect of intrinsic crumpling on the mechanics of free-standing graphene,” *Nat. Commun.*, vol. 6, p. 8789, 2015.

[47] W. Bao *et al.*, “Controlled ripple texturing of suspended graphene and ultrathin graphite membranes,” *Nat. Nanotechnol.*, vol. 4, no. 9, pp. 562–566, 2009.

[48] L. Meng *et al.*, “Hierarchy of graphene wrinkles induced by thermal strain engineering,” *Appl. Phys. Lett.*, vol. 103, no. 25, 2013.

[49] L. Yang *et al.*, “Self-assembly of suspended graphene wrinkles with high pre-tension and elastic property,” *2D Mater.*, vol. 4, no. 4, 2017.

[50] S. Y. Xu *et al.*, “Unconventional superconductivity in magic-angle graphene superlattices,” *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 108, no. 3, pp. 1–10, 2015.

[51] Y. Cao *et al.*, “Correlated insulator behaviour at half-filling in magic-angle graphene superlattices,” *Nature*, vol. 556, no. 7699, pp. 80–84, 2018.

[52] J. Lin *et al.*, “AC/AB stacking boundaries in bilayer graphene,” *Nano Lett.*, vol. 13, no. 7, pp. 3262–3268, 2013.

[53] E. V. Castro *et al.*, “Biased Bilayer Graphene: Semiconductor with a Gap Tunable by the Electric Field Effect,” *Phys. Rev. Lett.*, vol. 99, no. 21, p. 216802, Nov. 2007.

[54] A. O’Hara, R. E. Kahn, Y.-Y. Zhang, and S. T. Pantelides, “Defect-mediated leakage in lithium intercalated bilayer graphene,” *AIP Adv.*, vol. 7, no. 4, p. 045205, Apr. 2017.

[55] Y. Zhang *et al.*, “Direct observation of a widely tunable bandgap in bilayer graphene,” *Nature*, vol. 459, no. 7248, pp. 820–823, 2009.

[56] W. Regan *et al.*, “A direct transfer of layer-area graphene,” *Appl. Phys. Lett.*, vol. 96, no. 11, p. 113102, Mar. 2010.

[57] Y. Mao and J. Zhong, “Structural, electronic and magnetic properties of manganese doping in the upper layer of bilayer graphene.,” *Nanotechnology*, vol. 19, no. 20, p. 205708, 2008.

[58] A. Y. Borisevich *et al.*, “Suppression of Octahedral Tilts and Associated Changes in Electronic Properties at Epitaxial Oxide Heterostructure Interfaces,” *Phys. Rev. Lett.*, vol.

105, no. 8, p. 087204, Aug. 2010.

[59] M. M. van Wijk, A. Schuring, M. I. Katsnelson, and A. Fasolino, “Relaxation of moiré patterns for slightly misaligned identical lattices: graphene on graphite,” *2D Mater.*, vol. 2, no. 3, p. 034010, Jul. 2015.

[60] I. Vlassiouk *et al.*, “Large scale atmospheric pressure chemical vapor deposition of graphene,” *Carbon N. Y.*, vol. 54, pp. 58–67, 2013.

[61] “2um holes with 2um spacing (657-200-AU QUANTIFOIL SUBSTRATE, 200M)No Title.” [Online]. Available: https://www.tedpella.com/technote_html/656, 657, 658 TN.pdf.

[62] O. L. Krivanek *et al.*, “An electron microscope for the aberration-corrected era,” *Ultramicroscopy*, vol. 108, no. 3, pp. 179–195, 2008.

[63] A. Witkin, “Scale-space filtering: A new approach to multi-scale description,” *ICASSP ’84. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 9, pp. 150–153, 1984.

[64] Y. Nishiyima and G. Oster, “Moiré Patterns: Their Application to Refractive Index and Refractive Index Gradient Measurements,” *J. Opt. Soc. Am.*, vol. 54, no. 1, pp. 1–5, 1964.

[65] Q. McNemar, “Note on the sampling error of the difference between correlated proportions or percentages,” *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[66] G. Kresse and J. Furthmüller, “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set,” *Phys. Rev. B*, vol. 54, no. 16, pp. 11169–11186, 1996.

[67] J. P. Perdew, K. Burke, and M. Ernzerhof, “Erratum: Generalized gradient approximation made simple (Physical Review Letters (1996) 77 (3865)),” *Physical Review Letters*, vol. 78, no. 8. p. 1396, 1997.

[68] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, no. 15, p. 154104, Apr. 2010.

[69] A. D. Becke and E. R. Johnson, "A density-functional model of the dispersion interaction," *J. Chem. Phys.*, vol. 123, no. 15, p. 154101, Oct. 2005.

[70] I. V. Lebedeva, A. V. Lebedev, A. M. Popov, and A. A. Knizhnik, "Comparison of performance of van der Waals-corrected exchange-correlation functionals for interlayer interaction in graphene and hexagonal boron nitride," *Comput. Mater. Sci.*, vol. 128, pp. 45–58, 2017.

[71] G. Blöchl *et al.*, "Projector augmented-wave method.," *Phys. Rev. B. Condens. Matter*, vol. 50, no. 24, pp. 17953–17979, 1994.

[72] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B*, vol. 59, no. 3, pp. 1758–1775, 1999.

[73] C. Koch, "Determination of core structure periodicity and point defect density along dislocations," Arizona State University, 2002.

[74] Y. C. Fung, *Foundation of Solid Mechanics*. 1965.

[75] A. Politano and G. Chiarello, "Probing the Young's modulus and Poisson's ratio in graphene/metal interfaces and graphite: a comparative study," *Nano Res.*, vol. 8, no. 6, pp. 1847–1856, 2015.

[76] Z. Wang and M. Devel, "Periodic ripples in suspended graphene," *Phys. Rev. B*, vol. 83, no. 12, p. 125422, Mar. 2011.

[77] L. Tapasztó, T. Dumitrică, S. J. Kim, P. Nemes-Incze, C. Hwang, and L. P. Biró, "Breakdown of continuum mechanics for nanometre-wavelength rippling of graphene," *Nat. Phys.*, vol.

8, no. 10, pp. 739–742, 2012.

[78] K.-K. Bai *et al.*, “Creating One-Dimensional Nanoscale Periodic Ripples in a Continuous Mosaic Graphene Monolayer,” *Phys. Rev. Lett.*, vol. 113, no. 8, p. 086102, Aug. 2014.

[79] F. Liu and T. C. Wang, “Quantized phenomena in graphene nanoripples,” *Carbon N. Y.*, vol. 96, pp. 1175–1180, 2016.

[80] E. Cerda *et al.*, “Geometry and physics of wrinkling.,” *Phys. Rev. Lett.*, vol. 90, no. 7, p. 074302, 2003.

[81] X. Wei, B. Fragneaud, C. A. Marianetti, and J. W. Kysar, “Nonlinear elastic behavior of graphene: Ab initio calculations to continuum description,” *Phys. Rev. B*, vol. 80, no. 20, p. 205407, Nov. 2009.

[82] R. Bistritzer and A. H. MacDonald, “Moire bands in twisted double-layer graphene,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 30, pp. 12233–12237, Jul. 2011.

[83] J. S. Choi *et al.*, “Friction Anisotropy-Driven Domain Imaging on Exfoliated Monolayer Graphene,” *Science (80-.).*, vol. 333, no. 6042, pp. 607–610, 2011.

[84] P. Gallagher *et al.*, “Switchable friction enabled by nanoscale self-assembly on graphene,” *Nat. Commun.*, vol. 7, p. 10745, 2016.

[85] R. Zan, U. Bangert, C. Muryn, P. Mattocks, B. Hamilton, and K. S. Novoselov, “Scanning Tunnelling Microscopy of Suspended Graphene,” *J. Phys. Conf. Ser.*, vol. 371, p. 012070, 2012.

[86] E. Miranda *et al.*, “Disorder-driven non-Fermi liquid behaviour of correlated electrons,” *Reports Prog. Phys.*, vol. 68, no. 10, pp. 2337–2408, 2005.

[87] E. Abrahams and G. Kotliar, “The metal-insulator transition in correlated disordered

systems," *Science*, vol. 274, no. 5294. pp. 1853–1854, 1996.

[88] P. A. Lee, N. Nagaosa, and X. G. Wen, "Doping a Mott insulator: Physics of high-temperature superconductivity," *Rev. Mod. Phys.*, vol. 78, no. 1, 2006.

[89] X. Shen *et al.*, "Interlaced crystals having a perfect Bravais lattice and complex chemical order revealed by real-space crystallography," *Nat. Commun.*, vol. 5, 2014.

[90] W. Zhou, J. Lee, J. Nanda, S. T. Pantelides, S. J. Pennycook, and J. C. Idrobo, "Atomically localized plasmon enhancement in monolayer graphene," *Nat. Nanotechnol.*, vol. 7, no. 3, pp. 161–165, 2012.

[91] O. S. Ovchinnikov *et al.*, "Theory-assisted determination of nano-rippling and impurities in atomic resolution images of angle-mismatched bilayer graphene," *2D Mater.*, vol. 5, no. 4, p. 041008, 2018.

[92] D. T. Lee and B. J. Schachter, "Two algorithms for constructing a Delaunay triangulation," *Int. J. Comput. Inf. Sci.*, vol. 9, no. 3, pp. 219–242, 1980.

[93] A. D. Becke and E. R. Johnson, "A density-functional model of the dispersion interaction," *J. Chem. Phys.*, vol. 123, no. 15, 2005.

[94] J. Feng *et al.*, "Imaging of Single La Vacancies in LaMnO$_{3}$," Nov. 2017.

[95] S.-Z. Yang *et al.*, "Direct cation exchange in monolayer MoS2 via 'explosive' recombination-enhanced migration (under review)."

[96] S.-Z. Yang, Y. Gong, P. Manchanda, and Y.-Y. Zhang, "Re doped and stabilized MoS 2 alloy monolayers with basal plane catalytic activity (in press)."

[97] D. Liu, Y. Guo, L. Fang, and J. Robertson, "Sulfur vacancies in monolayer MoS 2 and its electrical contacts," *Appl. Phys. Lett.*, vol. 103, no. 18, p. 183113, Oct. 2013.

[98] M. Houssa, K. Iordanidou, G. Pourtois, V. V. Afanas'ev, and A. Stesmans, "Point defects in MoS_2: Comparison between first-principles simulations and electron spin resonance experiments," *Appl. Surf. Sci.*, vol. 416, pp. 853–857, Sep. 2017.