UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Cohesive Experimental and Analysis Techniques for Angle Resolved Photoemission Spectroscopy

Permalink

https://escholarship.org/uc/item/7fz887nk

Author

Stansbury, Conrad Holland

Publication Date

2022

Supplemental Material

https://escholarship.org/uc/item/7fz887nk#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Cohesive Experimental and Analysis Techniques for Angle Resolved Photoemission Spectroscopy

By

Conrad Stansbury

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alessandra Lanzara, Chair
Professor Feng Wang
Professor Kwabena Bediako

Spring 2022

Cohesive Experimental and Analysis Techniques for Angle Resolved Photoemission Spectroscopy

Copyright 2022

by

Conrad Stansbury

Abstract

Cohesive Experimental and Analysis Techniques for
Angle Resolved Photoemission Spectroscopy

by

Conrad Stansbury

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Alessandra Lanzara, Chair

This thesis focuses on phenomena arising in two dimensional materials and the software, statistical methods, and analysis tools materials scientists can bring to bear in understanding these phenomena. Concerning materials, we discuss on WS_2, a transition metal dichalcogenide, as well as moiré superlattices and heterostructures which contain at least one WS_2 layer.

Read front-to-back, it tells the story of software tools built to collect and understand photoemission data, which begins by situating the problems facing photoemission and the scientists who use it (Chapter 1) and culminates in two large, cooperative pieces of software built to furnish solutions: PyARPES (Chapter 2) for the analysis of ARPES data and AutodiDAQt (Chapter 3) which systematizes its collection.

The subsequent experimental chapters demonstrate that scientific insights can readily be drawn with these new tools in hand. Understanding the physics of transition metal dichalcogenides and superlattices constituted from several layers of them (Chapter 4) requires nanoscale photoemission experiments, plenty of exploratory data analysis, and careful treatment of data. We determine that electron localization in WS_2/WSe_2 is part of a richer moiré phenomenology including inhomogeneous hybridization, miniband formation, and multiple superlattices (Chapter 4). On the way, we include a treatment of the physics of multilayer WS_2 where we use ARPES to perform “momentum metrology”. These results lay the groundwork for the interpretation of superlattices, which lay epicycles on top of the multilayer picture. Leaving 2D materials, we explore where novel treatment of data can take photoemission science if hardware innovation stalls (Chapter 5). This section concludes with experiments testing the application of spectral unmixing to contaminated ARPES measurements.

Read back-to-front, the reader infers a different chronology: experimental necessities drove attempts to innovate in the interpretation of photoemission data. The reality is somewhere in between and reflects the author’s preferred aesthetic for the scientific process. First, methodology, interpretation, and the exigencies of scientific results are in constant dialog. Second, shortening the path for feedback between ideas, data, and experiments is the surest path to progress. Chapter 6 contains idle speculations about what might lie down that path.

Contents

Acknowledgements iv

Curriculum vitae vi

1 An Introduction to Angle-Resolved Photoemission Spectroscopy 1

1.1 Motivation and Context 1

1.2 Angles and Momenta 2

1.2.1 The Spectral Function, the Photocurrent, and the ARPES Measurand 3

1.2.2 The Basic Geometry of ARPES Experiments 6

1.2.3 A Practical Example: Extracting the Photoelectron Momentum from Measured Angles 7

1.2.4 Consequences of Measuring the Photocurrent Density 8

1.3 A Practitioner's Overview of ARPES 12

1.3.1 Two Models for Photoemission 14

1.3.2 Disorder 15

1.3.3 Surface quality 16

1.3.4 Surface aging 17

1.3.5 The light source 18

1.3.6 "Two photon" effects in the photoemission process 19

1.3.7 Propagation of the energetic electron through sample 20

1.3.8 Vacuum electrostatics and magnetostatics 20

1.3.9 Hemispheres: Final detection 23

1.3.10 At analysis time 23

2 PyARPES 24

2.1 Practical factors driving the necessity of common data analysis software for ARPES . 24

2.2 Software architecture 26

2.3 Data model 27

2.3.1 Data provenenace 29

2.4 Angle-to-momentum conversion 31

2.5 Data Corrections and Data Preparation 32

2.6 Interactive analysis 32

2.7 Common Analyses 33

2.8 Data exploration and data analysis driven by machine learning 33

2.9 Comprehensive usage documentation, example data, and tutorials 34
2.10 Theoretical modeling and detector response functions 35
2.11 Impact and Future Directions 36
3 AutodiDAQt 37
3.1 Scientist interaction with data acquisition 39
3.2 AutodiDAQt and Existing Software Solutions 40
3.3 User Interface Generation 40
3.4 Composability of Data Acquisition Programs with an Algebraic (Sum and Product) Structure 42
3.5 Analysis-in-the-Loop: Realization in AutodiDAQt and Applications in ARPES 45
3.5.1 Applications of Analysis-in-the-Loop to Nano-ARPES 47
3.6 Building Real Acquisition Programs with AutodiDAQt, a Few Examples 50
3.7 System Internals: Implementing a Data Acquisition Runtime with Remote Brokered Programming 54
3.8 Conclusions and Long-Term Prospects for AITL DAQ Systems 56
4 Emergent Behavior and Localization in Electronic Superlattices 57
4.1 A general approach to flat band emergence in 2D materials 58
4.2 Getting an Interaction Potential 59
4.2.1 Orbital Structure of 2H-TMDCs WS_2 and WSe_2 60
4.2.2 Long wavelength superlattices 61
4.3 Correlates and Consequences of General 2D Material Flat Bands 62
4.4 Electron Localization and Hybridization in WS_2/WSe_2 Superlattices 63
4.5 Preparation of Moiré WS_2/WSe_2 Samples for ARPES 64
4.5.1 Flip-and-place Polypropylene Carbonate Transfer 66
4.5.2 Polyethylene Terephtalate Transfer 67
4.5.3 Determination of the Local Interlayer Twist Angle after Fabrication 67
4.6 Band Structures of Semiconducting WS_2 and WSe_2 68
4.6.1 Motivating Continued Study of TMDC Thin Films 69
4.6.2 Nano ARPES Measurements of 1L WSe_2 and 1L-3L WS_2 69
4.6.3 Significance of ARPES Measurements on Multilayers 76
4.7 High Symmetry Spectra in Monolayers and Moirés 76
4.8 Intrinsic Linewidth Modification in Moirés, Ruling out Sample Quality and Measurement Effects 78

4.9 Flattened Electronic States in Long Wavelength WS_2/WSe_2 Moirés 81
4.9.1 Obtaining Supercell Spectral Functions by Unfolding 84
4.9.2 Applying Supercell Unfolding to Understand Moiré WS_2/WSe_2 85
4.10 Incommensurate Superlattice Hybridization at the **Γ** Point of Moirés 86
4.11 A Phenomenology for $MX_2/M'X'_2$ Moirés 86
4.11.1 Constructing a Bridge to the Optical Literature 87
4.11.2 Extending Results across $\boldsymbol{\lambda}$ and across Material Families 88
4.12 Accessing Intra-moiré Physics with Large Beams 94
4.13 Multiple Superlattices 96
4.14 Concluding on Moiré Superlattices 98
5 Reinterpreting ARPES Data and Standard Analysis Approaches 99
5.1 Machine Learning Enabled Exploratory Data Analysis in ARPES 100
5.2 Noticing Variations in a Diffuse Heterostructure Band with Dimensionality Reduction 103
5.3 Using Machine Learning to Extend the Capabilities of ARPES Experiments 105
5.3.1 Measurement Variance, Domain Structure, and One-Dimensional Simulations . 108
5.3.2 Applying NMF Unmixing to NanoXPS Data from an Artificial Domain Pattern 112
5.3.3 Exploring Regularization Functions for ARPES NMF Applications 117
5.4 Impact and Outlook for Spectral Unmixing in ARPES 121
6 Conclusion and Thoughts on the Future Coevolution of ARPES Data and Its Interpretation 123
6.1 Machine Learning to Establish Stronger Connections with Theory 123
6.2 Autonomous Experimentation 124
6.3 Machine Learning to Improve Performance of Experiments 125
6.4 Domain Specific Improvements in Analysis Tools 125
7 References 127
8 Appendix 1: Practical Reference for Deriving Angle-to-Momentum Transforms with Symbolic Algebra Packages 142

Acknowledgements

Foremost, I need to thank my advisor Prof. Alessandra Lanzara for her continuing support and for the bulwark she formed for our group against realities which threatened to distract us from pursuing open-ended science. Funding was never scarce, conferences always available, and these details gave me and other students room to exercise intellectual freedom. Above all Alessandra curated a group of highly curious students. Some, like myself, were just entering the field of condensed matter physics when they arrived at Berkeley, having pursued different sets of interests in their undergraduate years. If leaders are to be judged principally by the teams they build, Alessandra should take great pride in the dynamism in instrumentation and science she has collected in the group.

As is familiar to anyone who has spent time in a lab at night, it can be difficult to assume the role of scientist exercising mastery over lifeless hardware. An animating force lived in the gaps of my knowledge and rendered initial experiments in the lab more like an overnight in Bradbury's *The Veldt*. I weathered those nights, but the credit is due to a previous generation of students (Kenny Gotlieb, Greg Affeldt, and others) who generously donated their time and expertise when I was first entering the lab. These are debts which cannot be repaid, but which I have done my best to transfer to a next generation of students. Even so, I doubt my ledger would come up clean.

A walk past the beamlines surrounding the ALS storage ring yields many more names I need to acknowledge. At MERLIN, Jonathan Denlinger offered more generously of his deep expertise on photoemission than would have been due as an advisor. He is a resource the ALS is deeply lucky to have and a model for retaining deep seated curiosity throughout a scientific career. Sung Kwan Mo and Alexei Fedorov cheerfully endured naïve and pestering questions about molecular beam epitaxy when our group's students came running on errand. At MAESTRO, Chris Jozwiak has built a bridge from our group to the instrumentation expertise of what might be the most engineering minded beamline at the ALS. I was deeply spoiled, as Chris' expertise on spin-ARPES and hardware debugging would easily have sufficed on its own. Aaron Bostwick is as talented an engineer as he is a scientist—I am grateful to him for convicting me by example that my inclination to engineer and build systems might not be a defect and for the ease with which he dispatched instrumentation troubles that I and other students encountered. Eli Rotenberg's unflinching vision for nanoscale photoemission sciences is inspiring. He joined for many day-measurements and discussions in the MAESTRO hutch to teach and—as far as I can tell—out of the joy of collecting data and recollecting the long history of photoemission. I often needed the reminder that science can be a joyful process.

A few dozen meters west of the ALS, in Building 2, I watched older students graduate as later I watched new students arrive. Before they left, I enjoyed discussions with (or reaped the benefits of pestering) Tristan, Drew, Greg, Kenny, Chiu-Yun, and Jonathan. Doubtless the whole bunch of us were a nuisance for our postdocs Cassi, Claudia, and Ping. Sam Ciocys, Nick Dale, and I constituted whatever closest resembled my cohort during my tenure in the group. We were very loosely coupled and with interests and expertise as divergent as can be found in three members of the Berkeley student body. I owe them each deeply for their ambition in taking on challenging projects for the good of the group and for their providing a checksum on wilder ideas I occasionally proposed. Younger students of all persuasions—Kayla, Danny, Burhan, Prosper,

Sam “B”, Max, Akshara—contributed energy they also carried to endeavors outside the lab. If I was as energetic upon arrival as they continue to be, I no longer remember it.

I am very grateful for the support of the members of my thesis and qualifying exam committee—Professors Alessandra Lanzara, Kwabena Bediako, Feng Wang, and Ehud Altman—who in receiving my qualification exam talk graciously endured my first attempts to collect the threads I’ve spun over the past years at Berkeley into a cohesive whole. While it is never a straightforward task to combine a PhDs worth of work, my path has been particularly meandering: Alessandra’s uncanny sense for narrative science and effective communication shouldered some of the burden for the message I delivered in that talk, as it does now as I sit down to the task of writing this thesis.

From the patio between LBL Building 2 and the ALS mezzanine, I indulged in early morning or late evening moments of solitude as the sun transited the Marin headlands. Occasionally, during beamtimes, strange hypnagogic thoughts entered my head: LBL covered in its building and instruments and webbed with ethernet cable sometimes seemed like a stainless-steel giant laying on the hill. These days, the shoulders of giants being too short, we take our vantage on more numerous and complicated machines built by multitudes of scientists. Working at LBL, it’s impossible not to feel indebted to the long legacy of cooperative science, tool building, and engineering spirit unique to national labs.

Considering those nights, I realize I am also endlessly grateful for the beauty offered by the bay—above all food, natural beauty, and wonderment. As a student at LBL, the wonderment didn’t come free: it is easy to feel divorced from peers and happenings at Berkeley when you’re trapped on the hill. The Blue Line was an essential tether back to campus, coffee, and conversations with students and friends. For each successful or unsuccessful collaboration with a group on campus, I have several of the shuttle drivers to thank.

I don’t assign particular significance to endings or beginnings. My family and friends round out an incomplete account of the support I’ve received through the PhD program only because even after consideration I can’t summon words potent enough to describe their role. Emma, my wife and a decorated member of this year’s outgoing Applied Science and Technology doctoral cohort, knows without saying the inexplicable role that our love has played in delivering us to the doorstep of our degrees. When we encountered friction and frustration in our work, we found respite and counterpoint in the friendship and home we built together. My parents are proof that no rule is sharp enough to be without exception: they are giants whose shoulders I doubt I shall ever climb from.

Our family first expanded when Smudge and Moose arrived in our lives a few years ago and changed the tenor of our final years at Berkeley. A few months ago, on January 25th, our daughter Rowan was born. She changes everything that is to come.

Conrad Stansbury
Virginia Street
Berkeley, California

Curriculum vitae

Education

2022 University of California, Berkeley, CA, USA
Ph.D., Physics

2020 University of California, Berkeley, CA, USA
M.A., Physics

2015 Stanford University, Stanford, CA, USA
B.S., Physics

Publications

Two-dimensional Materials, Superlattices, and Heterostructures

1. "Comprehensive band structures and evolution from mono- to multilayer in the semiconducting transition metal dichalcogenides WS_2 and WSe_2", **C. H. Stansbury**, et al. In preparation.
2. "Visualizing electron localization of WS_2/WSe_2 moiré superlattices in momentum space". **C. H. Stansbury**, M. I. B. Utama, C. G. Fatuzzo, E. C. Regan, D. Wang, Z. Xiang, M. Ding, K. Watanabe, T. Taniguchi, M. Blei, Y. Shen, S. Lorcy, A. Bostwick, C. Jozwiak, R. Koch, S. Tongay, J. Avila, E. Rotenberg, F. Wang, and A. Lanzara. Science Advances, **7** 37, (2021).
3. "Correlation-driven electron-hole asymmetry in graphene field effect devices", N. Dale, R. Mori, M. I. Bakti Utama., J. Denlinger, **C. H. Stansbury**, C. G. Fatuzzo, S. Zhao, K. Lee, T. Taniguchi, K. Watanabe, C. Jozwiak, A. Bostwick, E. Rotenberg, R. Koch, F. Wang, A. Lanzara. npj Quantum Materials **7**, (2022).

Machine Learning, Software, and Data Acquisition

1. "autodiDAQt: Simple scientific data acquisition software with analysis-in-the-loop", **C. H. Stansbury** and A. Lanzara. Submitted.
2. "A new approach to nano-probe photoemission microscopy by leveraging fluctuations", **C. H. Stansbury**, C. Beamer, S. Kahn, A. Zettl, and A. Lanzara. In preparation.
3. "PyARPES: An analysis framework for multimodal angle-resolved photoemission spectroscopies", **C. H. Stansbury** and A. Lanzara. SoftwareX, **11**, 100472 (2020).
4. "Fuzzy Jets," L. Mackey, B. Nachman, A. Schwartzman, and **C. H. Stansbury**, JHEP (2016).

Invited Talks and Seminars

1. "Visualizing Electronics in WS_2/WSe_2 Moirés / Pushing ARPES Limits with Spectral Unmixing", UC Berkeley 290K Seminar, (2022).
2. "Visualizing Electron Localization in WS_2/WSe_2 Moiré Superlattices", Fritz-Haber-Institut, Max Planck Gesellschaft, (2021).
3. "Studying Transition Metal Dichalcogenide Superlattices with nano-ARPES" *Layered Quantum Materials Workshop: Electronic and Magnetic Properties*. ALS User Meeting (2019).
4. "Fuzzy Clustering Algorithms for Jet Reconstruction", *Connecting the Dots 2015*, UC Berkeley/LBNL, (2015).

Contributed Talks

1. "Visualizing Electron Localization in WS_2/WSe_2 Moiré Superlattices", EFRC PI Meeting Poster Session, 2021.
2. "Achieving Smaller Effective Spot Sizes in nano-ARPES with Machine Learning", *American Physical Society March Meeting*, (remote), 2021.
3. "Studying Transition Metal Dichalcogenide Superlattices with nano-ARPES", *American Physical Society March Meeting*, (cancelled, COVID-19 pandemic), 2020.
4. "Driving Topological Materials Towards New Physics", DOE Program Review Poster Session, MSD Ultrafast Materials Program, (2019).
5. "Orbital and Charge Participation in the Magnetic Phases of Intercalated $Fe_{1/3}TaS_2$, $Fe_{1/3}NbS_2$ and Their Alloys", *American Physical Society March Meeting*, Boston, Massachusetts (2019).
6. "Electronic Structure Dynamics in Semi-metallic TMD $Mo_{1-x}W_xTe_2$ using Time- and Angle-Resolved Photoemission Spectroscopy", *American Physical Society March Meeting*, Los Angeles, California (2018).

Teaching

2017 UC Berkeley: Graduate Student Instructor, 111B Experimental Lab

2016 UC Berkeley: Graduate Student Instructor, 8A Introductory Mechanics

2015 Stanford University: Course Instructor, Physics 91 SI, Practical Computing for Scientists

1 An Introduction to Angle-Resolved Photoemission Spectroscopy

1.1 Motivation and Context

ARPES is a technique which has for decades advanced on wave after wave of hardware improvements. Bright, high resolving power undulator sources at synchrotrons and highly monochromatic UV lasers made ARPES a powerful tool for understanding valence electronics where smaller energy scales are relevant. Hemispheres permitted band mapping, rapid experimentation, and secured a place of relevance for ARPES in materials characterization.

With the advent of nano-ARPES, the current wave may be cresting. From this vantage, it is worth surveying what beachhead ARPES might crash upon if not buoyed by future advances. What's apparent to the author is that ARPES has through countless hardware advances changed from a technique offering a few kB of energy distribution curves per experiment to one providing 10s or 100s of GB per session. Prior generations of photoemission scientists were parched for just a few energy distribution curves on critical systems like the cuprates, while the generation now being brought up on the technique are awash in data on ever more exotic systems, even complete band maps. High resolution electrostatic ToFs and PEEMs, ripples in the water for future generations of graduate students, may carry the technique further still, but only exacerbate this situation in the end.

This boon for the technique is also a critical challenge: if we don't know how to extract interesting and meaningful physics from the huge volume of data being produced on current generation instruments, experiments will slow, and materials understanding will be delayed. As a graduate student in a group deeply embedded with the ALS where "some extra beamtime" could often be scrounged, this surfeit of data probably hampered as much as it helped.

Solving this problem requires the consensus of scientists to agree on practices and approaches to dealing with data, at all levels. Collectively, we need to determine what formats we use, what approaches are acceptable in interpreting data, how data should be shared in collaboration and stored after publication, and what tools we will use to accomplish these aims. Consensus requires first exploring the possible solution space and proposing options. This framing elucidates one of two threads in the thesis: Chapters 2, 3 and 5 provide one holistic vision for how photoemission data might be collected, shared, and analyzed in the future with an eye to ongoing developments and computer vision applications in Chapter 6.

What about the second, experimental thread? This is simple too: good products rarely come out of organizations which aren't their own clients (jargon: "eat their own dogfood"). Taking this maxim and minimizing or personalizing it, the author is a scientist and real ARPES experiments on high data bandwidth experiments like nanoARPES are the ideal testbed to trial long term approaches to ARPES' "data problem". This risks minimizing the experimental side of this thesis, but it properly captures the crosstalk between tool development, which was strongly guided by the immediate needs of the author and colleagues' experiments, and experiment, which was sometimes enabled by software or machine learning tools.

Where does this leave us for the remainder of the thesis? The right first step is to motivate and explain ARPES, the experimental technique connecting the work. This chapter gives a comprehensive and high-level overview of angle-resolved photoemission spectroscopy (ARPES). Although ARPES has grown to include many complicated experimental subdisciplines, all applications refer to momentum-sensitive electron spectroscopy. As the name implies, ARPES measures electrons ejected from samples by the photoelectric effect.

This thesis is more concerned than others with the acquisition and analysis of ARPES data for its own sake. We use ARPES to study WS_2/WSe_2 superlattices later, but Chapters 2, 3, and 5 are essentially concerned with the technique itself and the properties and structure of ARPES data without grounding to any physics. This shift of focus makes it valuable to consider ARPES not merely at a theoretical level, but in terms of the practical effects which arise in ARPES data. Throughout this chapter we will give a hopefully holistic view of the technique which will be practical for new entrants into the field by considering how ARPES measures the photoelectron momentum with a complete consideration of the geometrical details and how we can implement coordinate transforms for the volumetric data produced by our instruments.

Even after this process, our data can have many issues arising from the instrumentation, measurement procedures, and physics of the measurement process. We include a practical discussion of some of these issues so that can be recognized, prevented in experiment design, or mitigated in analysis. The picture of ARPES presented here is much more practically oriented and therefore theoretically incomplete. Throughout, we have tried to provide a collection of resources which provide the missing theoretical underpinnings.

1.2 Angles and Momenta

Very briefly, angle resolved photoemission spectroscopy is a technique which provides direct access to the behavior of electrons in semiconductors, semimetals, and metals by measuring the momentum and energy of electrons ejected via the photoelectric effect (*1*, *2*). Through gradual refinement, experiments based upon the photoelectric effect were able to validate Einstein's theoretical picture and measure Planck's constant h (*3*), directly measure the energy spectrum of the core levels (*4*), and ultimately provide access to the entire valence band structures at high resolution (*5*).

The name "angle-resolved" photoemission spectroscopy indicates to an experimental scientist that there is work to be done before quantities of interest—expressed in terms of binding energies and momenta—can be recovered in terms of the quantities measured—expressed in terms of pixel units on raw measurement hardware or in terms of relative angles and electron kinetic energies.

The correspondence between these two is idealized by conservation relationships for the electron-photon interaction and are frequently quoted in the photoemission literature (*6*–*10*)

$$E_{kin} = h\nu - E_B - \Phi \tag{1.1}$$

$$\hbar \boldsymbol{k}_{\parallel} = \sqrt{2mE_{kin}} \sin\eta \, \widehat{\boldsymbol{k}}_{\parallel} \tag{1.2}$$

$$\hbar k_{\perp} \approx \sqrt{2m(E_{kin}\cos^2\eta + V_0)}. \quad (1.3)$$

The simplicity of presentation of these equations belies the complexity of the situation to photoemission novices and to physicists acquainted with other techniques who need to use ARPES to answer a question of limited scope and who want to avoid the occupational hazard of becoming a scientist using ARPES full-time. For one, there's a pesky modification to the equation for $k_{\perp}$ arising through the presence of the inner potential, V_0. Although we will not discuss it in depth, the assumption of a constant inner potential is only accurate for large input photon energies where we can approximate the final states as free electron like. Even at moderate energies however, this approximation is close enough to the truth to be useful. Were it not for this term, the out of plane momentum equation would be the simple geometrical analogue for the equation for $k_{\parallel}$. A deeper and less obvious problem is that the spherical declination angle η is not measured through the experiment, so the phrase "angle resolved" in ARPES does not directly refer to this this quantity either.

The issues arise because the clean theoretical picture above describes energy and momentum conservation for photoemission in the frame of reference of the sample in spherical coordinates (η, φ), but these are not the angles which are measured in real experiments. We will have to leave this nomenclatural mystery aside for a moment, so that we can discuss the issue of how ARPES extracts electrons from a sample surface.

1.2.1 The Spectral Function, the Photocurrent, and the ARPES Measurand

For now, we trust that after some data treatment the ARPES measurand $M(\boldsymbol{\theta}_m)$, expressed in terms of some measured coordinates $\boldsymbol{\theta}_m$, can be transformed into $M(\boldsymbol{k}, \omega)$, expressed in terms of the initial state crystal momentum. Because ARPES measures electron arrivals, the measurand M can be interpreted as the momentum and energy resolved photocurrent. If we had perfect ARPES apparatus, capable of exactly measuring the angle and kinetic energy of each outgoing electron, then the ARPES measurand would be proportional (*7*, *8*) to

$$M(\boldsymbol{k}, \omega) \propto I_0(\boldsymbol{k}, h\nu, \boldsymbol{A}) f(\omega) A(\boldsymbol{k}, \omega). \quad (1.4)$$

Definition up to proportionality arises because ARPES experiments must be performed over relatively long durations of time t (from seconds to hours, depending on conditions). While we stated earlier that M was the photocurrent, this practical aspect of experiments means that we should instead interpret the $M(\boldsymbol{k}, \omega)$ as the momentum and energy resolved photocharge. In practice, the distinction never matters since all quantities we care to extract are defined relative to a reference intensity. Directly comparing measurements means we must normalize the collected charge by the experiment duration $M(\boldsymbol{k}, \omega) = tI(\boldsymbol{k}, \omega)$. The quantity, $I(\boldsymbol{k}, \omega)$, is the momentum and energy resolved photocurrent and can be thought of as the most fundamental physical observable for ARPES.

Briefly analyzing the right-hand side of the equation above, I_0 is the matrix element term expressing the dipole interaction between the UV light used in the experiment and the sample, $f(\omega)$ is the Fermi occupation of initial state electrons in the sample, and $A(\boldsymbol{k}, \omega)$ is the spectral function for the sample. The Fermi function places a straightforward role: the photoemission

cross section for a state is weighted by its occupation. Both the matrix element term and the spectral function encode some sample physics. The matrix element

$$I_0(\boldsymbol{k}, h\nu, \boldsymbol{A}) = \left|\langle \Psi_f \mid \boldsymbol{A}\cdot\boldsymbol{p} \mid \Psi_i\rangle\right|^2 \tag{1.5}$$

summarizes the transition between initial and final states under the dipole approximation and where $|\Psi_j\rangle$ are the N-particle eigenstates for the entire system. As in the discussion of photoexcitation from atomic orbitals, the matrix element contains all the orbital selective physics. This means that the relative intensity of bands in ARPES can, with sufficient treatment of the geometry (which changes the orientation of the incoming light expressed through the vector potential $\boldsymbol{A}$), be used to map the orbital composition of bands in a material(*11–13*). Conversely, if the orbital composition is known, the experimental geometry can be adjusted into favorable configurations for measurement. In cases where geometric effects are subjects of study, performing photoemission in a fixed, high symmetry configuration is advantageous: this is one of many reasons why deflection mapping is becoming more popular.

To establish basics here, the spectral function can be thought of as mapping the location of bands in materials. Although this is extremely reductive, we will return to the physical underpinnings of the spectral function later. In the limit, we can consider instead using Fermi's golden rule to assign the transition probabilities between the initial states and the vacuum final states instead

$$I_0(\boldsymbol{k}, h\nu, \boldsymbol{A}) \propto \int_{i,f} \frac{2\pi}{\hbar}\left|\langle \Psi_f \mid \boldsymbol{A}\cdot\boldsymbol{p} \mid \Psi_i\rangle\right|^2 \delta\left(\omega_f - \omega_i - h\nu\right). \tag{1.6}$$

In the situation where we can assume that the electrons are functionally noninteracting, the N-particle eigenstates can be replaced for the single electron states $|\psi_j\rangle$.

In practice, the measurement hardware we use is far from ideal. If angular aberrations are corrected, real hardware can measure angles with error on the order $0.1°$ and energies with error ranging from 1 to 100 meV depending on circumstance. These show up in convolution with the measurand, but they are not the only sources of resolution degradation which threaten the theoretical link between the observable M and the sample physics.

This story is very academic, and it can be helpful to have an idealized picture in mind when we think of the photoemission process. Figure 1.1 provides this, by visualizing what the photocharge M looks like in the xz-plane if every electron were fixed in place and stained black after propagating away from the sample surface for a fixed time interval. The diagram shows two roughly spherical charge distributions corresponding to a VUV excitation energy of 10 eV in the top half of the diagram (a compromise for demonstration among 6 eV, 7 eV, and 11 eV lasers) and 60 eV (as might be found at a soft X-ray ARPES beamline) in the bottom half. Because the electrons at the Fermi surface are moving much more slowly for the 10 eV case, they are imaged at 50ns instead of at 25ns for the bottom cloud.

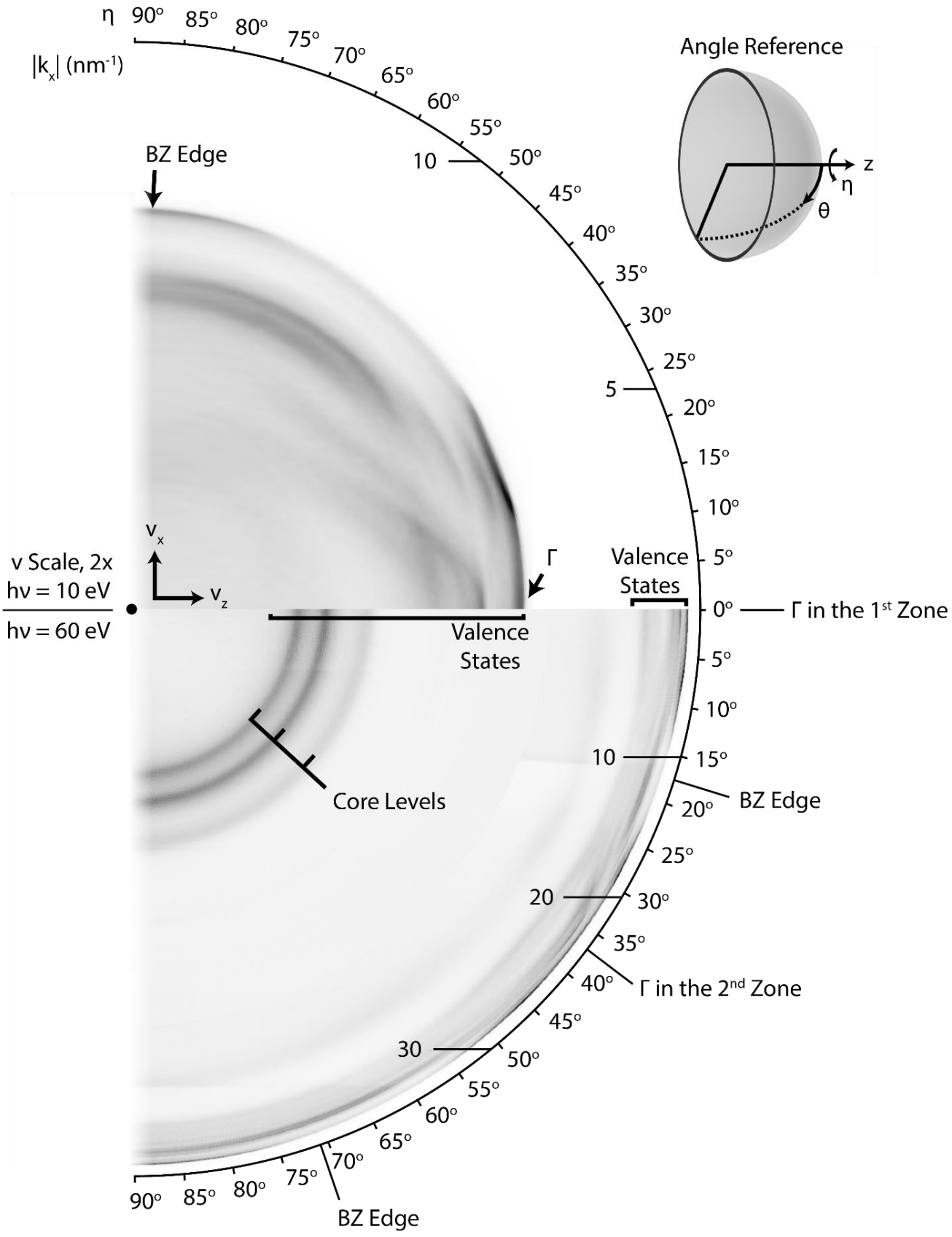

η
90° 85° 80° 75° 70° 65° 60° 55° 50° 45° 40° 35° 30° 25° 20° 15° 10° 5° 0° 5° 10° 15° 20° 25° 30° 35° 40° 45° 50° 55° 60° 65° 70° 75° 80° 85° 90°
|k_x| (nm^-1)
10
5
10
20
30
Angle Reference
z
η
θ
BZ Edge
v_x
v_z
v Scale, 2x
hν = 10 eV
hν = 60 eV
Γ
Valence States
Valence States
Core Levels
Γ in the 1st Zone
BZ Edge
Γ in the 2nd Zone
BZ Edge

Figure 1.1 **The Photoelectron Cloud's Density Imaged 25 ns (50 ns, top) after Photoemission** If we could take a picture of the expected location of electrons in the plane leaving a sample a short time after photoemission, it would look something like the representation above. Two representative photon energies—a VUV energy of 10 eV on the top and a soft x-ray energy of 60 eV on the bottom—are used to photoexcite electrons from a sample in vacuum. Their expected density as they propagate away in vacuum is shown 25 nanoseconds (50 ns, top/10 eV) after the photoemission process where the fastest electrons, travelling some 4400 km/s with nonrelativistic energies of 56 eV, have traversed 4 inches and halfway crossed the page. The structured, fastest travelling shell of photoelectrons comes from the valence band, with a hard outer cutoff due to the Fermi occupation. The structure is distorted but periodic in angle, because of the finite Brillouin zone in crystalline materials. Below the valence band there is a wide gap in velocity before the bright, nearly spherically symmetric core level states can be seen. Because these core levels are deeply bound in the material, in the vacuum final state they have energy reduced by their binding energy. At VUV energies, the valence states are less compressed in velocity and in angle and only one Brillouin zone can be seen. No core level photoelectrons can be emitted, because the photons are insufficient to pay the energetic barrier to remove them from the material. Some practical angles and momenta provided for intuition. The indicated momenta are calculated at the Fermi surface.

We can use a few features as waypoints in later discussion. At 25ns, after propagating halfway across the page, the valence electrons travel at $4400\ \mathrm{km/s} \approx 0.015\mathrm{c}$. This is very fast, but it indicates we generally don't have to worry too much about relativistic effects when we perform soft X-ray ARPES, as when building time-of-flight analyzers. We also see a great deal of interesting structure in the photocharge distribution. At 60 eV, the dispersive valence states near the Fermi level are a very narrow shell There is much more periodicity and structure in the declination angle η, which has been marked in terms of the Brillouin zone boundaries in the measured material. Because we are in a nonrelativistic setting and $\boldsymbol{p} = m\boldsymbol{v}$, drawing horizontal lines corresponding to constant v_x shows iso-momentum curves in the outgoing cloud. The spherical shape then says there is a sharp cutoff above which we cannot measure for a given input energy. The pinching in momentum and this cutoff are both direct consequences of the original energy and momentum conservation picture and explain a great deal about how ARPES is used in band mapping. VUV and UV photon energies permit high spectral resolution because features are spread out in angle-space and because of finite resolving power of the analyzer. However, UV energies are often insufficient to measure the entire first Brillouin zone of materials, and this places experimental constraints.

1.2.2 The Basic Geometry of ARPES Experiments

We've now had some flavor for what ARPES measures, built some graphical intuition, and motivated why we want to use ARPES to study materials. It's time to justify how ARPES does this. We need to understand which angles ARPES does measure, and how the real experimental geometric is connected to the semiclassical description of the photoemission process perspective proffered in the first section of review papers.

To answer this, we must take account of the geometry of a modern ARPES experiment. First, due to the difficulty of obtaining perfectly flat samples and fixing them flat upon the sample holders used or measurement, the sample is mounted relative to the goniometer and cryostat at fixed Euler angles. Second, the orientation of the sample holder is controlled relative to the electron analyzer—the physical apparatus we use to measure the photoelectron kinetic energy and one propagation angle in the lab frame. Control is achieved via rotational degrees of freedom on the goniometer which vary from apparatus to apparatus. All that has been described thus far is

only enough to specify the angles of the electrons in the lab frame as they propagate *in vacuo* towards the electron analyzer. Further coordinates need to be introduced to account for angular degrees of freedom resolved by popular hemispherical electron analyzers or two-angle imaging time-of-flight systems and to account for degrees of freedom of the analyzer relative to the lab frame. These usually consist of real rotations of measurement hardware, including on-axis rotation of the photoelectron analyzer as in the micro-ARPES beamline 7.0.0.1 (μ-ARPES at MAESTRO), whole chamber rotations, as in the analyzer-on-boom configuration of BL10.0.1.1 (HERS) of the Advanced Light Source (ALS) or as in the nano-ARPES beamline 7.0.2.2 (*14*) (nano-ARPES at MAESTRO) of the ALS.

However, it has recently become possible to virtualize rotations by steering the photoelectron cloud electrostatically with additional electron optics in the analyzer. This virtual steering approach, called deflector mapping, promises to reduce hardware complexity and to improve acquisition times. Most importantly, deflection mapping permits using a fixed sample geometry. To the reader, this should not be considered a total account of geometries; the writer merely hopes to convey the scope of the problem.

The best way to grapple with these different geometries is to work through an example, which sheds light on how to approach the general case.

1.2.3 A Practical Example: Extracting the Photoelectron Momentum from Measured Angles

An experimental geometry where the slit of the electron analyzer is oriented vertically and no deflectors are used is both straightforward to analyze and commonly instantiated in real hardware. To match the conventions used in PyARPES (*15*), we will call the angle resolved by the analyzer, ϕ, in accordance with the broader set of conventions depicted in Figure 1.2. We will consider a photoelectron traveling with speed $|\boldsymbol{v}|$. Although we must begin work in the frame of the analyzer, Figure 1.1 can be used for intuition. The velocity in the analyzer frame is

$$\boldsymbol{v}_a = \begin{bmatrix} 0 \\ \sin\phi \\ \cos\phi \end{bmatrix} |\boldsymbol{v}|. \tag{1.7}$$

The correspondence between the velocity in the analyzer frame and that in the sample frame, which is directly connected to the emission angles we sought after accounts to unwinding the stack of rotations we must apply to get from the sample frame to the analyzer frame:

$$\boldsymbol{v}_s = R(\chi, \hat{\boldsymbol{z}})R(\beta, \hat{\boldsymbol{x}})R(\theta, \hat{\boldsymbol{y}})\boldsymbol{v}_a. \tag{1.8}$$

The exact relationship depends on the order in which the rotations are implemented by the goniometer and scan hardware. This can be accounted for by changing the order of rotations applied, but in practice it has little bearing on ARPES experiments because most experiments have at least one of $\theta, \beta \ll 1$ in which case the rotations approximately commute with one another. Because $\hbar k_\perp = m_e v_{s,\perp}$ and similarly for the in-plane components, we have a way of applying the original momentum conservation equations. Together, we receive vectors equation for the parallel and perpendicular momenta in terms of the recorded angles

$$\boldsymbol{v}_s = R(\chi, \hat{\boldsymbol{z}})R(\beta, \hat{\boldsymbol{x}})R(\theta, \hat{\boldsymbol{y}}) \begin{bmatrix} 0 \\ \sin\phi \\ \cos\phi \end{bmatrix} |\boldsymbol{v}| \tag{1.9}$$

$$k_x = \hat{\boldsymbol{x}} \cdot \left(R(\chi, \hat{\boldsymbol{z}})R(\beta, \hat{\boldsymbol{x}})R(\theta, \hat{\boldsymbol{y}}) \begin{bmatrix} 0 \\ \sin\phi \\ \cos\phi \end{bmatrix} |\boldsymbol{v}| \right) \frac{m_e}{\hbar}. \tag{1.10}$$

The equations for k_y and $k_z \equiv k_\perp$ are similar. In the $k_\perp$ equation, we need to be aware that $\hbar k_z \neq m_e v_{s,z}$ when $V_0 \neq 0$. The process to derive the equation above is mechanistic: we need to take account for what is measured in the analyzer frame, and push results back to the sample frame by performing rotations in the order appropriate for the apparatus. This approach provides the means, in principle, to analyze any hardware geometry and to obtain an expression for the momenta in terms of the measured angles.

Does this take full account of the connection between the picture presented in the literature and the process undertaken by an experimenter to interpret ARPES data? Unfortunately, not yet. If we abstract notation so that $\boldsymbol{\theta}_m$ is the vector of measurement coordinates, including all degrees of freedom for the analyzer and goniometer, we have shown a process for building the relationship $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{\theta}_m)$. This means that if we record a single electron arriving at coordinates $\boldsymbol{\theta}_m$, we can calculate the momenta for that single electron. The prototypical ARPES experiment does not measure single electron arrivals though: instead, it measures the photocurrent density over coordinates resolved by $\boldsymbol{\theta}_m$.

1.2.4 Consequences of Measuring the Photocurrent Density

ARPES measures the photocurrent density as opposed to single electron arrivals because it is convenient to instantiate this in hardware. The traditional approach to tracking individual electrons requires amplifying each electron arrival into current pulses in a microchannel plate or microchannel plate stack (*16*) before measuring the current in crossed current-carrying delay lines (*17*). The time between current arrival in the delay lines makes it possible to determine which intersection of the lines originated the current pulse. A detector made this way consists of a grid of delay line intersections which act individually like pixels on a camera. A simpler approach does away with complicated timing hardware and instead converts the amplified current pulse by dumping the amplified electron arrival into a phosphorescent screen—made by sintering the popular x-ray scintillating ceramic Gd_2O_2S:Tb (P43) or another phosphor onto a transparent electrode—across a bias exceeding 3kV. By converting electron arrivals into current pulses and then into flashes of light, standard cameras can be placed outside of the vacuum chamber for final detection. The price that must be paid for simplicity is that the frame rate of standard cameras integrates the flashes of light on the phosphor screen from individual electron arrivals with others nearby, and that bloom on the phosphor slightly degrades the spectral resolution of the experiment. This substantiates the statement that the prototypical ARPES experiment, which uses hemispherical analyzers, measures the photocurrent density.

Using the phosphor and camera setup for detection comes with a few other issues which can be remediated with effort. For one, nonlinearity in phosphor and CCD response means that there is

a complex, saturating relationship between the true and the measured photocurrent densities (*18, 19*). If the framerate of the camera is sufficiently high, inspection of each frame can be performed to isolate individual arrival events. By splitting each frame into patches of pixels, algorithms to label individual electron arrivals can be run synchronously with acquisition on commercial graphics processing units (*20*). Practical limits on electron arrival rates are then placed by densities of the electron arrivals on the phosphor, and the decay time of the phosphor glow which range from tens of nanoseconds (*21*) to milliseconds. The second and more serious practical issue with using direct camera readout is that changes in the gain of the current amplification process or with the phosphor efficiency impact the measured photocurrent density. As spectrometers age, the microchannel plates used for current amplification and the phosphor must both be replaced, and the effects of their changing gain must be corrected to normalize the measured photocurrent at acquisition time by sweeping the angle and energy resolved cloud across the detector (sweeping and dithering) or by correcting for the effects post facto.

We started this discussion by pointing out that ARPES frequently does not measure individual photocurrent arrivals and instead measures the photocurrent density. We now know under what conditions this is the case. What's the practical distinction between this and electron event counting? There are two significant downstream consequences. The first can be appreciated by considering how volumetric data is stored in a computer for analysis. The most common representation makes the coordinates along the Cartesian axes implicit. Here, we imagine labeling a function on a regular rectangular grid by indexing with an integer number of steps from the edge. For a rectangular grid, we need only store the number of points in each direction, the location of one corner c_i, and δx_i—the change in coordinate by taking one step in the x_i direction in the volume—to know the coordinates of each point. The values of the function represented can then be stored as a dense set of numbers with no additional overhead. This system works because for all points having the same index n_i along the x_i direction, they have the same coordinate $c_i + n_i x_i$. The parsimony granted by this assumption breaks if we want to change the coordinate representation by a nonlinear transform g, such as the angle-to-momentum transform.

The second can be appreciated by looking at the units on quantities involved. If we are treating the photocurrent, the units on our measurement are those for current, but if we are treating the photocurrent density, we also have inverse units of the quantities we measure over: the angles in the experiment. Our data after conversion to momenta will be in units of current per unit momentum, which means that we need also to incorporate the volume dilation of the coordinate transformation. If we are treating a volume of data V, V can be seen as a function $V(\boldsymbol{\theta})$ which maps to the value recorded at coordinates $\boldsymbol{\theta}$. Then, the standard machinery of differential geometry can be applied so that we can write

$$V(\boldsymbol{\psi}) = \frac{V\big(g(\boldsymbol{\theta})\big)}{\left|\partial g_j / \partial \theta_i\right|_{\boldsymbol{\theta}}}. \tag{1.11}$$

The necessity of Jacobian determinant in the denominator can be seen in two simple ways. Mathematically, we need to account for the volume element when we change variables: $\boldsymbol{\psi} =$

$g(\boldsymbol{\theta})$, and so $d\psi_i = \partial g_i/\partial\theta_j d\theta_j$. Alternatively, we can think of V as a pigmented rubber sheet where the density of pigment represents the value of the density function V. To coordinate change, we distort the sheet and in the process the rubber gets thicker in some areas and thinner in others. In places where g is contractive, the rubber gets thicker and the pigment density increase. The opposite occurs where g is dilative.

These two distinctions—that we assume regular gridded representations of the photocurrent density and that the geometry of the coordinate transform requires a Jacobian correction, are shown in Figure 1.3A, while the procedural steps in Figure 1.3B show how they can both be addressed. The non-intuitive thing is that to move volumetric data $V(\theta)$ to $V(\psi)$, we need the inverse coordinate transform $\theta_i = g_i^{-1}(\psi_j)$. We simply build a regular grid in ψ (Figure 1.3B 1&2) and then pull coordinates back into the recorded volume $V(\theta)$ to determine how to populate the $V(\psi)$ grid (Figure 1.3B 3) after normalizing by the Jacobian determinant.

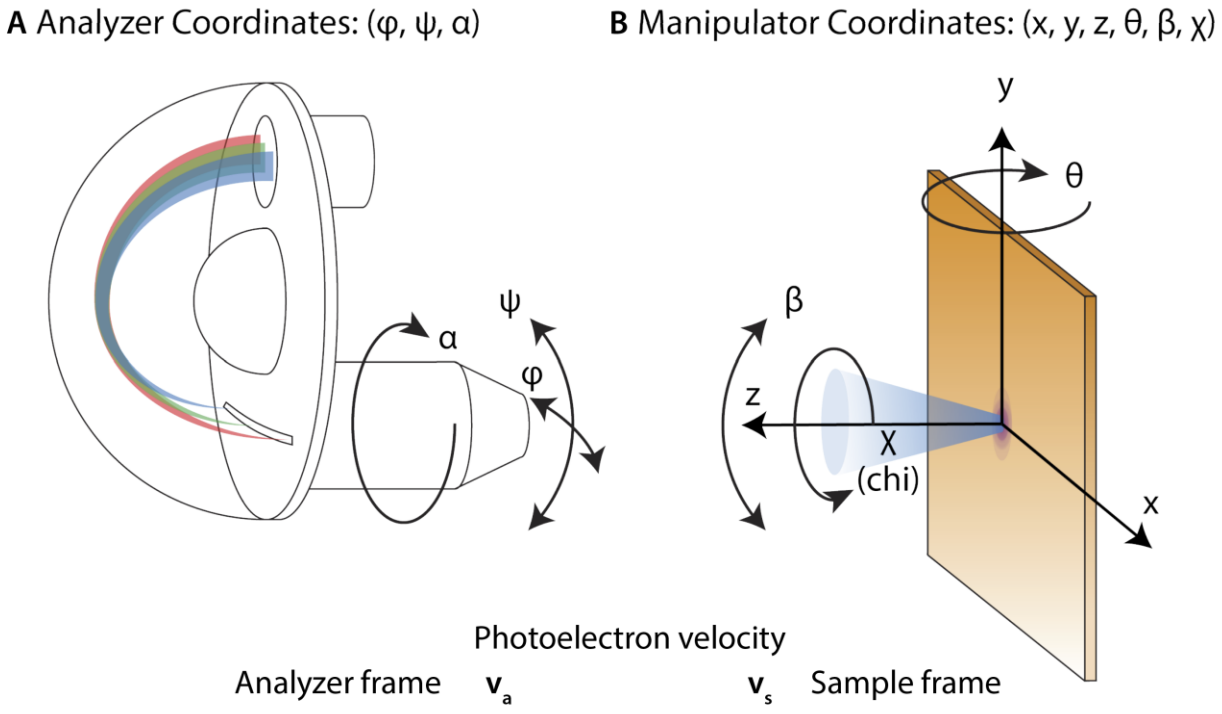


Figure 1.2 **ARPES experimental geometry and coordinate naming convention. [A]** The coordinates assigned for hemispherical analyzers which can be rotated axially around their lens by angle α, measure a finite angle ϕ along the slit of the analyzer and can steer the incident photoelectron cloud perpendicular to the entrance slit of the analyzer by angle ψ. As depicted, the analyzer is operating using a horizontal slit configuration with the value of $\alpha = \pi$. **[B]** Additional coordinates on the sample goniometer or manipulator. Depicted is a six-axis manipulator showing the standard complement of three scanning/spatial axes ($\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$) and three rotational axes. χ rotates the sample about its face, while θ is the rotation around $\hat{\boldsymbol{y}}$ and β is the rotation around $\hat{\boldsymbol{x}}$. As depicted, the three spatial degrees of freedom agree with chosen conventions for the analyzer frame, though they would no longer agree after rotations by either β or θ.

We are now at the point where we can collect the separate threads running through a discussion of what ARPES measures and specify how an experimentalist can sit down at ARPES equipment

and claim to have measured $I(\boldsymbol{k}, \epsilon)$, the momentum resolved photocurrent density. ARPES measures $I(\boldsymbol{\theta})$, the photocurrent density in degrees of freedom of the experiment which may differ across measurement hardware. For the situation above, we determined $\boldsymbol{k}(\boldsymbol{\theta})$ which plays the role of the abstract forward coordinate change g. We saw that to obtain a regular grid of volumetric data $I(\boldsymbol{k}, \epsilon)$, we need the inverse coordinate transform $\boldsymbol{\theta}(\boldsymbol{k})$ and the Jacobian determinant. To conclude our example, we need only invert the coordinate transform we derived previously. To make progress from our example, we will assume the small angle approximation, which permits permuting rotations. Then,

$$\beta \approx -\arcsin\left(\left(R(\chi, \hat{\boldsymbol{z}})\right) \cdot \hat{\boldsymbol{y}}\right) \tag{1.12}$$

$$\phi - \theta \approx \arcsin\left(\frac{\left(R(\chi, \hat{\boldsymbol{z}})\right) \cdot \hat{\boldsymbol{x}}}{\sqrt{1 - \left(\left(R(\chi, \hat{\boldsymbol{z}})\right) \cdot \hat{\boldsymbol{y}}\right)^2}}\right). \tag{1.13}$$

We will return to this process again when we discuss PyARPES, an analysis framework for ARPES. For now, we note that an appendix at the end of this thesis—used for PyARPES—provides derivations and a reference library for the angle-to-momentum (forward) and momentum-to-angle (reverse) coordinate transforms for common experimental geometries. These can be adapted according to the readers' needs if they encounter some novel geometry, or an issue in their computer's implementation of momentum transforms.

1.2.4.1 Modifications for Event Based Data

Event counting data, which usually comes from time-of-flight electron analyzers, consists of an unordered set of arrival events $\{\boldsymbol{\theta}_i\}$. To convert the data, we need neither the inverse transform nor the Jacobian. We can convert directly under g to yield the collect of events in ψ: $\{\boldsymbol{\psi}_i\} = \{g(\boldsymbol{\theta}_i)\}$. It is typical that at some stage, these events will need to be accumulated into a dense array format from the tabular event format, but this can be deferred arbitrarily which removes some sources of analysis complexity we have already confronted.

1.2.4.2 Practical Necessity of the Jacobian Correction

Frequently, it is not necessary to apply the Jacobian correction for analysis. Whether it necessary to apply depends on whether the experimenter cares to compare absolute intensity values in parts of the energy-momentum spectrum which are distant to one another. Most often, ARPES concerns itself with quantitative effects local to one or a few bands, in which case the Jacobian determinant is not changing rapidly in the region of interest. Because of the energy dependence for $|\boldsymbol{v}|$, the experimental necessity to account for volume dilation occurs more universally at low photoelectron kinetic energies—as in DUV and VUV ARPES with flight energies below 10 eV or so. In the DUV/VUV regime however, experimental shortcomings arising from the difficulty in constructing effective lens tables for electrostatic optics at extreme retardation ratios place harsher constraints on experiments. Overall, experimentalists working with DUV/VUV ARPES must account for a broader set of factors which must be controlled to make quantitative statements about the measured photocurrent density.

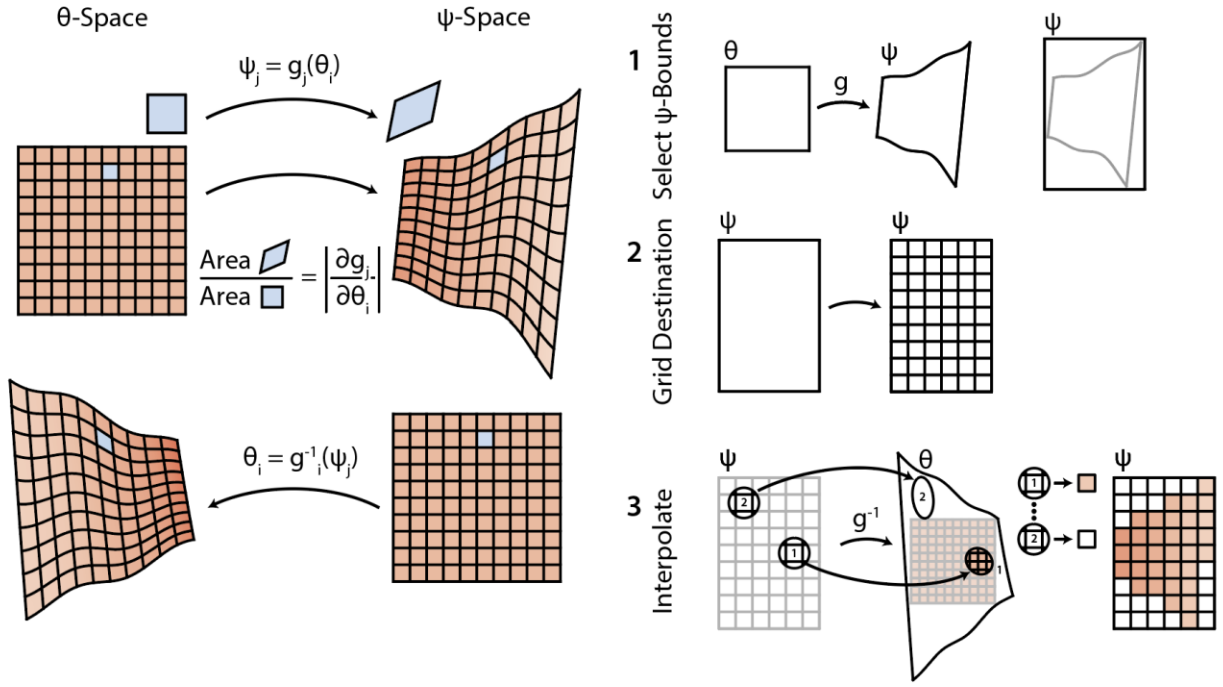


Figure 1.3 **Coordinate Conversion and the Angle-Momentum Relationship in ARPES. [A]** A coordinate transform from θ coordinates to ψ coordinates is defined through a vector transform $\psi_j = g_j(\theta_i)$. A coordinate plane in θ-space maps to a distorted plane in ψ-space (top), as it does going back to θ-space under the inverse coordinate change g^{-1} (bottom). In general, a transform can expand or contract the input space, with the dilation at a point $\boldsymbol{\theta}_0$ being equal to the determinant of the Jacobian matrix of g evaluated at $\boldsymbol{\theta}_0$. The grids above can be viewed as Cartesian axes or as a volumetric dataset. The former perspective is essential to build intuition while the latter has practical relevance for ARPES to understand how we handle experimental data and extract the band structure as a function of momentum. **[B]** The practical set of steps to take a gridded dataset in θ and turn it into an equivalent gridded dataset in ψ. First, we determine the range in ψ which our dataset covers. If we assume that g is injective for all θ represented in our data volume, it is sufficient to convert the edges and corners only, to obtain the outline in ψ, any larger rectangular volume in ψ will contain the image of all the points of our dataset under g. Second, we grid ψ to obtain the empty data volume in the target space at some desirable resolution. This resolution can be picked by heuristic or using the max of the partial derivatives $\partial g / \partial \theta_i$. Finally, we convert each pixel $\boldsymbol{\psi}_0$ to $g^{-1}(\boldsymbol{\psi}_0)$ and use the value of the gridded dataset in this neighborhood to populate the value at $\boldsymbol{\psi}_0$. If you wish to preserve semantics about the dataset, you must correct by scaling by the inverse Jacobian determinant.

1.3 A Practitioner's Overview of ARPES

In the previous section, we discussed some of the history and theory of ARPES, and thoroughly grounded how to think about actual data coming from measurements as it should be analyzed on a computer. In this section, we will discuss practical aspects of actual ARPES experiments which must be understood before you can use the technique to learn about the physics of materials.

Some of the difficulty stems from contradictions and counterintuitive aspects of how the technique is used in practice. ARPES is a local probe but requires extended spatially homogenous samples. ARPES measures only a thin slice on the surface but excels in identification and measurement of bulk electronic states. ARPES produces a huge distribution of photoelectrons but throws away most before measurement and most practical analysis focuses on

an even smaller cohort. ARPES uses electron analyzers which are in ways like cyclotrons and in others like a simple parallel plate capacitor. In this way, ARPES is the same as any other characterization technique in physics, we need to keep in mind both its strengths and weaknesses to bring ARPES measurements to bear on problems of interest.

To make matters more complicated, ARPES has been constantly evolving over the last few decades with the steady removal of hurdles to experiments. Early ARPES experiments were constrained by the brightness and monochromaticity of light sources: undulators at soft x-ray synchrotrons. These gradually improved while the resolving power and fundamental resolution limits to electron spectrometers also improved by several orders of magnitude. With highly monochromatic lasers (*22–24*) and hemispherical electron analyzers capable of resolving below 1 meV (*25*), modern ARPES experiments can access previously unresolvable structure at low energy and can be run on (comparatively) shoestring budgets.

As ARPES experiments have improved dramatically along resolution, they have also multiplexed acquisition by several orders of magnitude (*23*, *26*). Whereas early photoelectron analyzers only collected a one-dimensional vector of data from a fixed angle-space point $\boldsymbol{\theta}_0$ and over a limited energy range, hemispherical analyzers collect two-dimensional datasets simultaneously by dispersing photoelectrons along one angular degree of freedom while still resolving the photoelectron energy. Hemispheres resulted in improvements in acquisition time of ARPES data by around 100-fold, but their longest lasting impacts were qualitative rather than quantitative: the wide adoption of hemispherical analyzers changed how scientists conducted ARPES experiments. Visual inspection of entire two-dimensional energy and angle cuts allowed spectroscopists to rapidly assess surface variations in sample quality, to determine the coarse material phenomenology, and to make better use of the acquisition time on instruments by quickening the time from introduction of samples to the point where desirable high-symmetry ARPES data was being acquired. Currently, a revolution of the same kind and magnitude is ongoing with the introduction of two-angle imaging time-of-flight photoelectron analyzers. These analyzers again improve the throughput of ARPES experiments by a factor of around 100 and permit decision making based on the structure of the constant photon energy Fermi surface, rather than a single energy and angle slice.

Critics of the utility of imaging time-of-flight systems point out that we are rarely interested in the whole Fermi surface. We typically only want the high symmetry directions in the band structure which are one dimension smaller. Hemispheres have traditionally excelled at exactly this function. Despite this assessment, time-of-flight systems are perfectly adapted to the material systems poised to preoccupy the technique over the next half decade. Superlattice systems, particularly long wavelength ones, put additional emphasis on the Fermi surface because the interaction with the superlattice potential $V_s(\boldsymbol{q}, \epsilon)$ creates new photoemission intensity offset by wavevectors $\boldsymbol{q}$ where V_s is large. These novel spectral features are most simply captured by collecting the whole Fermi surface simultaneously. Another reason stems from pragmatism: reversible electrostatic control of sample doping via *in operando* measurements requires biasing sample regions and modifying the electrostatic configuration seen by the vacuum photoelectrons. This distorts the photoelectron cloud so that if the region collected by a hemisphere was a high

symmetry one at doping level n_0, it is not likely to remain high symmetry if the doping is adjusted to $n_0 + \delta n$ (*27*). Once again, capturing a 2D region containing the 1D momentum directions of interest solves this problem. Finally, as we will see in the chapter on TMDCs, access to the entire spectral function allows for analysis to better constrain values of parameters on symmetry axes—for example, by global fitting.

1.3.1 Two Models for Photoemission

A practical way to keep in mind the strengths and limitations of the technique is consider the photoemission process. *Pro forma*, this is done by obtaining and reading a recent review article on ARPES by professor Z. X. Shen (*7*), my advisor Alessandra (*28*), professor Damascelli (*8*), or one of countless others depending on the technique and material specialization of your choice. When you do this, you will learn of a nested hierarchy of theoretical pictures of the photoemission process.

The simplest picture is of energy and momentum conservation which describe the transition probabilities from initial to final states in the photoemission process according to Fermi's golden rule. Under this view, ARPES is a beautiful and simple band mapping technique. This level of sophistication also grants practitioners a useful weapon. The dipole matrix element M^2 permits hiding and eschewing complexities in real experimental data.

Read a few more paragraphs and it is explained that ARPES does not merely measure the location of infinitely narrow tight-binding like bands, but instead the angle-resolved photocurrent $I(\boldsymbol{k}, \epsilon)$ is related to the single-particle spectral function $A(\boldsymbol{k}, \epsilon)$ and the electron self-energy

$$A(\boldsymbol{k}, \epsilon) = -\frac{1}{\pi} \frac{\mathrm{Im}\Sigma(\boldsymbol{k}, \epsilon)}{\left(\epsilon - \epsilon(\boldsymbol{k}) - \mathrm{Re}\Sigma(\boldsymbol{k}, \epsilon)\right)^2 + \left(\mathrm{Im}\Sigma(\boldsymbol{k}, \epsilon)\right)^2}. \tag{1.14}$$

Read further and you see that this is motivated and substantiated in treating ARPES, like optical response and tunneling, among many other techniques, in linear response theory thereby making a direct connection to the Green's function $G(\boldsymbol{k}, \epsilon)$.

We will avoid a discussion at this level of complexity going forward. Other authors handle and develop the connection between many-body effects exceptionally elsewhere. Additionally, the rigorous and valid application of the self-energy connection requires confronting a second model for photoemission. This second model is less about rigor: it is a mental picture which everyone who engages with photoemission for long enough inevitably builds through the trials and faults of performing ARPES experiments. Unfortunately, it is impossible to fully develop this model in any other way than through experience. ARPES—like so many other measurement techniques of its complexity—is riddled by complex hardware-sample interactions. Each apparatus is inflected by a unique personality which superstition suggests imprints from whichever scrappy group designed and built it.

This intent of this section in the thesis is in part to point out to the reader details which the author noticed and incorporated into their mental picture. If the reader is at the start of their journey with photoemission, this document can serve as an accelerant to developing the reader's mental

model as they gain experience through a first set of experiments. If the reader is an expert, they should feel free to deride its contents for being crude or stylistically different than the picture they drew when they learned ARPES.

Mental models are hierarchical, and there are many transpositions of the second model for photoemission. We can spite Pendry (*29*) and take a "chronological" view of the photoemission process. First, we consider what happens in the initial state due to problems in sample preparation, in surface quality, and in vacuo degradation. Next, we consider what occurs during the photoemission process itself and in the propagation toward the sample surface. Before detection, we must next consider what occurs on the electron's way through vacuum. Finally, we consider what happens through resolving the photoelectron energy and marking it in the experimental record through detection. Other transpositions of this mental model which may be useful to consider during experiments are initial state/final state, static configuration/dynamic configuration, and photon-induced/electron-induced among others. Keep these alternative organizations in mind, even as we take a roughly chronological catalog of effects.

Finally, the catalog of relevant effects we document below is woefully incomplete. This is by necessity, since I promised only to document major components of the model for ARPES I developed, and no graduate student spends enough time with the many subdisciplines in ARPES to develop deep expertise. I have a reasonable idea of where shortcomings lie, however. I won't discuss complicated pump interactions like Volkov and Floquet states, the intricacies of pump-probe photoemission, the importance of the photoexcited final state, or the sudden approximation.

For brevity, I have also neglected to include a discussion of the resolution and aberrations in hemispherical analyzers. These are generally excellently described in the manuals accompanying your analyzer. Furthermore, calibrating the recorded angles and energies of your spectrometer is probably best handled empirically rather than by recourse to a theoretical model of the lenses—details here depend on manufacturer, pass energy, and are exacerbated at extreme retardation ratio which means that the situation is extremely complicated for UV ARPES.

1.3.2 Disorder

When we develop tight binding theory and the band structure in elementary solid-state physics, we operate on the basis that the material in question has a characteristic set of symmetries. The word characteristic is doing overtime here because these symmetries help set the character of the material, and because they place the material in one of many different crystal families which share those symmetries and confluent material properties.

Strictly speaking, any disorder at all breaks these perfect symmetries, and yet, while there are almost no perfectly symmetric systems that come across the desks of physicists this does not pose a problem. The most common type of disorder which affects the work of photoemission scientists is chemical disorder in the form of vacancies, intercalants, substitutional defects, and their like. One perspective to take is perturbative: seen from the perspective of an electron in an eigenstate of the parent material, the introduction of the disorder creates a small potential $\delta V(\boldsymbol{k}, \epsilon)$ which modulates and can scatter the electron. Not surprisingly, if we talk about diffuse

disorder, there's no reason to expect that the location of one vacancy affects the site of any other. This means that δV is mostly peaked around zero momentum and does not have any other peaks to speak of: its affect on the photoemission observable will be to smear the photoemission signal in accordance with the different local environments seen by the photoelectrons we happen to record.

This stochastic aspect of things shows up everywhere in photoemission, as we will see. We cannot really get around measuring averages of quantities and averages show up even in unexpected places, for instance in the binning in the transverse momentum set by a choice of slit size on a hemisphere. It also points out one way of trying to understand the effect of disorder which consists of taking the empirical average of the spectral function—or any other property—over different disorder realizations $\langle A(\boldsymbol{k}, \epsilon)\rangle_D$. As we will see in the discussion on spectral function unfolding in moiré superlattices, this can be done by performing a calculation like DFT or tight binding over a supercell, projecting states back to the parent Brillouin zone via unfolding, and finally averaging over different configurations sampled, as necessary to capture relevant phenomena. Alternatively, when we take the free electron final state approximation, the ARPES intensity is essentially the Fourier transform of the LCAO states (*30*), which explains why ARPES measurements retain their character even when disorder or broken symmetries are present.

1.3.3 Surface quality

At low electron kinetic energies, like those used for ARPES, the photoemission process is very surface sensitive due to survivorship bias. Soft X-rays have very long comparative penetration depth—100 eV corresponds to 12.4 nm from the photon dispersion relation—but electrons which are excited deep in the material rarely make it to the surface without inelastically scattering away the energy the photon granted (*31*). This surface sensitivity has made ARPES exceptional for distinguishing surface from bulk physics, for proving topological phases (*32–35*), and for application in the study of two-dimensional materials (*36–38*). The same surface sensitivity burdens study design by requiring an exceptionally clean surface for photoemission. Other techniques, like scanning microscopies and scanning tunneling microscopy, have even greater requirements on surface cleanliness but study only a minute region of the sample surface.

ARPES has only high rather than exceptional local requirements, but it makes these requirements over as large an area as is covered by the incident light used for photoemission. This means that "surface quality" covers not just the density of atomic defects on the surface, but also morphological features of the sample. In the past, samples uniform over a square millimeter were required. In 2022, you can get away with uniformity over a square micron or so, so long as you consign yourself to measure at one of only a small handful of facilities around the world (*14*, *39*, *40*) and are willing to accept concessions to spectral resolution and total electron gathering power. Otherwise, samples are with a chemically and morphologically uniform region of around 1×10^{-3} mm^2 will do. The author has seen many samples where even this requirement is strenuous despite invitingly uniform presentation under an optical microscope. An unwelcome and common surprise is to find that a cleaved crystal surface is not even flat (the normal vector $\hat{n}$ varies over the sample surface) which imprints by smearing the emission angles η and φ roughly

by $\sigma(\hat{n}) = \sqrt{\mathrm{Var}(\hat{n})}$. This rather simple issue, taken together with the nonlinearity $\boldsymbol{\theta}_m(\boldsymbol{k}, \epsilon)$ means the resulting broadening is even dependent on photon energy. Many issues of practical significance in ARPES are caught up in the geometry of the measurement for this reason which complicates running experiments. At the same time, the dependence on geometry also provides a handle with which certain issues can be wrangled or suppressed: if surface flatness issues are expected, for instance, measuring at lower photon energies will help.

Other possible issues with surfaces abound. A crystal may have several cleavage planes in the unit cell. A novice may assume these are all parallel with one another so that the normal emission directions are parallel, but even this cannot be taken for granted for some crystals. Different cleave surfaces may be more reactive in general due to their surface composition, may be more reactive to certain gases present in the measurement chamber, or may be more easily damaged by incident x-rays. Certain cleaved surfaces may consist of species whose orbitals do not contribute to pertinent states the experimenter hoped to resolve. Crystals may cleave along imperfections in the lattice where another phase, or even disordered or polycrystalline nonsense, grew. ARPES thankfully combines symmetry sensitivity via the Brillouin zone and chemical sensitivity via core level spectroscopy that most of these issues can be discovered, if not handled, at experiment time.

Leaving the realm of cleaved crystals, the diversity of effects grows. TMDCs, where via accident the author has spent most of their time, are forgiving but nonetheless instructive to the character of exfoliated material and device surfaces seen by photoemission. This discussion will be deferred to the discussion of WS_2/WSe_2 moirés. Certain materials present additional, unique challenges: they may grow thin, grow narrowly, grow embedded in a host crystal, or may be available only in a thin-film.

1.3.4 Surface aging

Surface quality is not static. ARPES experiments frequently occur on relatively short timelines, and not just because time on beam at user facilities and synchrotrons is scarce. The measured surface typically ages during measurement and this may place a barrier to indefinite collection of high-quality data from a single sample.

The most significant cause of damage to the surface during measurement is due to imperfect vacuum. Applying the kinetic theory of gases, the areal flux of particles F is given by

$$F = \frac{\langle v \rangle n}{4}, \tag{1.15}$$

with number density n and mean velocity $\langle v \rangle$. Although we have not derived it here, the proportionality constant $1/4$ can be derived or recovered simply via Monte Carlo. Plugging in $\langle v \rangle = \sqrt{8k_BT/(\pi m)}$, again from kinetic theory and applying the ideal gas law we can express F in terms of the pressure P:

$$F = \frac{P}{\sqrt{2\pi m k_B T}}. \tag{1.16}$$

Even if our sample is cold, the gas in the chamber is hot. Plugging in representative numbers for the areal density of a monolayer of particles, mass of relevant species, and room temperature we arrive (roughly) at the common rule of thumb for surface exposure: a monolayer takes 1 second to form at a pressure of approximately 1×10^{-6} torr (one Langmuir of exposure). This units are impractical for photoemission: at 1×10^{-10} torr, the formation time for a monolayer of vacuum chamber cruft is a few hours. Base pressures in UHV chambers for ARPES range from 2×10^{-11} to 5×10^{-10} torr, but this simplifies matters too much because it ignores gas composition. In clean chambers, that 2×10^{-11} constitutes only hydrogen which is less reactive and damaging than water, hydrocarbons, or whatever the last group grew and measured. Nevertheless, we can't expect photoemission measurements to last longer than 24 hours without resilient samples. By stroke of luck, this is about the limit of young scientists' endurance for a single session anyway.

The consequence of this discussion is deeper because it influences study design. If ARPES feels like a race against the clock, it's good to reserve some time before passing the torch to the next experimenter to double check initial findings lest interesting phenomena measured towards a sessions end go undiscovered as surface aging.

Surfaces can also age because of temperature dependent effects. That particle flux just discussed sticks so well because we measure cold. ARPES uses in vacuo goniometers which are colossally large. When they're cold, they pump the chamber and help preserve the sample, but if we heat the sample and the manipulator with it, all the adsorbed detritus previously cryo-pumped can find its way to your sample's surface. This is a factor in determining how to do temperature dependence studies: return to the original temperature to verify there are no aging effects and consider outgassing hardware over the temperature range you need in advance.

A final set of surface aging phenomena concern the beam interaction. Plenty of species in vacuo are reactive enough to damage your sample on their own, but some of the more inert ones can be more vigorous under exposure to a bright soft x-ray beam. This leads to a host of interesting phenomena where UV or x-ray exposure can dope samples (*41*), even over the course of seconds (*42*, *43*). Even without residual gases, the x-rays can damage your sample directly. Whether this is relevant depends on the material in question, the energy of the x-rays or UV, and their intensity (meaning nano-probe ARPES exacerbates these issues (*44*)).

1.3.5 The light source

Having discussed many of the possible issues with the surface we want to measure, we can turn to the light we use for the photoemission process. Unlike sample issues which may differ between each sample, the character of the laser or undulator you use for photoemission is essentially fixed. From the standpoint of spectroscopy, the most important feature of the light is how monochromatic it is. Many light sources produce a monochromatic beam by selecting a narrow band from a broadband source which permits tuning the photon energy. Undulators at third-generation storage rings behave this way, and the improvement in brightness of synchrotron-based sources is a source of major improvement in the resolution for ARPES. Brighter beams can deliver more photons to the sample while operating with a narrower

passband. For ARPES, the effect of an incident photon linewidth of $d\omega$ is to broaden the spectrum—by convolving with the actual distribution of intensity by $d\omega$—during the actual photoemission process. Some types of monochromators, like diffraction gratings, do not completely suppress light at frequencies $2\omega, 3\omega, \ldots, n\omega$. Although the delivered flux of these harmonics is typically very low, the cross section against atomic core levels can be very high and serendipitously can cause the kinetic energies of a core level struck by a harmonic to have similar kinetic energy to the valence region of interest.

In practice, this linewidth can now be chosen in most cases to be narrower than other sources of resolution broadening, although few meV total resolution from all sources remains a challenging target due to the difficulty in delivering a high photon flux necessary for operating electron spectrometers at low total resolution broadening while offering a highly monochromatic beam.

UV lasers suitable for photoemission come in both continuous wave and pulsed varieties and offer the ability to do ultrahigh resolution spectroscopy or study electronic dynamics, respectively. For a Gaussian pulse, the time-bandwidth product specifies that

$$d\omega d\tau \approx 1824\ (\mathrm{meV} \cdot \mathrm{fs}). \tag{1.17}$$

This is a helpful rule of thumb in considering the tradeoffs between time- and energy-resolution in photoemission. Using 30 fs pulses short enough for resolving excited electron dynamics means paying around 60 meV for energy broadening from linewidth alone. In fact, the situation is somewhat worse on both accounts. In pump-probe experiments like time- and angle-resolved photoemission spectroscopy (TARPES) the effective resolution is set by the quadrature sum of each of the pump and the probe which degrades time resolution further than the bandwidth of the probe would suggest on its own. This is often not a serious problem, as the pump can be made temporally narrow than the probe. On the other hand, short pulse duration probes exacerbate space charge effects between outgoing photoelectrons, to be discussed later.

1.3.6 “Two photon” effects in the photoemission process

In pump-probe photoemission, it is possible for photoemission to occur only from the infrared pump beam even if $h\nu_{\mathrm{pump}}$ is much less than the work function, Φ. This can occur because of direct multiphoton effects when the incident pump power densities are in the range of $1\mathrm{GW/cm^2}$ (*45*), as happens when even the output of an oscillator is focused to $50\mu\mathrm{m}$ spot for TARPES.

In the context of pump-probe photoemission, multiphoton effects between the pump and probe also occur, giving rise to laser-assisted photoemission (LAPE) where a UV photoemitted electron interacts with the infrared field and creating sidebands of the photocurrent offset by the infrared photon energy (*46*). This provides a kind of autocorrelation measurement in photoemission permitting the accurate determination of temporal overlap for dynamics studies. LAPE is, from this perspective, the reverse process to the scenario of interest in TARPES where hot, IR-excited carriers are photoemitted by the UV probe.

1.3.7 Propagation of the energetic electron through sample

After excitation, the photoelectron must make its way to the surface if it is to be seen in the experiment. On the way, electrons can scatter elastically and inelastically against the lattice, against imperfections, and against superstructure in the sample. Scattering or diffraction against the lattice and superstructure can create sidebands or replicas of the photocurrent signal offset by the momentum change (*47–49*). In the case of the lattice, this means scattering from zone to zone and which leaves the observed photoemission signal essentially unchanged, but in the case of superstructure another set of scattering vectors $\boldsymbol{q}_s$ are involved. For now, we will leave superstructure unconsidered. Inelastic scattering can rob an electron of the energy required to overcome the work function and cause it to reflect against the vacuum terminated surface. Inelastic scattering can also multiply the number of low energy electrons emitted from the sample surface which exacerbates space charge effects and creates a very intense, nondispersive background away from the Fermi level. The inelastic mean free path λ_{MFP} specifies the probability that an electron won't scatter after propagating a distance λ

$$P(\lambda) = e^{-\frac{\lambda}{\lambda_{MFP}}}. \tag{1.18}$$

Much has been considered about the inelastic mean free path and the predominant model for its character is that it is photon energy *dependent* but material *independent*, which is summarized by the universal curve (*31*). In truth, ARPES mostly considers the inflection point of the universal curve, it studies materials which were not in the same class as those used to prove out the universality of the model, and it operates squarely at the absolute minimum λ_{MFP}. For these reasons, a conservative attitude towards the mean free path is probably warranted, unless you have strong prior knowledge of how large it should be for your sample.

Remarkably, there are other consequences of inelastic scattering which a photoemission expert should be aware of which extend beyond extinguishing the photocurrent signal. Inelastically scattered electrons can make it into vacuum but scattered to new energy and momenta. When measured, these electrons constitute a diffuse background signal whose intensity increases at lower photoelectron kinetic energy. The second effect concerns momentum resolution. The finite mean free path means we measure a λ_{MFP}-thin pancake on the sample surface and must pay for our lunch: $k_\perp$ is broadened proportionally to $1/\lambda_{MFP}$. For very short mean free paths at soft x-ray energies which are highly dispersive along the perpendicular axis, this may be a dominant source of resolution broadening. Consider: $1\ \mathrm{nm}^{-1}$ is much larger $0.06\ \mathrm{nm}^{-1}$, a conservative estimate for the parallel momentum resolution[1]. Furthermore, every source of momentum broadening is also a source of energy broadening, via the band velocity.

1.3.8 Vacuum electrostatics and magnetostatics

Although the journey from the surface toward the analyzer may seem straightforward, there are many details which affect the recorded photocurrent signal. First among these are Coulomb

[1] A 50 eV photoelectron carries $36.2\ \mathrm{nm}^{-1}$ of momentum. Using the small angle expansion and an angular resolution of 0.1^{o} gives $36.2\ \mathrm{nm}^{-1} \times 0.1^{\mathrm{o}} \times \frac{2\pi}{360^{\mathrm{o}}} = 0.06\ \mathrm{nm}^{-1}$. The situation is even better at more resolving angular modes or at lower kinetic energies.

interactions between the photoelectron, other electrons in vacuum, and the sample. These two interactions are referred to as the space charge effect and mirror charge effect respectively. Quasi-CW and pulsed sources may cause the emission of not one but several photoelectrons from the sample surface in rapid succession.

The photoelectrons emitted concurrently Coulomb interact while propagating in vacuum before their different emission angles and emission energies disperse the photoelectron cloud. Pretending that the spatial and temporal profiles of the UV beam are tophat rather than Gaussian, just as the electrons are leaving the surface, we can imagine expected charge density as occupying a pancake at the surface. The pancake's shape is given by σ_x and σ_y—surface coverage of the probe beam in two directions—and $\langle v \rangle \tau_{\mathrm{pr}}$ in the direction normal to the surface. Here, $\langle v \rangle$ is a representative velocity for the photoelectrons and τ_{pr} is the duration of the probe pulse used for photoemission. Recalling Figure 1.1, at 6 eV photon energies or around 1 eV kinetic energy, photoelectrons travel in vacuum roughly at $\langle v \rangle \approx 593$ km/s. At 60 eV photon energies or around 55 eV kinetic energies, photoelectrons travel[2] at $\langle v \rangle \approx 4399$ km/s. The prototypical 6 eV source uses Ti:Sapph lasers with pulse durations of 50 fs or higher. At synchrotrons, pulse durations vary widely but are closer to 100 ps.

A good approximation to the distance between two electrons in the pancake its diameter. We can further approximate the diameter of the pancake by the largest dimension $d \approx \max\left(\sigma_x, \sigma_y, \langle v \rangle \tau_{\mathrm{pr}}\right)$. Two electrons at a distance r have an electrostatic energy

$$E = qV = \frac{e^2}{4\pi\epsilon_0 r} \tag{1.19}$$

which amounts 1.44 meV · micron. Half this energy is shed to each of the participating electrons in their center of mass frame as they disperse in vacuum. When the pancake is flat, this has the tendency to accelerate the faster of the two electrons and decelerate the slower one. How does this scale approximately when there are more electrons? The situation for electrons in the middle of the cloud is far from clear: some electrons are "in front" of it and others behind so it is not sufficient to anticipate that the space charge energy broadening δE_{sc} scales proportionally to the number of electrons per pulse, n. In a line of n electrons, the typical electron has $O(n)$ more of them on one side than the other. On balance, this gives the scaling factor for the broadening, so that very approximately

$$\delta E_{\mathrm{sc}} d \approx n \text{ meV} \cdot \text{micron}. \tag{1.20}$$

This picture is a little naïve for two reasons. First, it turns out that the scaling dynamics for the mirror charge are different than for space charge. This means that the space charge effect can be partially compensated by the mirror charge effect in certain regimes of pulse duration. Second, this is too rosy a picture because the factor n is not rigorously correct. If a more precise picture of the influence of space charge broadening is required, its effects can easily be simulated in

[2] In both cases, sufficiently nonrelativistic that we can use $\langle v \rangle \approx \sqrt{2(h\nu - 5 \text{ eV})m_0}$.

Monte Carlo in the appropriate emission geometry. Nonetheless, this gives some practical limits on experiments which are useful to keep in mind. Because the effective diameter of the photoelectron cloud $d \approx \max(\sigma_x, \sigma_y, \langle v \rangle \tau_{\mathrm{pr}})$, space charge effects are self-limiting in the nano-ARPES regime and in the ultrafast pulse regime when considered separately.

Conversely, conducting pump-probe studies with very small spot sizes requires effort to stay in the one electron per pulse regime[3] where broadening doesn't occur. Space charge effects are also significant relative to other sources of broadening even with moderately large beams when used for femtosecond pump-probe studies.

Other vacuum electrostatic effects arise in unique contexts. With the recent introduction of electrostatic gates for studying doping transitions and electric field induced symmetry breaking nanoscale devices, exposed low voltage contacts close to the photoemission region can lens photoelectrons leaving the sample. Consider a photoelectron emitted normal to the sample surface with energy E_k at a distance d from a contact at voltage V. The photoelectron follows the parametric trajectory $vt\hat{y}$. Its distance to the contact is

$$r(t) = \sqrt{d^2 + v^2 t^2}. \tag{1.21}$$

Over a period $\tau = d/v$ the electron is about distance d from the contact and is accelerated laterally with force approximately qV/d . As a result, the photoelectron's momentum is perturbed such that $\delta k_x \approx Ft/\hbar \approx -\frac{1}{\hbar}\frac{d}{v}\frac{eV}{d} = eV/(\hbar v)$. At the soft x-ray photoemission velocity of 5000 km/s and using a contact voltage of 5V, this gives $\delta k_x \approx 1.5\ \mathrm{nm}^{-1}$! The real deflection will be somewhat smaller, due to screening by metallic regions on the sample surface and faster falloff of the dipole field. Nevertheless, we can conclude that the local electrostatics matter a great deal for photoemission if we hope to recover the electron's crystal momentum in the initial state. Further complexities arise because electrons which are not normally emitted will come closer or further to any sources of fields in vacuum, meaning that $\delta \boldsymbol{k}$ is dependent on the emission parameters (E_k, η, ϕ), and where under the UV beam the electron was photoemitted from. In the case where $d\delta \boldsymbol{k}/dx$ is large, then this may contribute substantially to momentum broadening observed in the photocurrent signal. Local electrostatics may be influenced by different material work functions in the vicinity of the sample, transient charge distributions created by the pump beam, and static charge accumulation arising from photoemission from nearby insulators or poorly grounded pieces of sample. At UV energies, these problems are exacerbated roughly according inversely to v (to $E_k^{1/2}$) as is seen in the discussion above.

Although we have estimated the seriousness of effects stemming from vacuum electrostatics here, it is not difficult to make estimates with a higher degree of realism to your experiment. For

[3] The number of electrons per pulse is Poisson distributed around a central value λ proportional to the incident UV pulse energy. To prevent broadening, we must suppress the product $(1 - e^{-\lambda}[1 + \lambda])\delta E_{\mathrm{sc}}$ where δE_{sc} is the anticipated space charge broadening for two photoemitted electrons. After expanding and throwing away constants, this means we must keep $\lambda^2 \delta E_{\mathrm{sc}} \ll \delta E_{\mathrm{ins}}$. Satisfying this requirement demands that pump-probe experiments be conducted at extremely high repetition rates—as are available from Ti:Sapph and ytterbium fiber oscillators—for photoemission.

instance, in many practical scenarios the geometry is close enough to a high symmetry situation that the electrostatic problem can be solved analytically, and even in the absence of symmetry, numerically integrating the electron trajectory is tractable.

1.3.9 Hemispheres: Final detection

To record the electrons after resolving their energy and angle, we need some way of detecting their arrival at the exit of the analyzer. In 2021, this is mostly done indirectly by converting the angle and energy resolved photocurrent into light. First, the individual electrons are accelerated over a moderate potential at the entrance to a micro-channel plate (MCP). The MCP is biased front to back with a potential of around 1kV per plate and may consist of two or three plates stacked in series. Each plate consists of bonded, extruded glass pores cut off-angle and acts like an array of photomultiplier tubes sharing a common ground current (*16*). When an electron hits the interior wall of one of the pores, it creates secondary electrons which are accelerated by the bias potential which collide with the pore wall several times. An MCP or MCP stack therefore acts as a location sensitive current multiplier[4].

1.3.10 At analysis time

Finally, decisions made during analyzing ARPES data can have ramifications on the effective resolution obtained by the experiment. Although it is reasonable with event-counting datasets produced by delay lines and ToF analyzers, analysis rarely operates on collections of individual counts and instead operates on histograms of the photocurrent density resolved by experimental parameters like the electron energy and momentum. To report the binding energy of a feature in an ARPES spectrum, the experimenter needs to integrate over some momentum region $\delta\boldsymbol{k}$ to produce the energy-resolved photocurrent density inside that window. In the case where the band velocity is large, this results in an equivalent energy broadening which can distort or even move the center of the lineshape under study.

On the other hand, setting $\delta\boldsymbol{k}$ to a small value during analysis does not avoid the problem, because it can also subtly degrade resolution by resulting in a selection where the statistical uncertainty is large enough that it affects downstream analysis. Treating analysis in the case of under-sampled spectra is difficult, especially when the correspondence between Poisson distributed electron arrivals and the image intensity is unknown. The best that can be said in general is that in the regime where binning (integration on one of the coordinates of the density) is necessary for robustness of results, the effect of the binning should be studied to ensure it is not a dominant term.

[4] A single MCP may have gain around 10^3 or 10^4, and therefore a single electron in a stacked MCP can be multiplied into a large charge bunch of 0.1 pC or so. If there is perfect efficiency of the energy transfer from 3 keV electrons to 3 eV photons and the charge bunch illuminates a square millimeter of phosphor, a single pixel with 5 μm pitch receives 10^4 photons which is more than sufficient for detection after considering CCD/CMOS sensitivity and typical phosphor conversion efficiency.

2 PyARPES

A sea-change has affected the way in which scientists conduct condensed matter physics. As the discipline of condensed matter physics matured, revolutions in the phenomenology of materials have grown rarer, though they still occur. A desire to realize materials with superlative performance by a particular metric, for instance ever hotter superconducting transition temperatures, has been subsumed by a pursuit for general understanding. Instead of instantiating a high-T_c material, we want to know what common factors underpin the high-T_c phenomenon. This is one of many shifts in focus which material physicists hope will unravel the field's enduring challenges. Their solution offers the opportunity to avoid a meticulous search through the space of all materials. Instead of hill-climbing from cuprate to cuprate, we could jump to paragon materials capable of delivering on the dream of energy and compute applications we have discussed for decades.

The way in which we conduct experiments has changed to match our changing goals. To find a material with a higher superconducting transition temperature, we only needed to grow the material and measure its resistance. Yet superconductivity, like all complicated material phenomena, does not exist in a vacuum. In the complex cuprates and iron-based superconductors, it accompanies countless other phenomena which may elevate transition temperatures or antagonize the formation of superconducting phases. This material reality necessitates that scientists leverage more sophisticated and more numerous experimental probes to build a core understanding. Experimentalists now combine measurements of the same materials and samples from several international user facilities, or combine insights made by correlating the phenomenology observed in several distinct experimental probes at the same university.

2.1 Practical factors driving the necessity of common data analysis software for ARPES

This abstract discussion has direct ramifications on who performs ARPES and how ARPES experiments contribute to our understanding of materials. While the photoemission community was more siloed in the past, scientists who do not consider ARPES their primary research tool now collaborate on photoemission experiments or collect data themselves. Friendlier user facilities for ARPES, nano-ARPES, and even extreme ultraviolet TARPES simplify and even democratize access to photoemission measurements.

Simultaneously, ARPES has advanced along the practical dimensions which a spectroscopist might quote to third party, curious what photoemission might tell them about their sample. Spatial resolutions have improved by several orders of magnitude and now push into the sub-micron realm (nano-ARPES), ARPES can without so much compromise on photoelectron yield resolve the photoemitted electron spin (SARPES), and in combination with pump-probe spectroscopy techniques, ARPES delivers insights into electron and structural dynamics. Improvements in the resolving power of hemispherical electron analyzers, the introduction of

two-angle imaging “momentum microscopes,” DUV lasers, and brightness increases at synchrotron X-ray sources improve the spectral resolution of all these downstream techniques. To name just a few current applications, high-quality ARPES data can be collected from two-dimensional materials (*36*, *37*, *50*, *51*), operating nano-electronic devices (*27*, *52–54*), from buried interfaces (*55–57*), from superlattices with momentum structure stretching the limits of the instrumentation resolution (*58–60*), and from emerging and familiar quantum materials (*7*, *61–63*).

Yet collecting high-quality data comprises only a narrow part of the scientific process. ARPES, more than other techniques, relies on complicated and time-consuming data treatment and analysis to turn its multidimensional datasets into physical insights. While experiments have grown more accessible, the tools used for ARPES analysis remain esoteric, proprietary, and challenging to use for photoemission outsiders. We saw this as an opportunity to develop the software tools paralleling instrumentation advancements which could ensure photoemission remains a core and accessible part of the condensed matter physics toolset. To date, there have been extremely limited efforts to develop open and modular data analysis frameworks which are convenient for the high-dimensional photoemission data experiments record. Despite this, the neighboring disciplines of scanning probe microscopies (*64–66*), super-resolution microscopy (*67*), and x-ray science (*68*) have demonstrated the viability and usefulness of this approach in advancing the scientific impacts of their respective techniques, and signal an opportunity for letter more scientists drawn insight from ARPES if they are granted appropriate tools.

Our requirements for a data analysis system centered on the trend towards multimodal experiments factored under four principal headings. First, and most directly, ARPES analysis requires uniform and predictable representations of data collected across instruments and facilities. Second, we should seek automation and fault tolerance of essential data preparation procedures and the analysis methods common to all applications of photoemission. Third, a new and broader cohort of scientists using ARPES require highly documented codebases which can introduce them to ARPES and guide them through the application of data processing routines. Fourth, the data analysis system should as best as possible remove the tension between careful analysis and exploratory data analysis to encourage a higher standard of rigor in data treatment. Finally, a modern data analysis system for ARPES is a foundation providing primitives for next-generation experiments which combine machine learning and theoretical modeling approaches constrained by real spectral data.

The net result of this effort was the development and release of PyARPES, a Python data analysis framework structured around typical photoemission spectroscopy analyses and workflows. PyARPES is distributed as open source, and relies heavily on the PyData software stack as a backbone, allowing integration into a large variety of cross-discipline analysis tools, including robust image processing and basic machine learning support provided by scikit-image and PyOpenCV (*69*). Open source and industry standard tools for data science including SciPy (*70*) for linear algebra and signal processing, numpy (*71*), xarray (*72*), dask, and pandas for numerics support, and Matplotlib and seaborn for plotting are used internally by PyARPES to make working in the environment familiar and productive. With the advent of DFT calculator

packages and physics libraries in Python, including the Atomic Simulation Environment (*73*) and Python Materials Genomics (*74*), PyARPES offers a data analysis toolkit for experimentalists complementing those for theory and simulation. PyARPES is more than just a collection of analysis routines however, it is also a platform to perform interactive data exploration and analysis with minimal friction, enabling the sophisticated analysis of photoemission data during data acquisition which can guide and constrain ongoing experiments.

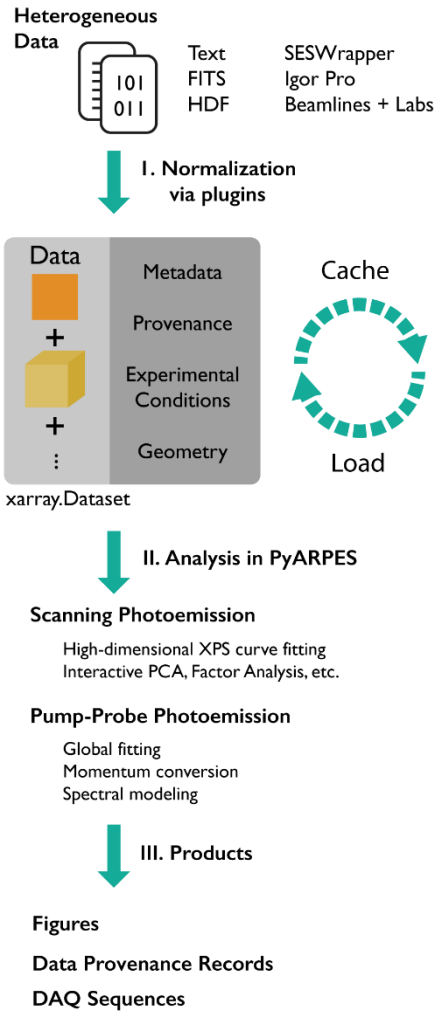


Figure 2.1 **Overview of the data analysis process in PyARPES showing [1]** the data loading and supported formats, **[2]** selected examples of analysis facilities for nano-ARPES and pump-probe ARPES data and **[3]** work products that PyARPES can generate.

2.2 Software architecture

PyARPES is structured as a loosely coupled collection of Python modules comprising common ARPES analysis techniques, providing primitives for building data analysis routines based on volumetric data selection and iteration, and creating data visualizations (see Figure 2.1). These decoupled modules are linked by a common data model for spectroscopic data, built on top of the xarray library (*72*) and providing an object-oriented interface both to raw photoemission data

and associated metadata attached by the data acquisition software or experimenter. In addition to independent modules for analysis and data preparation, PyARPES contains an extensible data ingestion package, a spectral function simulation package, a package for designing and exporting DAQ sequences to guide ARPES experiments, and a large number of interactive applications that can be run inside the Jupyter analysis environment, as interactive Matplotlib visualizations, or natively on a desktop as Qt applications.

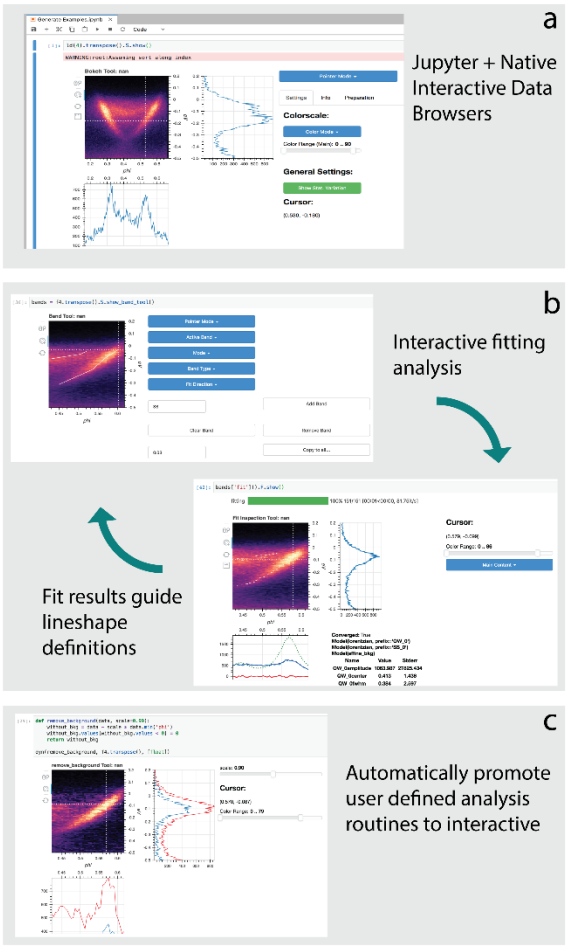


Figure 2.2 **Screenshots of selected interactive analysis and data exploration tools. [a]** A Bokeh-based data browser for 2D, 3D, and 4D data available in Jupyter Notebook via PyARPES. Projected marginals at the cursor location are shown. PyARPES also provides a high-performance native browser built on PyQt5. **[b]** A pair of interactive curve-fitting tools allowing the experimenter to lay approximate lineshapes for fitting, and to interactively view the results of multidimensional curve-fitting. **[c]** PyARPES also features the ability to turn any user defined analysis function into an interactive tool, with standard widgets for viewing data like color saturation, and providing UI elements for all analysis parameters. Here, a background subtraction method is tested on a Bi_2Se_3 spectrum.

2.3 Data model

The data format used for storage of experimental data after collection in labs or at beamlines is a source of unexpectedly serious challenge for scientists who do ARPES. Engineering challenges like how to read binary file formats derivative of those developed in astrophysics over decades

and how to read proprietary formats stemming from WaveMetrics—who makes Igor Pro—and Scienta—who makes most of the analyzers used at soft x-ray energies—only just scratch the surface. These different formats all use different geometry conventions as well, often at the discretion or whim of the first scientist to create the data acquisition software for a beamline but sometimes due to genuine differences in the geometry of the experiment and hardware used for ARPES.

This means that scientists traveling to new beamlines for experiments confront novelty not just in experimental hardware differences as they their data but also in loading and momentum converting their data when they arrive home again. Attempts to standardize data formats for different experimental sciences, like NeXus, will cut this Tower of Babel problem off at its foundation, but rely on long term agreement on basic conventions about photoemission experiments. This consensus will take time, perhaps a decade or longer, as instruments and software are phased out. In the interim, it's clear that current heterogeneity in data formats creates an urgent need for analysis tools which are geometry agnostic and can handle reading data formats from many different facilities.

PyARPES satisfies these constraints by adopting a universal data format for spectral data after it is loaded. This universal format is needed anyway since data analysis happens primarily with in-memory representations rather than on-disk ones which have different design constraints than those considered by current standardization efforts. Given the diverse geometries and container formats for data, PyARPES accomplishes a universal runtime data format—the PyARPES spectral data model—by performing geometry normalization as a step in the data-loading process. The data model also makes data exploration simpler by removing novelty and surprise from the data container formats used for ARPES. This further makes analysis performed on one dataset portable to another with minimal modification, creating a virtuous cycle where it becomes worthwhile to invest analysis and software development time to creating reusable, high quality analysis code which can be shared with the photoemission community.

The PyARPES data model permits analysis routines, primarily the angle-to-momentum conversion routines, to operate across data sources in a reliable way. With included modules, PyARPES provides a uniform interface to ARPES data collected at synchrotrons around the world—notably the Advanced Light Source, SOLEIL, Elettra Sincrotrone Trieste, National Synchrotron Radiation Laboratory, and Shanghai Synchrotron Radiation Facility—data stored in NeXus format, data stored in Igor Pro packed experiment and propriety binary wave formats, and data produced in proprietary formats by commercial electron spectrometers. These diverse data formats are supported through a plugin model which makes it easy for end users of PyARPES to add additional data sources for their experiments in a straightforward documented way. Optionally, these plugins then become part of the PyARPES data compatibility code so that other users can also access data in these same formats. Via the plugin system, limited support is even available for other experimental data which is frequently required in order to support the analysis of ARPES data. Most notably, PyARPES supports loading X-ray backscattering data by direct ingestion of the proprietary binary format from the NorthStar Laue backscattering data acquisition software. Several plugins, including those for the Shanghai Synchrotron Radiation

Facility and National Synchrotron Radiation Facility, have been contributed by end users without the aid of the author.

Although the geometry is the most challenging part of the PyARPES data model to understand and design, another critical function of the data model lies in exposing metadata in a programmatically accessible way. Data analysis and the general interpretation of ARPES data requires making sense of the high-dimensional photoelectron spectrum in the light of diverse metadata. The most important pieces of this metadata include information about the experimental configuration, temperature, source flux, and analyzer configuration and mode, but extend to minutiae like the timing of a piece of data in the larger acquisition sequence and the vacuum chamber pressure. PyARPES provides a predictable and uniform interface to all of this experimental metadata, meaning that analysis requiring certain information about the experimental conditions can be expected either to work reliably, or to inform the user with a meaningful description of what data is missing.

So far, the aspects of the data model we have discussed are common to all photoemission experiments. Narrower disciplines which use ARPES to resolve electronics while more carefully controlling or manipulating sample conditions or physics require specific metadata. TARPES and pump-probe studies require information about the pump and probe wavelength, fluence, pulse energies, temporal duration, polarization, and bandwidth. Electrostatically doped studies in nano-ARPES require the contact potentials and nano-ARPES itself has a richer coordinate set due to extra degrees of actuation on focusing optics and hierarchical sample positioning which needs to be resolved to a unified coordinate scheme.

2.3.1 Data provenenace

Layered atop concerns of data access are issues of reproducibility and consistency of scientific analysis. At their core, reproducibility amounts to creating a record of the analysis steps performed and data products created through the process of analysis which can be grouped together under the heading of data provenance. PyARPES tries to be practical concerning data provenance by retaining complete records of analysis sessions using IPython logging and attaching records directly to data when built-in analysis routines, including data loading and momentum conversion, are performed. Full logs are useful for reproducibility, while data-specific records are useful for an experimenter to quickly check what pieces of analysis code have touched a piece of data or to determine how a particular analysis in the past was performed. The most typical task is that a scientist wants to remember how they created a particular figure, so PyARPES also provides an opt-in figure provenance system where the data used to produce each figure is retained in a Python accessible binary format with each figure, as well as a JSON record of the script originating the figure, and the code that was run to produce it. Above improving figure introspection, this also makes it possible to replot data even when it would be too costly or complicated to re-run an entire analysis.

Ultimately, data provenance systems in a dynamic language like Python require the consent and cooperation of the end user: there are no foolproof ways to ensure that structured records are retained in a world where both data and analysis code are mutable. A broader challenge concerns

the way in which scientists formulate the data reduction process. In a perfect world, many scientists would use purpose-built data reduction graphical user interfaces and would avoid writing too much code on their own. This seems good in theory, but it complicates provenance because records of analysis steps are typically textual rather than graphical, even leaving aside engineering issues of supporting scientific workflows with growing complexity in a fluent graphical user interface. These issues are foundational for how PyARPES conceptualizes interactive analysis's role. While interactive tools are exceptional for initial data exploration and for determining basic facts like the coordinate offsets for high symmetry points in the recorded data volume, reproducible text-based analysis should be used all substantive analysis work. Reducing the coupling between interactive analysis and code-based analysis has been a large aim for PyARPES, with the intention that the resulting analysis code better reflects a reusable and reproducible single narrative terminating a publication quality figure.

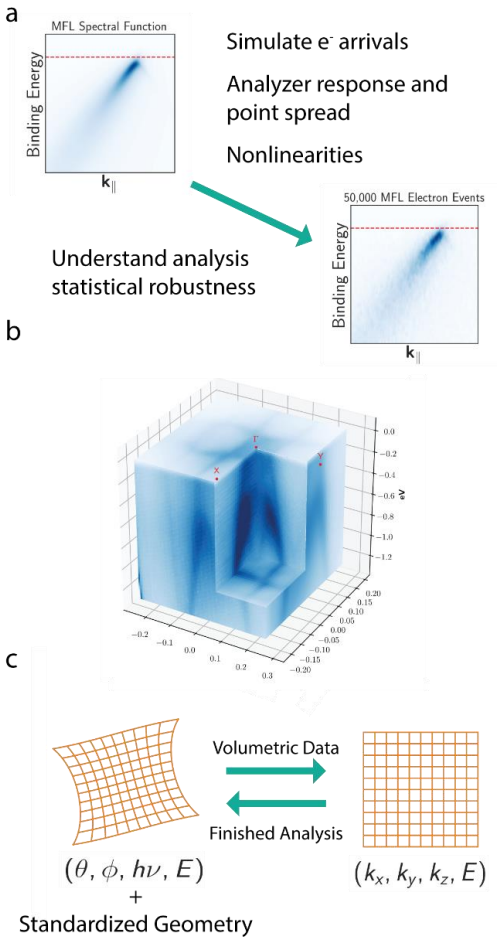


Figure 2.3 **Selected examples of features from PyARPES. [a]** PyARPES provides support for modeling ARPES spectra from theoretical or phenomenological models of the electron self energy, including by simulating electron events and detector point spread and non-linearity. **[b]** A variety of common and sophisticated plotting schemes for ARPES data, including cut surfaces through volumetric data (shown), are supported. **[c]** PyARPES provides an automated coordinate transformation scheme, based on the geometry defined by plugins in the data ingestion system. A uniform API is provided regardless of detector type or geometry. Coordinate meshes indicate the inverse transform coordinate warping over a roughly $\pi/3$ range for vertical slit detector geometry.

2.4 Angle-to-momentum conversion

Because PyARPES performs geometry normalization when loading data according to the PyARPES data model, PyARPES can automate the process of choosing and inverting the coordinate transform functions which are used to convert from the acquisition coordinates, expressed in angles, to physically meaningful momentum coordinates (Figure 2.3c). To leverage this automation, the user only needs to provide coordinate offsets corresponding to normal emission in their dataset, which vary from sample to sample and cannot be inferred automatically in a general way, after which PyARPES can convert the dataset and any others collected under the same geometry, to the photoelectron momentum. Tedious procedures like choosing appropriate momentum bounds and granularities which minimize dataset size and coordinate transform compute time without sacrificing resolution can also be conducted automatically based on the coordinate resolution in the parent dataset and approximate derivatives in the momentum coordinates with respect to acquisition coordinates. Furthermore, because ARPES datasets now routinely include very large datasets with many dimensions which do not participate in the momentum transformation process, like temperature dependent datasets or scanning microscopy datasets, PyARPES can automatically handle batching momentum transformation calls for datasets which are too large to fit into memory and can automatically pass-through these extra dimensions so that the momentum-space data retains them. Despite Python's reputation in some circles as a slow language, it has excelled in data intensive applications because expensive computations can be performed in a low-level system like LAPACK or BLAS while providing high-level operation chaining. Using numba (*75*), PyARPES achieves momentum conversion throughput in excess of several GB/s on typical analysis hardware like laptops—significantly faster than existing approaches in MATLAB or Igor Pro, according to the author's experiments. Extremely fast momentum conversion makes writing interactive visualization tools which rely on live updating a momentum space view feasible even for large, volumetric datasets.

Volumetric angle-to-momentum conversion routines implement the generic picture of gridded coordinate transforms which we discussed in the introductory chapter. To recapitulate, the forward conversion routines $k_i(\theta_j)$ are used to determine the bounds in momentum spanned by an angular dataset, a grid in momentum space is created, and the grid is filled by projecting each point $\boldsymbol{k}$ to the value in the measured coordinates using the inverse transformation $\theta_j(k_i)$. Although writing down $k_i(\theta_j)$ is relatively straightforward for simple experimental geometries, deriving inverse coordinate transforms and cataloguing appropriate routines supporting the geometries currently in use in labs and at popular synchrotron beamlines is not a small task. Because these are nevertheless problems an ARPES scientist may face when setting up a new experimental apparatus, I've devoted an appendix to covering this topic and providing Mathematica code which derives the transforms for standard goniometer scanning and deflector scanning geometries.

While bulk volumetric coordinate transforms provide everything needed for robust analysis of ARPES data, bulk conversion is computationally expensive and is inconvenient when the relevant momentum direction for analysis cuts diagonally through the data volume. These situations are common enough that PyARPES provides specific routines which perform a

coordinate conversion along a particular path in momentum space or which cut through a point in momentum space at a requested angle in momentum. This makes it extremely straightforward to look at various high symmetry momentum directions passing through a zone corner or to look at cuts through a superlattice mini-Brillouin zone which is rotated with respect to the parent material's Brillouin zone. It is also ideal for studying materials like graphene where the dispersion is similar along momentum cuts radiating from a high symmetry point and where comparative analysis of the dispersions in these different directions reveal details about the sample physics.

2.5 Data Corrections and Data Preparation

Quantitative interpretation of ARPES data requires assessment of the experimental apparatus on the data collected. PyARPES contains implementations of semi-automated corrections and data preparation tasks that make these quantitative analyses simpler and less prone to error. These include symmetrizing data across an axis, merging multiple datasets with common coordinates, applying corrections to the chemical potential across the spectrometer window or as a result of miscalibration of beam energy, normalizing data by the recorded angle-integrated photocurrent, background estimation, removal of the trapezoidal edge of an image recorded on a hemispherical analyzer and aberrations, and converting time-of-flight data from flight time to electron kinetic energy.

2.6 Interactive analysis

Interactive analysis clearly has a tenuous role in science because it makes it more difficult to keep detailed and reproducible records of the analysis process. Nevertheless, data exploration is the single most important aspect of analysis. On the one hand, exploratory analysis facilitates decision making during experiments, and therefore serves a critical role in enabling efficient use of hardware and samples. On the other, it allows the focused direction of efforts after data collection and therefore remains an invaluable feature of more rigorous analysis. PyARPES seeks a balanced approach to this conflict by providing rich interactive data browsers using an in Jupyter UI toolkit based on Bokeh and using native application UI support via Qt5 and PyQtGraph, but by restricting the analysis functions of these interactive exploration tools. A few example of the user interfaces, and a high level description of their function, can be seen in the screenshots of the analysis tools in Figure 2.2. The interaction between interactive tools and data analysis proper is limited to finding parameters and constants which can be easily exported for use in a text-based analysis session or to allowing an experimenter to validate aspects of their analysis, such as curve fits, by visualizing them more simply.

Although necessarily a compromise, experience has shown that this is a relatively fluent path towards engineering reliability and repeatability into analysis workflows which use interactive user interface during the exploration stage. As a positive follow-on effect, because the end result is a text-based analysis, the analysis record is executable and can be verified even on a system without installed UI dependencies, which are typically more complicated, expensive to install, and occasionally platform locked.

As with plugins, PyARPES encourages end users to build whatever interactive analysis tools they require to wrangle their data. To help motivate users to do this, PyARPES includes a high level UI programming abstraction for spectral data in Qt, in Bokeh, and in Matplotlib together with a detailed guide in the API documentation which illustrates by example building a new analysis tool from scratch.

2.7 Common Analyses

PyARPES implements the second derivative, image curvature (*76*), and minimum gradient (*77*) methods for band visualization, in addition to wrappers for common filters and data smoothing tools such as the Savitzky-Golay filter. Modules are also included for the analysis of band dispersion created by curve-fitting. To analyze Fermi surface structures and low energy phenomena, a module provides facilities to unwrap and select datasets around paths that follow the Fermi momentum across the Fermi surface, in addition to the robust, high level momentum conversion APIs we have already described. Meanwhile, robust support for spin-ARPES data together with common spin-ARPES plots is provided.

Many common but complex data visualization approaches are supported in a way which is aware of high symmetry locations in the data. For example, three dimensional volumetric plots with a segment removed to visualize the dispersion along high symmetry directions while also presenting the Fermi surface can be constructed in angle or in momentum coordinates, see Figure 2.2b. Because no single set of visualization approaches will be sufficient for the entire photoemission community owing to different stylistic approaches to the presentation of data and the broad set of subdisciplines concerned with different portions and properties of the electron spectral function, PyARPES takes a compositional approach to the creation of figures by providing modular utilities for layout, plotting, and annotation.

Finally, because curve-fitting of lineshapes is an essential component both of understanding the chemical information available in nano-XPS and also in understanding gap and valence band dynamics for pump-probe ARPES, interactive and tools are provided as part of PyARPES both for defining lineshapes that can be used in curve-fitting, and for inspecting the results of multidimensional curve-fitting analyses (Figure 2.2). Once lineshapes are defined, PyARPES can perform curve fitting conveniently and automatically over large and high dimensional datasets, allowing for efficient analysis of large amounts of photoemission data. After curve fitting, PyARPES supports many tools for understanding curve fits—e.g. tools to find the worst fitting curves among a large collection of automatically performed fits—and to transport good parameters to subsequent fitting runs. Curve fitting also automatically supports multiprocessing on a worker pool which improves the performance when hundreds or thousands of fitting runs are undertaken, as for dispersion fitting along a band in an image plot.

2.8 Data exploration and data analysis driven by machine learning

We’ve seen how interactive analysis tools permit rapid exploration of high dimensional data volumes using a data browser. Being able to efficiently pan through data permits scientists to quickly begin to understand their data; if the data varies sufficiently along the experimental degrees of freedom—for instance if the photoelectron spectrum changes dramatically across a

phase transition in temperature, or if spatially resolved measurements reveal different phases separated in space—then manual exploration rapidly motivates next steps for data analysis. What happens if changes are very subtle or are layered atop a varying background signal? In situations like these, critical insight might be delayed until more sophisticated analysis can remove confounding effects and demonstrate a signal. This is too late for the information to be useful for decision making during an experiment, however.

For years, machine learning has been considered for the ways in which it can replace human expertise in various tasks. Despite preconceptions of outsiders, science is far too soft and encompassing a domain for this replacement process to mirror what has occurred in well defined environments like perfect information adversarial games—chess and shogi—or in automating repetitive tasks in translation, annotation, and object recognition. Instead, machine learning provides the opportunity to augment the expert insight of scientists by digesting large data streams into smaller ones which retain the core features and sources of variance. This amplifies the utility of interactive analysis in the data exploration process. An original motivation for creating PyARPES lay in making ARPES spectral data available in a format where industry standard ML analysis tools could be used to investigate it in exploratory data analysis, at experiment time, and through the long process of digesting ARPES data into ground truth details of a sample's electronics.

Examples of how PyARPES satisfies these objectives are best seen by example. Later, we devote a whole chapter to considering this issue and the broader one of applying machine learning to ARPES. For now, we consider only how PyARPES situates itself in the broader machine learning and Python data analysis ecosystem. For exploratory data exploration, PyARPES integrates interactive data browsers with machine learning driven decompositions of high-dimensional data using the facilities provided by scikit-learn and scikit-image (*69*). More sophisticated tools from machine learning which attempt to bridge the divide between theory and experiment, or which attempt to predict material properties given some prior understanding of a material class, for instance by directly measuring a spectral gap in a cuprate from the entire raw ARPES spectrum and the experimental geometry, require large volumes of data to establish the statistical relationship between the measured spectrum, theory, or material properties. The simulation tools and detector models we have already considered are exceptionally useful to create synthetic of semi-synthetic datasets with an appropriate degree of realism. PyARPES makes digesting these datasets and predicting material properties possible by providing integrations—a DataLoader and a Dataset—for PyTorch, a popular deep learning library in Python.

2.9 Comprehensive usage documentation, example data, and tutorials

Complex software is only as good as the documentation which facilitates its use. This truism is as valid in scientific software and data-oriented software, even if documentation rarely materializes outside of very large, established, and well-funded efforts because of the comparatively smaller groups of people involved in making and using scientific software. Two further reasons made providing high quality usage and tutorial documentation a requirement in PyARPES. First, PyARPES marks a departure from established tools like Igor Pro, which makes

providing community support essential. Second, we claimed at the outset that one design ambition for the software was to facilitate a future where more non-expert users might begin being able to perform and analyze data coming from photoemission experiments. How will these users know what to do with their data without being able to match their needs to common analyses and data workflows? The PyARPES Documentation Site meets this need by providing full executable analysis tutorials as Jupyter notebooks covering the basics of Jupyter usage, data loading, data preparation, momentum conversion and visualization, and multidimensional curve fitting. It further provides "case studies" where simple but realistic datasets covering nanoXPS and temperature dependent ARPES measurements are treated in detail from exploration to drawing physical understanding. These high-level tutorials operate on data included with the PyARPES installation which makes them editable and executable on every end user's analysis computer. Complementing high level tutorials, PyARPES includes internals documentation and comprehensive API documentation.

2.10 Theoretical modeling and detector response functions

ARPES experiments frequently need to appeal to phenomenological models to distinguish between genuine physical effects, and byproducts of the measurement process when the evidence is marginal, as when trying to measure the size of a gap in the presence of finite experimental resolution, or when trying to measure the self-energy within a narrow energy band around the Fermi level. These models are very different from the physical models used by theorists—with which they are often used in conjunction—because they are concerned with the accurate representation of the measurement process rather than physical phenomena.

PyARPES provides a growing library to simulate realistic photoemission spectra which can include detector response, point spread, and finite sampling of the photoelectron spectrum. These effects can either be applied to a known spectral function, or applied to simulations of the spectral function from the self-energy and bare band dispersions (demonstrated for a marginal Fermi liquid in Figure 2.3a).

Statistics support also exists in PyARPES and allows for bootstrapping analysis procedures that consume volumetric data. For applications which use electron counting detectors, these tools allow for the attachment of statistically meaningful uncertainties on quantities that in ARPES analyses have generally only been conservatively estimated. These tools are especially helpful in the context of spin-ARPES and time-resolved or pump-probe ARPES, which operate with significantly fewer photoelectrons per integrated spectrum than equilibrium "band mapping" applications.

In the case of Tr-ARPES, It was quickly realized that extracting meaningful information from experimental data requires a higher degree of care to avoid experimental pitfalls including deconvolution of the detector response and incident ultraviolet linewidth, detector response calibrations, corrections to the chemical potential, space charge (*78*), and experimental nonlinearities (*18*). As the community moves towards novel detector and light source schemes, including pump probe experiments at free electron lasers, these challenges will only intensify further (*79*, *80*). As ideas for the study of Tr-ARPES data emerged, it also became apparent that

tools for studying gap dynamics and size in superconductors and charge density wave materials produced conflicting results. This contraindicates their use without comparison to other methods and, ideally, a study of their performance on data drawn from theoretical models (*81*). The phenomenological modeling tools and statistical bootstrapping tools provided in PyARPES are ideal for exactly this use case, where realistic simulated ARPES spectra must be compared to experiments under challenging configurations.

2.11 Impact and Future Directions

Seamless and immediate access to data enables the experimenter both to develop varied analysis approaches—detailed above—and to make informed decisions during an experiment—detailed in the next chapter. With the release of PyARPES we aimed to improve the state of angle-resolved photoemission on both accounts. Because of the consistent data interface and plugins, it allows an experimenter to apply preferred preparation and analysis routines while conducting experiments at synchrotron light sources, allowing interactive assessment of data quality, statistics, and experimental geometry before valuable time is lost. End users at synchrotrons and university labs have already extended the compatibility of ARPES with plugins for data sources beyond those original provided by using this plugin interface.

Later at analysis time, in leveraging simple APIs to common ARPES corrections and analysis routines, it offers to the experimenter a friendly work environment that encourages best practices in the analysis of a variety of photoemission datasets. By building on and providing tight integration with scientific computing tools in Python, PyARPES enables rapid development and trial of new analysis techniques that would otherwise be prohibitively expensive to implement in environments with more meager scientific computing support. This structure also allows for its use in cooperation with existing theory and simulation codes without shuttling data between separate environments and formats.

While PyARPES was developed with the ARPES community in mind, its structure, utilities, and plugins for data ingestion make PyARPES appropriate for any chemical or physical spectroscopy, providing fertile ground for the development of analysis tools in synchrotron-based science and optical spectroscopies.

Through PyARPES, scientists practicing ARPES are provided simple access to their data, tools to explore it, and sophisticated analysis and modeling modules. However, PyARPES is by no means a finished product. Releases following the initial public version documented in the PyARPES paper (*15*) have expanded the scope of the analysis capabilities of the software and refine the experience of data analysis conducted inside PyARPES. In future releases, we hope to improve coverage for next generation photoemission apparatus like multidimensional time-of-flight detectors and to provide broader support for data formats at other synchrotrons. Progress on both goals will help fulfill the premise of universal access to data across instruments while the community moves to adopt common data formats.

3 AutodiDAQt

It would be easy to point out that analyzing the data is only one part of the story of experimental physics, we also need to collect relevant data in the first place. This gives the wrong impression. The acquisition and analysis of data are not separate endeavors; instead, most experimental disciplines operate via two cyclic feedback systems. The first, at the level of a discipline or high-level research direction describes how knowledge is created, spread, summarized, and used as the basis for future research. When we think and speak informally about scientific progress, we are most likely thinking about this loop. The operation of this loop and in particular the creation of new scientific knowledge requires the cooperation of a second cyclic feedback loop, which describes how a small group of scientists refine their ideas incorporating experimental data or modeling results. In this loop, the result of incorporating new data is that the space of unknowns where additional data is warranted shifts, and more experiments need to be done.

Experimental disciplines which can rapidly triage and understand experimental data are in the best position to be disciplined about supporting concrete scientific claims with experimental data and an isomorphic theoretical model. In the limit, these disciplines can provide an adaptive approach to collecting data with each experiment conditioned on what was learned in prior experiments. Their data tells a cohesive story about a scientific claim. In the opposite limit, the feedback cycle is slowed to the point that experiments can be conducted once. In this limit, their data is the narrative subject of a story told to explain a limited observation. Because ARPES produces large volumes of high dimensional data over very short experimental windows of just 8 to 48 hours, ARPES has fallen into the slow limit.

The challenges of inefficient experimentation are not limited only to efficiency of knowledge generation; they also apply to its range. One way of organizing the landscape of modern experimental physics is through the set of experimental tools used by its practitioners. Physics adapts to new unknowns and challenges by developing novel instrumentation and hardware capabilities capable of answering questions outside the operational limits of a prior generation of experiments. Well deservedly, this perspective places the burden on hardware and experiment for advancing what we confidently can know. In recent decades, however, the share of the burden hardware is saddled with has shifted: software and statistical advances now make up a larger portion of the increased range of modern experimental physics. We do not need to look very far afield from photoemission to see this happened. In astrophysical imaging, computational techniques like lucky imaging, laser adaptive optics, and machine learning based blind deconvolution and analysis are permitting physicists to obtain higher quality numerical and imaging insights with ground-based telescopes of the same apertures as were previously available. Cryo-electron microscopy only works because reconstruction algorithms exist which can build a least surprisal estimate of the 3D structure given many independent observations of 2D projections. Various approaches to super-resolution and nonlinear microscopy blend software and hardware approaches to defeat traditional optical resolution limits, requiring sophisticated approaches to the collection of data and to the way biologists analyze this data. Ptychographic approaches to scanning transmission x-ray microscopy and high-angle annular dark-field

scanning transmission electron microscopy are revolutionizing nanoscale imaging, and require real-time coordination between sophisticated analysis of experimental data and the hardware which acquires it.

Perhaps spurred by the advent of high-performance computing, the ease with which numerical computing can be performed using high level routines in Python, MATLAB, and Julia, or for a subtle confluence of reasons, great advances have been made in the development of data analysis software suites and toolkits for individual physical disciplines. PyARPES, our contribution to the wrangling and analysis of high dimensional photoemission datasets, has peers in *chinook*, in tools which integrate machine learning based denoising of spectral data, and in the development of experimental data repositories on materials like FAIRmat. Each of these signal ambition in the field towards making high dimensional spectroscopy data more digestible after it has been collected and reducing the burden of moving from experiment completion to publication of a scientific paper and a progression in collective materials knowedge.

At the same time, the development of hardware for ARPES continues to make great strides. Whereas in previous decades it was the spectrometers which made serious gains in both resolution and throughput with the commercialization of hemispherical electron analyzers and the refinement of low kinetic energy lens tables for soft x-ray and ultraviolet photoemission, purpose designed ARPES experiments now push boundaries along all imaginable dimensions of the experiment. Sub-micron beams are now available for photoemission: the experiments which use them are rightfully called nano-ARPES. Two-angle resolving time-of-flight electron analyzers have further increased throughput by permitting the simultaneous collection of the entire surface band structure. Together, nanoARPES and two-angle momentum-microscopy are beginning a revolution in how we study two dimensional materials, heterostructures, and electronic devices.

Achieving continued progress demands more than more of the same. We need a fusion of the hardware we use for data collection and the software we use to understand it. Ptychography, cryo-EM, astrophysical advances, and the other techniques already mentioned bear out this principle. Where they differ from other experimental approaches which have not adopted this hardware/software fusion is in the intensiveness and difficulty of the experiment. Tight coupling of software has occurred where benefits are highest and where experiments are hardest.

Limited penetration of tighter coupling indicates that the friction in adopting this style of experiment design remains very high, because benefits exist for all types of experimental physics. While fundamental qualitative improvements may not be possible as they have been in the examples given, increased experimental throughput and improvements in hardware allocation to the collection of high value data stand to gain even in very mature experimental domains. ARPES straddles a line here: it is intensive enough to expect genuinely new kinds of experiments can be conducted by better control and tighter coupling of hardware to software, but it also stands to gain dramatically by parsimonious use of very limited resources, especially for TARPES user facilities and emerging nanoARPES installations.

Further advantages arise because data collection is a highly specific software domain. Data collection is tasked with responsibilities which either do not exist or are less important in other application and software contexts. These responsibilities include the measurement or removal of systematic sources of bias in experiment, event logging and fault recovery, the creation and maintenance of data provenance, which is foundational to reproducible science, and interfacing and managing the lifecycle of complex and expensive experimental hardware. The most complex aspect of data acquisition software lies in furnishing user interfaces where scientists can make decisions about what data to collect and introspect data as soon as it is collected by the experimental apparatus.

In calling this a software domain, we mean that these requirements are universal and applied in a relatively undifferentiated way across all data acquisition tasks from optical spectroscopy to neutron scattering. In a less academic context, we would talk about these effects in terms of market size. There are comparatively fewer common needs for all software, despite essentially everyone requiring some software. In a very narrow discipline like TARPES, there are many requirements of software which are common to all TARPES setups, but only a few practitioners. Scientific data acquisition is a narrow enough discipline that it has complex common requirements but covers a large enough audience that common software systems which fulfill these common requirements in a simple and trustworthy way is valuable without considering the broader impacts of hardware/software fusion.

In the current scientific environment, scientific DAQ software is purpose built to suit the needs of a single or at best of a few experiments. Consequently, software tools which could fulfill the basic functions of data acquisition software are continually rebuilt at great expense or are not implemented at all. Because the most immediate need in any lab is for data and eyes to look at it, underbudgeted labs especially in smaller universities may not implement necessary data archiving and provenance tools at all which degrades the quality and integrity of scientific findings which demand reliable data. In the best case, DAQ software can consume substantial fractions of instrumentation costs. Every PI and program lead in this position prefers to see their grant money and time spent elsewhere.

3.1 Scientist interaction with data acquisition

Data acquisition does not solely consist of moving hardware and writing numbers down from instruments. Scientists must actively introspect their data during acquisition; this makes data acquisition closer to interactive design work than to writing computational code. Experimenters cannot perfectly forecast the region in the experimental configuration space—which may be a location on a sample surface or a temperature pertinent to a phase transition—that will yield the desired data. They can only react and refine what they are measuring based on what has transpired during experimentation. This ties into the earlier goals of acquisition and analysis fusion already promised. DAQ is tightly integrated to both user interfaces—this is a necessary part of keeping a human being in the loop about decision making—and to domain specific analysis tools. This gives a second inviolable design constraint for the scientific data acquisition domain. Data acquisition needs to be tied closely to hardware and to modular scientific analysis suited to determining what data a scientist needs to collect next. The close ties are required to

increase efficiency in the data collection analysis feedback loop and to shorten the knowledge generation cycle, while modularity is required to account for differing styles of experimentalists and requirements placed by individual experimental physics projects.

3.2 AutodiDAQt and Existing Software Solutions

This is a lot of boxes to check for an ideal data acquisition software system. Are there existing tools which satisfying them and are appropriately geared towards physicists? I asked this question when I sat down to write the prototype data acquisition software for a research and development electron spectrometer and PEEM being developed by my colleagues Ping Ai and Jacob Gobbo. Unfortunately, there was no ideal fit, despite many previous efforts to marry the requirements of scientific data and its acquisition to user interfaces, code generation, or any number of other application programming details. Nonetheless, two systems stood out for clearly moving towards the platonic ideal I had traced out. PyMeasure, a data acquisition framework for physics, provided user interface generation primitives, but did not attempt to handle automation for data handling and provenance. Furthermore, PyMeasure took only limited steps towards composing acquisition routines and refining acquisition sequences without requiring the experimenter to write lots of (PyMeasure specific) code. In another camp, Bluesky completely eschewed user interface generation, but captured all of the essentials of composition acquisition routines while offering robust metadata support. Unfortunately, unlike PyMeasure, Bluesky requires significant software engineering effort to handle configuration and provide user interfaces before it can be easily deployed in an academic lab setting. It was designed and introduced for use in synchrotron-based experiments with dozens or a hundred software addressable degrees of freedom at places like Argonne National Lab. At that scale of software, the involvement of some engineering time on the data acquisition side is inevitable.

In the absence of a sufficiently simple alternative which would provide both user interface generation and modularity of acquisition sequences, I opted to finish development on research software system—AutodiDAQt—which combined the approaches of Bluesky and Pymeasure that I had developed a few years before. The remainder of this chapter presents AutodiDAQt, which follows as closely as possible in the model suggested by prior discussion. Notably, I will show how AutodiDAQt provides composable data acquisition primitives, that AutodiDAQt automates the generation of experimental UIs which are isomorphic to the controlled hardware and the acquisition routines which use that hardware, that it provides the ability to control what is being measured with user analysis code during measurements, and that it embodies the simplicity required for use in flexible, lab-scale experiments.

At the end of the chapter, we explore how AutodiDAQt permits control of acquisition remotely by analysis software, provide full code examples for complete data acquisition systems, and discuss prospects for how systems like Bluesky and AutodiDAQt may change the data acquisition paradigm in ARPES.

3.3 User Interface Generation

AutodiDAQt uses several tools specific to the data acquisition programming niche which permit it to generate user interfaces for experimenters without code. Because user interface code is

exceptionally error-prone, complex, stateful, and platform dependent even when solutions like National Instruments' LabVIEW are invoked, UI generation relieves most of the burden of typical application programming for data acquisition. DAQ tasks, which consist of experiment parameters, collected data and their streams, and the state of apparatus have a finite set of enumerable and predictable components. By registering hardware components into AutodiDAQt with data schemas which indicate whether a piece of hardware reads values like a CCD or temperature diode or writes values like a motor or temperature controller, what kind of data they accept, and how to perform data validation, you not only permit communication and control of that hardware, but automatic representation in the application UI.

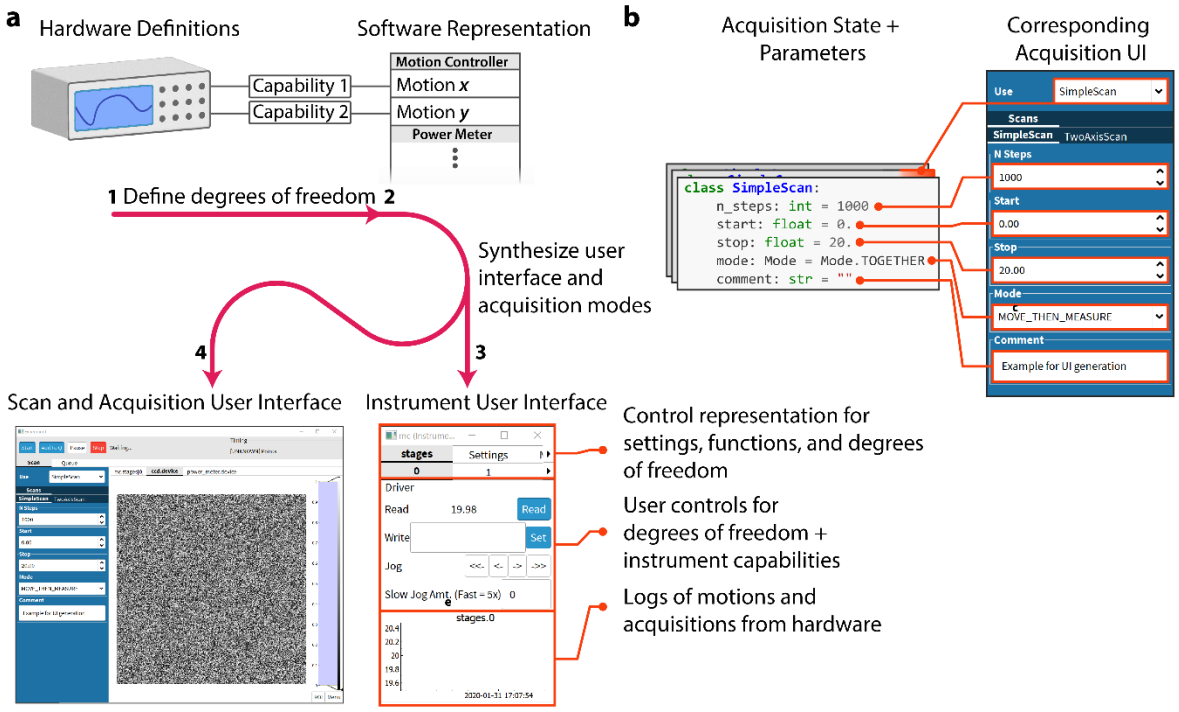


Figure 3.1 ***User Interface Generation.*** **[a, 1-4]** Selected portions of the default UI provided by AutodiDAQt and its isomorphism onto data acquisition hardware and instruments. **[a1]** Software drivers express degrees of freedom (axes) and capabilities (settings, functions) for instruments. **[a2]** Hardware capabilities are mapped to a uniform representation for experimental degrees of freedom and experimental state in AutodiDAQt. In limited contexts, as in driver libraries where uniform structure can be assumed or limited possibilities enumerated, this mapping can be built automatically by introspection. Otherwise, an expressive API for defining the hardware semantics of an experiment is available. Semantics describe how to invoke instrument methods to control hardware, enforce limits on values, determine their concrete types, and assign their role in an experiment. **[a3]** The data acquisition program state—progress, collected data, current configuration—are linked to automatically generated user interface elements for each degree of freedom in the experiment. **[a4]** AutodiDAQt generates controls and user interface for whole acquisition programs—their parameters and the interim collected data—from their software representation [a2]. The left panel shows the main acquisition window connected to a test CCD generating uniform noise with program status, data streams, queued future acquisitions, and user-defined parameters for data collection. **[b]** As more control may be required in some circumstances, users can use the UI generation primitives for composite data structures (classes) to encapsulate program state which must be represented on screen. If acquisitions are parameterized over this state—e.g. by the acquisition range, the number of points, or preferences of acquisition control flow like whether to wait after motion or whether to perform motion simultaneously with hardware reads as in depicted in [b]—AutodiDAQt will generate appropriate controls.

Because it would be unnecessarily difficult and inflexible to insist that all user interface definition had to happen automatically, AutodiDAQt also lets expert users who would prefer to control the interface for a particular instrument or for the entire acquisition suite to do so while plugging into and reusing as much of the automatically generated application UI as they would like.

The most straightforward way to see how this works, and to see the simplicity it provides for data acquisition programming is to look at how AutodiDAQt generates real user interface in practice (Figure 3.1a) and how fields in the user interface connect back and are synchronized with data in the application (Figure 3.1b). As we have discussed, hardware capabilities defined by instrument drivers (Figure 3.1a1) are the glue underlying both the software representation of an instrument and the resulting UI. The experimenter only needs to define the representation of these capabilities (Figure 3.1a2) in AutodiDAQt, which can sometimes be accomplished automatically, for AutodiDAQt to be able to communicate with and control hardware during an experiment. AutodiDAQt generates monitoring and control interfaces for hardware with a "virtual front panel" and streaming plots (Figure 3.1a3) together with an acquisition session manager (Figure 3.1a4) from these representations of the experimental hardware. This combination removes the obligation for UI programming for all but unusual experimental requirements.

In the next section, I will show that composability also follows from the definition of degrees of freedom of the experiment in software. This permits AutodiDAQt to describe common acquisition programs over the configuration space of an experiment defined by its hardware automatically, and for the internal state of acquisition programs to be connected and updated with user interface elements (Figure 3.1b). Previously, creating front panels for instruments which were coherently linked to the acquisition system and creating control interfaces for the acquisition program itself were demands made of scientists. AutodiDAQt subsumes each of these responsibilities by cutting away everything but the irreducible complexity of specifying your data acquisition hardware, its capabilities, and its semantics in your experiment (Figure 3.1a1 and Figure 3.1a2).

3.4 Composability of Data Acquisition Programs with an Algebraic (Sum and Product) Structure

Programs and systems compose when they have well defined borders which permit different orientations of assembly and reassembly. Often, these structures can be mathematically described, as in the type systems of strongly typed languages with algebraic data types (ADTs). ADTs compose data through a kind of sum—a piece of data can be an *A* or a *B* or a *C*—and a kind of product—the product of a thing which is a *D* and a thing which is an *E* is both a *D* and an *E* together. Naturally, as the term algebra would imply, sums and products compose. The distributive property $(A + B) \times C = A \times C + B \times C$ reads as: a piece of data which is either an *A* or a *B* and which is also a *C* can be viewed as either an *A* and a *C* or as a *B* and a *C*. This seems rather abstract and disconnected from the topic of composability of data acquisition programs at

hand, but AutodiDAQt embraces a form of composability of data acquisition programs which has exactly sum and product algebraic structure.

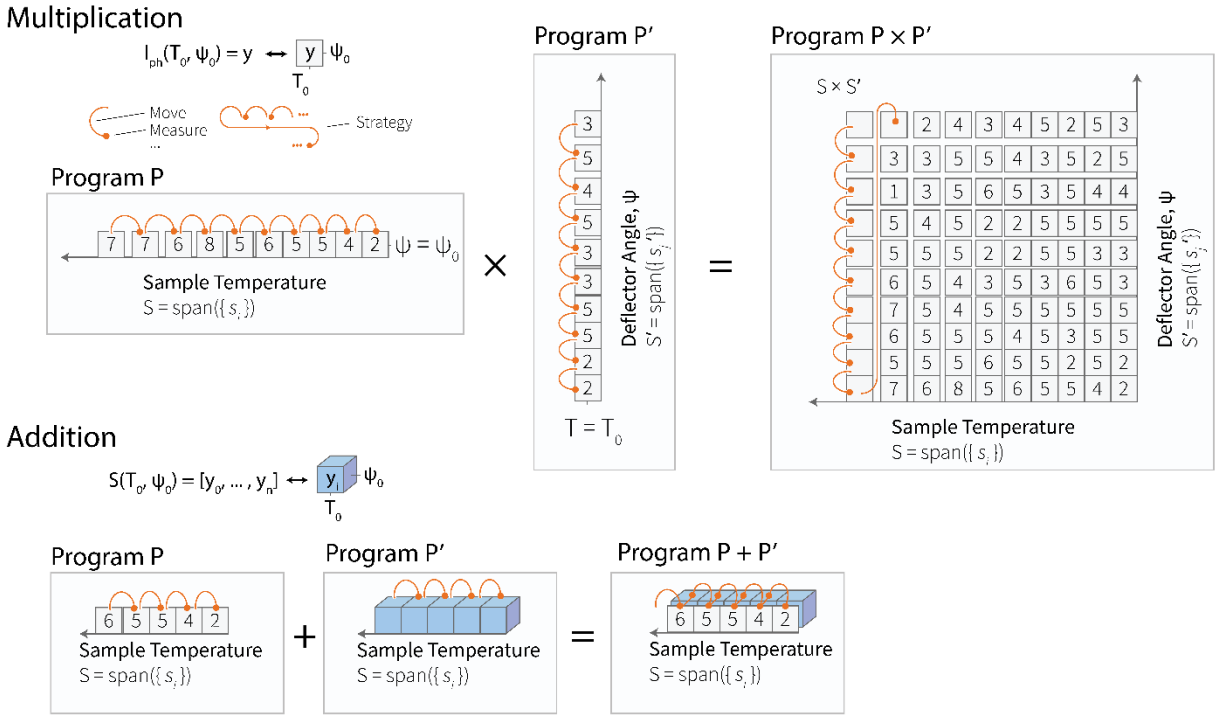


Figure 3.2 **Products (top) and sums (bottom) of configuration spaces and programs.** AutodiDAQt can use products and sum structures to composite acquisition programs as is shown by example in the diagrams above. **[Top]** Two one-dimensional acquisition runs R and R' acquire their data by running acquisition P over spaces S and S'. Their product consists of recording with the acquisition P while iterating over the Cartesian product $S \times S'$. There is some ambiguity in terms of the order of traversal over the product space. One option, yielding a noncommutative product, is to run R' in the inner loop of R by convention, which is what is shown in the first row. Other strategies exist depending on what is desirable in the context. As a convenience in AutodiDAQt, this can be configured using different kinds of product structures over S and S' including random traversal and a zigzag traversal which minimizes the travel time between adjacent coordinates. [**Bottom]** The result of taking the sum of runs is shown in the bottom row. The second run differs than the first in that it records $S(T, \psi)$, the entire ARPES spectrum, instead of just the photocurrent $I_{ph}(T, \psi)$. $R + R'$ records the union of the data by iteratively changing the experimental configuration to new temperatures T_0 and new angles ψ_0 before recording $I_{ph}(T_0, \psi_0)$—required by R—and then recording the data $S(T_0, \psi_0)$—required by R'. After collecting at (T_0, ψ_0), $R + R'$ continues to the next required configuration (T_1, ψ_0).

The minimal view of a data acquisition run R is that it traverses some configuration space $S = \{s_j\}$ and at each configuration in the configuration space s_i it runs a program P. Frequently, the configuration space is itself an outer product of degrees of freedom of the experiment, so that $S = \text{span}(\{f_i\})$ is the lattice span over degrees of freedom f_i. To be concrete about this idea, consider an ARPES experiment R which consists of measuring a single ARPES cut at different temperatures and another R' which consists of measuring ARPES cuts at fixed temperatures but at varying chamber rotations θ. In this case, the lattice we collect points over during run R is just a discrete set of temperatures, so $\{f_i\} = [f_{\text{Temperature}}]$ while for R' it is $\{f_i'\} = [f_\theta]$. The product

$R \times R'$ naturally consists of taking ARPES cuts along $S \times S' = \mathrm{span}\left([f_{\mathrm{Temperature}}]\right) \times \mathrm{span}([f_\theta]) = \mathrm{span}\left([f_{\mathrm{Temperature}}, f_\theta]\right)$, i.e. at every combination of temperature requested by R and θ requested by R'. This is what we mean by the product of acquisition programs.

What about sums? There are actually a few different sum structures we can put on acquisition programs (the simplest, and not very useful: do everything run R says to do and then do everything R' says to do). A very useful formulation of the sum structure can be considered when $S = S'$ but when R and R' run different internal acquisition programs at each visited point. If our acquisition program P merely collects a set of data $D = \{d_i\}$ by reading from attached instruments, then $R + R'$ can read from each of the instruments $\{d_i\} \cup \{d_i{}'\}$. Imagine two acquisition runs R and R' which record from the ARPES spectrometer and from a grounded femto-ammeter on the sample, respectively. Then $R + R'$ would consist of recording both pieces of data, the ARPES spectrum and the photocurrent, at each visited point. In surprisingly many cases, the combination of these sums and products permit expressing the desired data acquisition routine in terms of irreducible, atomic operations which AutodiDAQt knows how to perform. In such cases, no user programming of acquisition is required at all and the specification of the experimental R can be made declaratively. Diagrams which trace out exactly the situations of sums and products of simple acquisition programs just discussed are shown in Figure 3.2.

Internally, AutodiDAQt permits this kind of composability via synthesis of the runtime steps required to follow an acquisition sequence. Because AutodiDAQt separates high-level instructions required to specify the acquisition from the asynchronous runtime required to orchestrate hardware, it can string together high-level operations transparently without requiring the scientist to concern themselves with the details about how low-level operations, like controlling a stepper motor or waiting for temperature readings to settle, get implemented. In most cases, the scientist only cares about the high-level operations anyway, especially when records of the low-level operations are also retained automatically so that questions about correctness can be assessed in the event of any irregularities in the data. In addition to providing simplicity, this declarative approach—espoused also in Bluesky's take on data acquisition—has distinct advantages which range from making writing experiments more expressive, improving intent when recycling and reusing acquisition code, to improving the durability of the acquisition code and making stronger guarantees about the correctness of results.

Although the picture we have painted the modularity approach in AutodiDAQt sounds simple, there are many details which must be handled. Largely, these map onto the generic application programming concerns we already mentioned. AutodiDAQt records full acquisition sequences with detailed logs of high- and low-level operations. This behavior means that acquisition sequences can be replayed at a later date, either to collect new data or to view the behavior of the program in operating conditions for the purpose of debugging and either in reality or in simulation with virtual instruments. These aspects would be challenging to accomplish and would likely place much higher engineering burden on the scientist without the declarative separation of concerns. As alluded to above, other advantages of loose coupling arise at test and debugging time: because a piece of hardware can be transparently swapped with a simulacrum

test instrument, even automatically from the schema of the data it records, the entire acquisition software can be tested on computers without any real hardware, with slightly varied hardware, or with the full complement of instruments which would be used during the course of an actual experimental run.

A final advantage of the declarative approach stems from making it possible for an external analysis routine to provide the sequence of acquisition steps, fulfilling one of the core objectives we articulated for a data acquisition system at the beginning of this chapter. In the next section, we will examine the concrete issue of how the distributed systems architecture of AutodiDAQt permits attaching remote analysis clients which can be used to direct data acquisition. We will also examine some prospects for efficiency and correctness improvements this can bring to the acquisition of ARPES data.

3.5 Analysis-in-the-Loop: Realization in AutodiDAQt and Applications in ARPES

Providing application programming primitives which remove the difficulty of implementing correct and reliable data acquisition programs was only one of two urgent challenges in scientific DAQ software which we motivated at the beginning of this chapter. I've now demonstrated how AutodiDAQt was designed with precisely these concerns in mind and provides a straightforward approach to addressing these complexities. The second challenge we signaled concerned how experimenters interact with their data acquisition system when they go to collect data. Is the system sufficiently flexible to permit the experimenter to record only what data they need and to conserve valuable time on hardware at user facilities? Is the data provided immediately available and digestible so that the experimenter, or their analysis tools, can evolve their plan for subsequent data collection based on a dialog with the data? Can the data acquisition system express all the reasonable modes of acquisition supported by the hardware, even if they weren't considered by the original system designer?

Modularity moves towards answering these questions affirmatively but stops short on the issue of creating a dialog between the data and the experimenter. Existing systems accommodate dialog by integrating exploratory data analysis directly into the data acquisition suite. For instance, high resolution ARPES setups used to study cuprates and other superconductors usually integrate a display where a constant momentum spectrum containing the Fermi level is shown, together with the result of a curve fit from a simple edge-like model so that the experimenter can assess the presence or absence of a gap and the strength of the quasiparticle peak. In AutodiDAQt, I've taken the position that this approach is fundamentally flawed because it requires data analysis to be integrated in the system used to write the DAQ system—LabView, systems languages, and asynchronous runtimes—rather than with the tools for analysis an experimenter fluent with—MATLAB, Python + numpy, Julia, Igor Pro. Moving the application boundary further into the data acquisition software should also be avoided because it is unsafe and can cause errors: analysis code is written to a lower standard of rigor than code which controls hardware, by placing them in the same system errors in analysis and visualization code can propagate to the data acquisition system writ large.

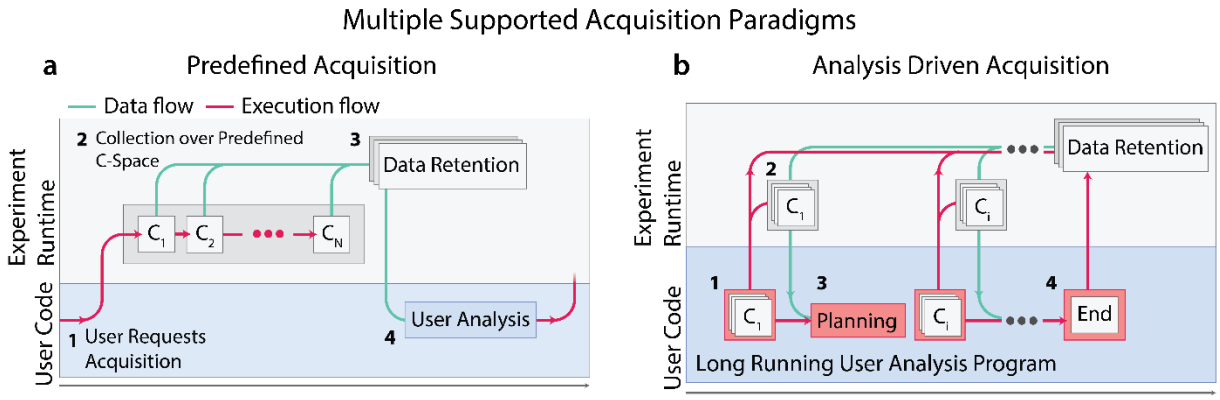


Figure 3.3 **Analysis Driven Acquisition. [a-b]** Diagrams showing the support data acquisition paradigms in AutodiDAQt. [a] In most data acquisition a scientist selects from a limited menu of static data acquisition programs. When new acquisition programs are required which do not fit, they have to be programmed specifically or modified from an existing option. Once requested (label 1), the system collects data over the states C_i (label 2) until all the anticipated data has been read from hardware. At this point, the run is finished, flushed to disk or other archival storage, and the data is provided for offline user analysis or inspection (label 3). After making decisions based on offline analysis with the newly recorded data (label 4) the experimenter can request another acquisition and the process is repeated. [b] An alternative acquisition paradigm which consists of iterative decision making with an embedded analysis program. A scientist submits a program to the acquisition system (label 1) which describes what data is required and how recorded data should influence further collection. The program can issue high-level acquisition instructions which are validated for safety and are scheduled on the acquisition runtime (label 2). Any intermediate data collected is provided back to the program synchronously. In response to the acquired data or as part of a pre-existing acquisition strategy, the scientist's program can issue additional instructions (label 3). This iterative process of collecting and performing a small, routinized analysis continues until the data acquisition hardware or the experimenter's analysis program decides that the acquisition is finished (label 4).

The only way to address this issue is to refuse the inclusion of programming interfaces for the analysis of data in the runtime itself. AutodiDAQt does this by providing a client library, AutodiDAQt Receiver, which runs concurrently with AutodiDAQt across a message broker as a distributed system. The receiver, which can run either on the same acquisition computer or on the computer of an experimenter on the same local network, receives and collates data during an acquisition so that the experimenter can perform whatever visualizations or exploratory analysis they choose. Data is published across the broker as soon as it is available, so that the partial results of acquisitions lasting longer than a few seconds are always updated and available on the experimenter's analysis computer. The receiver is more than just a replicated follower for the data logging, however. The receiver can also emit high level instructions to the experiment runtime indicating which instruments should record data, or how to reconfigure the experimental hardware to prepare for the next data point to be collected.

Broadly, AutodiDAQt supports two very different data acquisition paradigms. In the traditional paradigms, the experimenter selects from a predefined acquisition program, which may be parameterized by the range of values to measure over during the run. This scheme for data acquisition is shown in terms of its implementation on the AutodiDAQt runtime in Figure 3.3a. Once they have data, they must inspect or analyze it before making further decisions about what to do next.

The alternative is that fine- or coarsely-grained decisions about the next data point—even during a single collection run—can be made by a user analysis program which is running asynchronously on AutodiDAQt receiver, shown in Figure 3.3b. Being able to perform acquisition driven by analysis provides a leap in both experimental capability and efficiency.

3.5.1 Applications of Analysis-in-the-Loop to Nano-ARPES

I am an experimental physicist who uses electron spectroscopy. It's inevitable that some of the design constraints and decisions made are influenced by that background and that AutodiDAQt may be slightly overfitted to that regime (e.g. in the internal assumptions of slow but concurrent access to instruments). ARPES is nonetheless a demanding experimental discipline because of the complex hardware and large data volumes. These properties make it ideal for us to examine the impacts of analysis-in-the-loop (AITL) as implemented in AutodiDAQt by considering some theoretical modifications to how ARPES data is acquired.

NanoARPES and nanoXPS are emerging subdisciplines of photoemission spectroscopy which are rapidly gaining mindshare and beamline space at synchrotrons because they provide direct access to microelectronic devices, inhomogeneous materials, and two dimensional heterostructures. Although ARPES goniometers have always included spatial degrees of freedom for sample alignment, beam sizes were too large and positioning stages too imperfect to permit studies of spatially varying electronic properties on sample surfaces outside truly exceptional circumstances. For this reason, we should think of nanoARPES as adding additional degrees of freedom to those resolved in the ARPES measurement which must potentially be scanned over and tuned during acquisition. This makes the configuration space in nanoARPES and nanoXPS measurements vastly larger, and places significant emphasis on efficient use of the time budget for an experiment, which can be set by sample degradation or by the realities of scheduling allocation at oversubscribed facilities.

First, we consider how very rudimental applications of AITL can improve efficiency over typical control procedures. In Figure 3.4a, we plot the tungsten $4f_{7/2}/4f_{5/2}$ core level doublet XPS spectrum collected from a muti-layer piece of WS_2. This spectrum was collected by coarsely rastering—Figure 3.4b—over the surface of an exfoliated WS_2 flake on graphite using a data acquisition software suite programmed with LabVIEW at the MAESTRO beamline of the Advanced Light Source (BL 7.0.0.2). What happens if the experimenter concludes that the resolution they picked for the scan in Figure 3.4b is insufficient to determine a spatial offset where full electronic structure measurements should be taken? Increasing the spatial resolution in each direction by a factor of 3 results in a total acquisition duration which is around 9 times longer than that used for Figure 3.4b, if the dominant contribution to the acquisition duration is in acquiring electrons on the spectrometer and not in moving the motors to position the sample under the beam. A simple solution is to only collect data on points which are close to sampled locations in Figure 3.4b where the $4f$ level is intense. A simple approach for defining the cutoff is to use the ratio of the intensity on one peak to the XPS background between the peaks as a masking criterion, as is shown. Using this approach, a 9x higher areal resolution can be acquired over the entire sample surface using only 3.3x the acquisition time of the first scan.

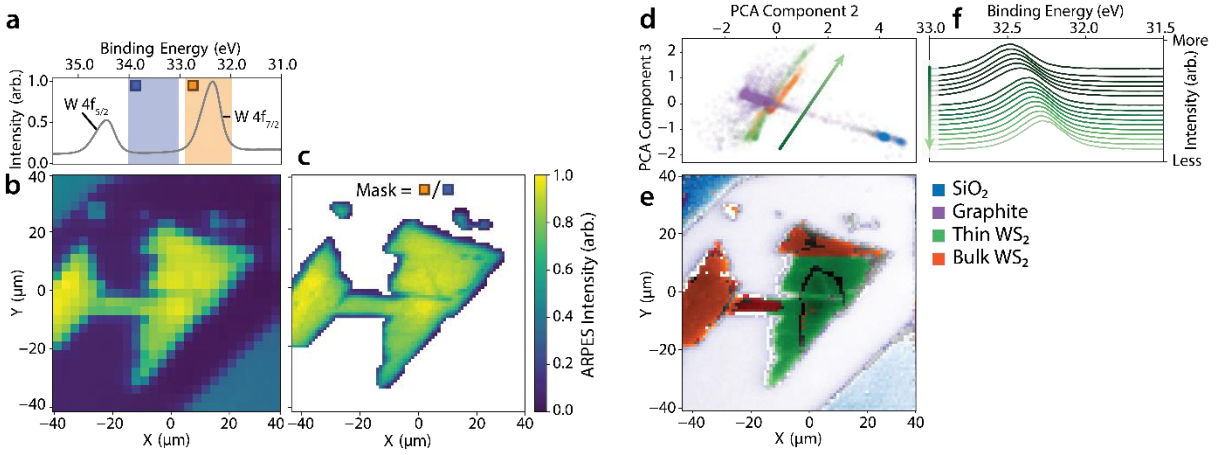


Figure 3.4 **nanoARPES Opportunities for Analysis Driven Acquisition. [a-c]** A concrete example of how a nano-ARPES data acquisition performed with traditional approaches can operate at much lower efficiency than with analysis driven acquisition. [a] X-ray photoelectron spectrum of the Tungsten 4f spin-orbit split core levels. The orange and blue shaded regions are integrated to use in ratio as a mask condition in [c]. [b] A low resolution nano-XPS image as might be used for sample finding and sample topography, with complex region-of-interest consisting of a WS_2 flake on graphite. [c] High resolution experiment trajectory taken exclusively on sample regions with intense W 4f peak (blue divided by orange energy region). Because the core level ratio rejects regions which consist of bare substrate graphite, the collected area occupies 37% of total scan window. **[d-f]** Integrating machine learning into the analysis-acquisition loop not only improves efficiently but permits rapid understanding. [d] PCA component projection for XPS curves across the WS_2 sample. The XPS curves are colored according to their composition (labels at right of [f]). [e] A coloring of the sample surface showing the assignment of different XPS curve clusters to morphological features on the sample surface. Equal color regions in [e] correspond to scattered points in [d]. [f] By observing that the green colored cohort of points in [d] is correlated with a strong factor of variation from the PCA, we can investigate its meaning by correlating the cluster to the equivalent location on the sample. In this case, variation corresponds to inhomogeneous doping across the sample surface which manifests as a peak shift seen by nanoXPS. Even simple applications of ML- and analysis-in-the-loop provide actionable understanding which can improve experiment outcomes.

Other more sophisticated schemes, falling under the broader umbrella of machine-learning-in-the-loop have been explored. One approach class is based on Gaussian Process regression (coarsely: approaches which maximize surprisal and sampled variance in the data so that regions of sample which change slowly are sampled sparsely, corresponding to white regions in c, which are not sampled at all). GP regression has already demonstrated progress towards autonomous experimentation in photoemission spectroscopy (*82*). It is unlikely however that autonomous experimentation can cover even a majority of the experimentation needs in ARPES spectroscopy in the short term. Human-in-the-loop methods which improve the decision-making power of scientists are a more capable middle ground which promise to improve throughput while targeting the specific needs of scientists. Figure 3.4d-f illustrates an approach covering the same problem space as discussed in Figure 3.4a-c. Figure 3.4d shows a principal component analysis (PCA) decomposition of the XPS curves from the same sample region. Colored scatter cohorts are identified by visual clustering and correspond to sample morphology and contents in Figure 3.4e. Although not principled, rudimentary decompositions like PCA rapidly provide insight which drives efficient use of acquisition time. Decomposition features are also frequently highly correlated with relevant sample physics: in Figure 3.4c it is apparent that PCA has identified a

proxy for the inhomogeneous doping variations on the WS_2 sample. In this scheme, machine learning is used for rapid surveying and summarization of datasets which makes the best use of available scientific expertise for decision making.

Crude machine learning approaches serve this vision ideally. In the above, PCA serves as a surrogate for a spatially resolved curve fitting analysis which would be costly and unstable to perform at experiment time. Surrogate analysis sufficiently simple to be robustly applied at experiment time and which concentrates data acquisition on the regions targeted by the desired analysis provide all the experiment time benefits. For instance, to collect data on highly gapped regions of a spatially inhomogeneous superconductor, using the leading edge of the photoemission signal as a surrogate for the gap is sufficient because the leading-edge location is a monotonic function of the gap size: with larger gap, the leading edge also recedes.

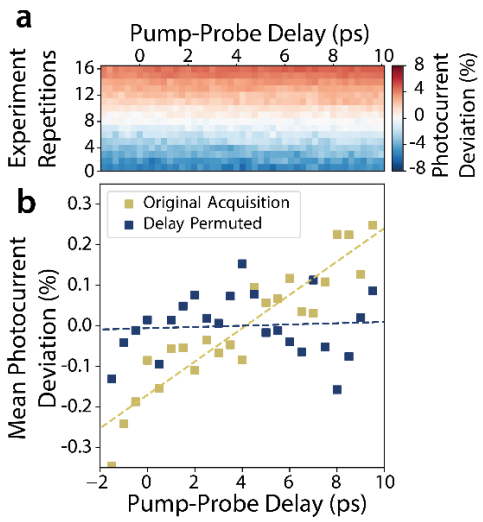


Figure 3.5 **Using analysis-in-the-loop and scan modularity to improve experimental reliability. [a, b]** Some sources of systematic bias in very sensitive measurements, like time- and angle-resolved photoemission spectroscopy measurements, can be reduced or removed by issuing custom scan instructions via analysis or via the modular composition of simpler acquisitions. [a] Tr-ARPES measurements using the fourth harmonic (or higher) of a fundamental laser have much larger power transients because the UV power scales with the fourth power (or higher) of the input. UV transients can bias the results of experiments which collect data "in order"—even when using many short iterations of the experiment intended to average out fluctuations—and invalidate the conclusions of studies which aim to address subtle effects. [b] By using AutodiDAQt receiver, or the internal modularity tools to perform a randomly ordered product scan, to perform acquisition in random order, the systematic bias can be removed. The yellow curve and trend show the power fluctuations occurring with a 4^{th} harmonic BBO + BBO system generating a probe UV pulse from the 835 nm output of a Coherent Mira fs Ti:Sapph oscillator. Transients exceeding 0.5% occur even when conducting the average of many pump pulse sweeps [a]. By simulating reacquiring the data in random order, real laser power fluctuations are no longer correlated with other degrees of freedom in the experiment.

Alternatively, the analysis-in-the-loop approach gives scientists the ability to rapidly adapt to changes in experimental conditions and to remove dataset bias, by treating software as a malleable tool rather than a fixed constraint. We now turn to issues of systematic bias arising in pump-probe ARPES, where changes in laser intensity throughout an experiment can cause confounding effects where infrared dose delivered can be correlated with the measured delay

time. A typical approach to minimizing this effect, if stabilizing the source power is not feasible, is to repeat the experiment in many short repetitions so that transients are better spread across different delays. The resulting total dose delivered can be visualized in this scheme for an actual experiment in Figure 3.5a.

Despite the appearance that this minimizes the effects of transients, when we average data across repetitions, we see that there is still a pernicious dependence of mean dose, measured by total photoelectron yield, as a function of the experimental delay, as in Figure 3.5b. Properly removing the bias requires stratifying individual experimental runs by dose cohort, and in randomly shuffling the acquisition order so that there can be no correlation. Adaptively accommodating these kinds of responses to issues arising during acquisition requires a dynamic and cooperative approach between acquisition and analysis. In this narrow case, AutodiDAQt provides support for acquisition shuffling in either of the supported acquisition paradigms. In the broader context, deeper cooperation between DAQ aware user analysis programs, such as PyARPES(*15*) in the context of angle-resolved photoemission spectroscopy, permit scientists to rapidly define acquisitions over configuration trajectories which are otherwise challenging to define without the expert knowledge of the experimenter. In this approach, it is straightforward to collect data along a particular path in reciprocal space, in ARPES, or to perform coordinate correction at acquisition time arising from deflection of photoemitted electrons on samples which complex electrostatic environments like gates for electrostatic doping.

3.6 Building Real Acquisition Programs with AutodiDAQt, a Few Examples

Software systems are best demonstrated with actual code samples, so we will give an example here of a system—using virtual hardware—which collects data from a scalar detector over a two-dimensional space. To ground this in the photoemission world, we will pretend that the two-dimensional control space is the sample XY plane, and that we are recording the photocurrent as a function of the relative alignment of the sample and the beam. Code Listing 1 shows complete application code.

In the lines initializing *dx* and *dy* we configure ranges which can be scanned or swept using the 1st and 2nd axis on a virtual motion controller, respectively. The actual preconfigured scans are built by combining these sweep definitions in the *scan_methods* attribute of the custom *Experiment* class. There, we define two scan methods, “dx Scan” and “dx-dy Scan” which use *dx* and *dy* as control variables and which read from the virtual power meter. Finally, in the last section we start the experiment.

Running the code shown above results in a long running data acquisition application, shown in Figure 3.6. Panels for each of the virtual motion controller and the power meter are shown which can be used to manually control hardware and which provide per-instrument management and data records. The largest window is the data acquisition window. In the blue sidebar, we can see controls which have been generated for the scan dimensions which we named *X* and *Y*. The acquisition panel is shown after having run a 2D acquisition with the settings shown in the left panel. The staircase structure for *mc.stages[0]* which was named *X* for the purposes of “dx-dy Scan” results from running the *Y* coordinate as the inner loop in the acquisition.

```
from autodidaqt import AutodiDAQt, Experiment, scan
from autodidaqt.mock import MockMotionController, MockScalarDetector

dx = MockMotionController.scan("mc").stages[0](limits=[-10, 10])
dy = MockMotionController.scan("mc").stages[1](limits=[-30, 30])
read_power = {"power": "power_meter.device"}

class MyExperiment(Experiment):
    scan_methods = [
        scan(x=dx, name="dx Scan", read=read_power),
        scan(x=dx, y=dy, name="dx-dy Scan", read=read_power),
    ]

AutodiDAQt(
    __name__, {},
    {"experiment": MyExperiment},
    {"mc": MockMotionController, "power_meter": MockScalarDetector},
).start()
```

Code Listing 1 **A full example acquisition program** which actuates a motion controller while recording from a scalar detector, like a power meter (shown) or ammeter.

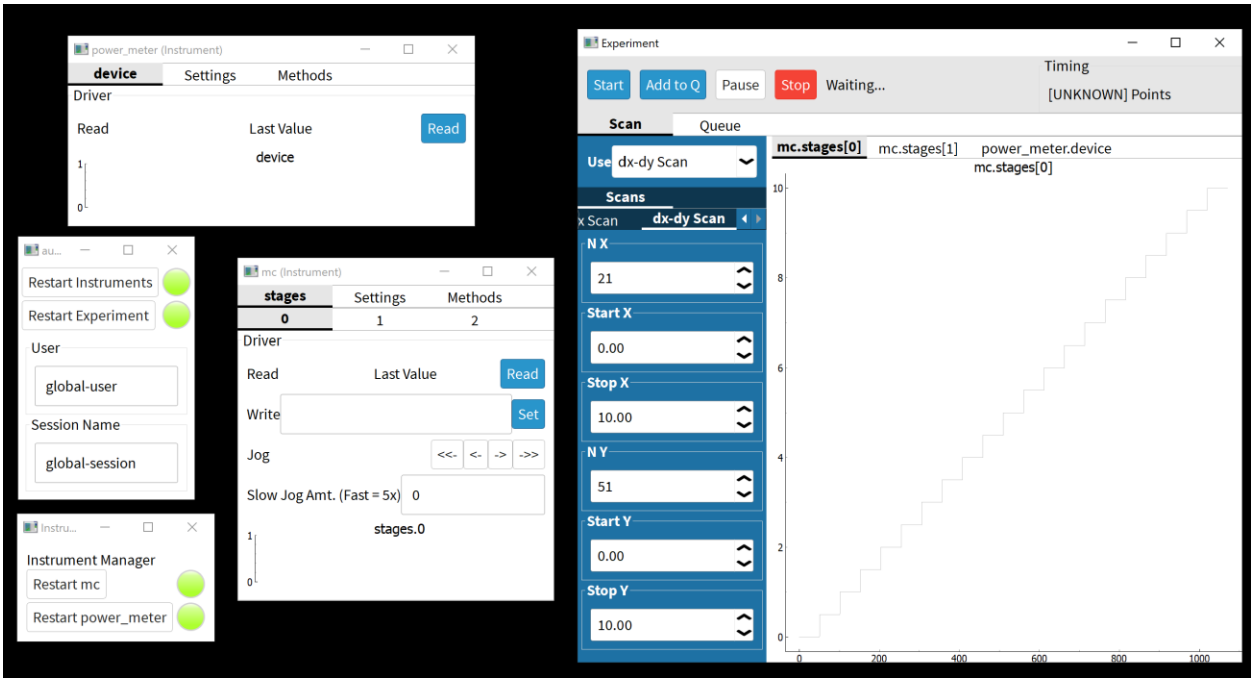


Figure 3.6 **The generated user interface and running data acquisition program for the code shown in** *Code Listing 1***.** On the right-hand side, we see the main acquisition window with configuration options for the currently selected scan mode and a selector for the scan which should be used (blue sidebar). The large plot shows the data recorded by each degree of freedom moved or read during the measurement as a function of a discrete time index labeling points visited (bottom of white subpanel). The gray bar at the top includes queue, stop, and measurement start controls as well as a progress indicator for the current experiment. On the left-hand side, we find control windows for the motion controller (*mc*) and the power meter (*power_meter*), with absolute positioning and jog controls. Additional application state windows on the far left allow modifying experiment metadata (like the experimenter) and reopening closed panels which correspond to long-lived background processes in the distributed runtime.

The default output data format in AutodiDAQt uses a zarr serialized xarray.Dataset. A simple representation of the data and the raw acquisition logs, as would be shown in Jupyter during a corresponding analysis session, are shown in Figure 3.7. The output data is a 2D grid with the expected *x* and *y* coordinates given the requested scan range, while the 2D *power* variable consists of randomly generated numbers over this grid. The raw acquisition logs, shown in Figure 3.7 bottom, are indexed by the wall clock time when each step, read, and write in the experiment was performed. Although all the data can be reconstructed from this record, the usefulness of this acquisition log is primarily in being able to retrospect the acquisition sequence for irregularities after an experiment has concluded. Partial data in the same xarray format is available over AutodiDAQt receiver during the course of a measurement so that it can be inspected in whatever analysis system the user desires during a long acquisition.

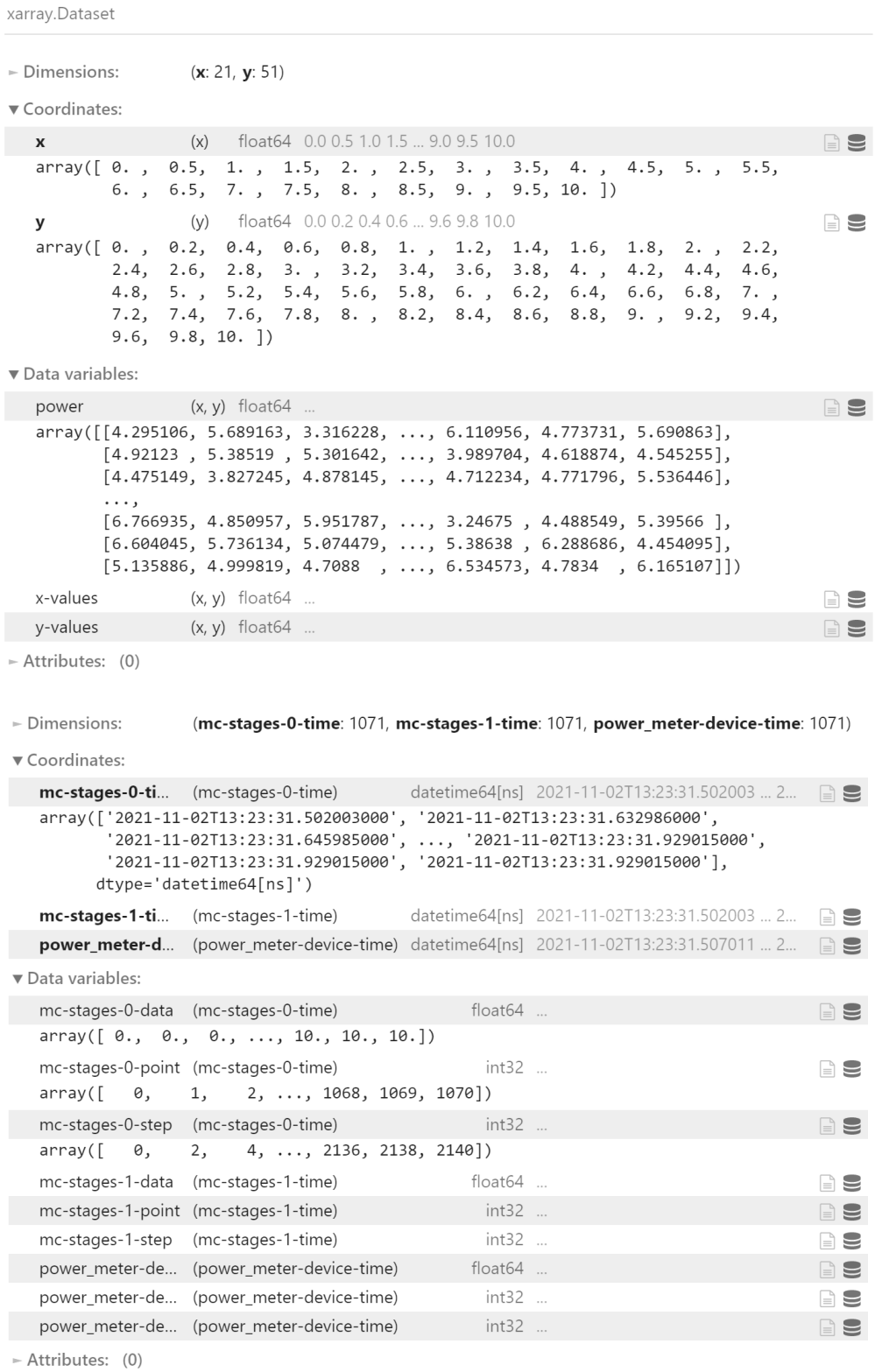


Figure 3.7 ***xarray*** **structured output data for the DAQ program shown in** *Code Listing 1* **when executed with the configuration and parameters which are shown in** *Figure 3.6* **[Top]** There are two degrees of freedom or scan variables which were changed during the "*dx-dy Scan*". These variables, x and y, each have an entry in the

output file together with the recorded data requested *power* which is a two-dimensional scalar array of the variables (x, y). Each of the variables x and y are indexed only by themselves, whereas the measured power is indexed by each of x and y. x and y can be thought of as the coordinates of the axis-aligned lattice of measurement locations. **[Bottom]** While the physically meaningful record of the data is the sampled scalar field $power(x, y)$, we may need to know the order points were collected in, the wall-clock time of each point's acquisition, and metadata emitted from the runtime like warnings or user defined comments which can be used to inspect irregularities. AutodiDAQt retains JSON and *xarray* structured command logs which include the acquisition order of each granular read or write instruction (e.g. *mc-stages-0-step*) the equivalent orders binned by output "point" (e.g. *mc-stages-0-point*) and the raw data which are indexed by the actual time when that command was processed (for the 0^{th} axis on the motion controller: *mc-stages-0-time*). Full application logs are also copied with data output (for brevity, not shown).

3.7 System Internals: Implementing a Data Acquisition Runtime with Remote Brokered Programming

Achieving reliable performance and fault tolerance in a nearly real-time distributed system is a serious challenge in the general case, but it is made easier for AutodiDAQt and small scale data acquisition systems because a single computer is sufficient for almost all cases, thereby ruling out unreliable networks as the most common cause of unpreventable faults. Despite their challenges, distributed systems are advantageous for long running data acquisition, especially over monolithic acquisition programs which run serially on a single thread, because distributed systems can isolate user code and other unreliable parts of the system so that critical components are unaffected. AutodiDAQt exploits this structure to improve overall system reliability, and to improve fault tolerance to errors caused by the acquisition sequencing code which must inevitably get written by scientists in a hurry.

Three principal design choices are responsible for increased resilience in applications running AutodiDAQt. First, AutodiDAQt uses a central monitor or watchdog process which is owns the actual user interface thread and which dispatches communication. Other code runs with some degree of isolation from the main thread—as an exception checked coroutine on the main thread, in a separate thread but on the same process as the monitor, or on an entirely separate process. Not all process isolation is for safety purposes: retention of data to disk happens on a separate process because it is both computationally expensive and expensive in terms of I/O to collate experimental data and save it to disk. Most isolation permits leniency on AutodiDAQt's part to recover from faults—by letting the monitor possibly restart them or gracefully shut down the rest of the system—or by reducing the chance of resource starvation on critical code segments. This last point is why hardware instruments run with their own coroutine on AutodiDAQt: it is never a user's responsibility to make sure that startup and cleanup code for a piece of hardware gets run.

The second main design decision is, unless required, to isolate the code written by an experimenter from the application code. User code should not be trusted, and it's harmonious that this complements the modularity strategy we've already discussed: user scan code does not control hardware directly but tells the runtime what to accomplish instead. This is also the strategy adopted in Bluesky, for similar reasons. A further advantage of this approach is that the separation between user and application code makes it transparent to test and develop data acquisition programs without having physical access to hardware, or when only a subset of hardware is present (as when all but one of your vendors have delivered hardware for your new

experiment, but a critical component is still missing). The separation between user and application code means that data flows into the runtime (instructions) and from the runtime (data) through the course of an experiment. This situation is shown in Figure 3.8c which depicts the DAQ command/data flow in a running acquisition. Further technical discussion along these lines is left for Figure 3.8c's caption.

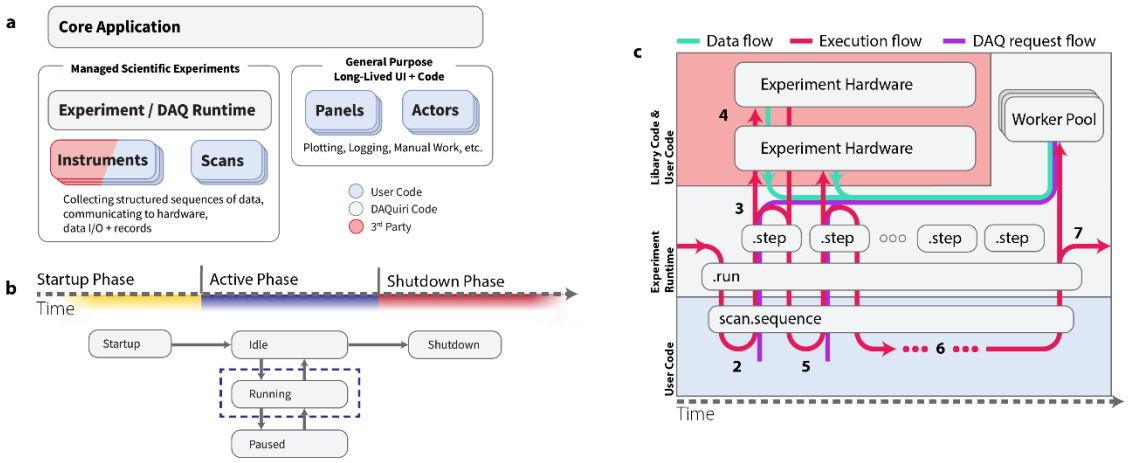


Figure 3.8 **System Component Diagram and Detailed Experiment Control Flow Diagram for AutodiDAQt [a]** The core application provided by a data acquisition system using AutodiDAQt is composed of parts related to the direct management of scientific experiments and associated hardware and auxiliary components required to maintain long running processes like data logging and retention and generated user interface. Under the framework of managed scientific instruments, the experiment runtime operates as a finite state machine [b] with transitions induced by interaction with the user interface, by dequeueing of an already queued acquisition request, or by remote command. Declarations of managed instruments can come either from user or library code. In addition to scan modes and hardware, the user can also plug in additional components to the distributed runtime by defining an Actor which can pass and receive messages with the runtime and which runs for the entire program lifecycle, or by creating a Panel which is a simple abstraction around a user interface window with a functional reactive programming model. Internally, instruments are modeled as an Actor (the instrument lifecycle, value stream, etc.) and a Panel (the user interface which provides controls to the experimenter, and which sends messages or calls methods on the Actor). **[b]** Simplified transition diagram for the experiment runtime. **[c]** The control flow between user and library code. The diagram depicts the flow of data (teal) and commands (purple) during acquisition operating without AutodiDAQt receiver. The flow of program execution (red) is passed between the runtime, hardware, and the planning sequence sequentially to perform reads and writes from hardware. Boundaries between user code, AutodiDAQt library code, and instrument driver code are shown in shaded regions. Individual execution "steps" corresponding to an atomic group of reads or writes are shown at the narrowest granularity and correspond to the **-step* subarrays in the example output data of *Figure 3.7 Bottom*. The numbered sequence can be followed to understand the chronology during an acquisition. (1) The experiment transitions into the running state. (2) The runtime asks the user scan code what it should do. User scan code provides a command for the next step in the acquisition and may issue further instructions ranging from output data format requests to a request to add comments into the output record and logs. (3) The command is executed which may involve (4) interaction with the hardware. (5) the process is continued until no further commands are issued (6). Finally, the data is finalized to disk and the hardware is available for another session (7).

The third decision is to adopt state machines to coordinate the experiment and remote command flow (simplified version showing correspondence to common notions of experiment status, Figure 3.8b).

So far, the design decisions we have discussed thus far concern AutodiDAQt running in local experiment planning mode—Figure 3.3a. All parts of the distributed runtime run on the same computer, although at times in different coroutines, threads, or processes. To attach another computer running AutodiDAQt receiver, AutodiDAQt waits for access requests over a network from a client which is running the receiver. Once this connection is made, a somewhat difficult hurdle needs to be overcome in establishing that the two ends of the connection are speaking the same language. This is a challenge because *a priori* the connecting remote does not know what kind or what structure of data is produced by the hardware, and in principle custom data types may have been added to the application code to support one or more of the static scan modes. We climb this Tower of Babel by first sending a message to the remote client which describes all of the different types which are defined, either by AutodiDAQt or as custom types by the user, by recursively encoding the schema for the instruments and experimental runtime, thereby establishing knowledge of all attached instruments, predefined scan modes, and data types on the remote client. Further communication which uses the custom types' values are serialized in JSON according to this schema.

At first glance, this is more complicated than is required to support remote execution. Python supports marshalling Python data via *pickle* and other serialization formats. Why can't we use those? This presents a problem where we place constraints on the remote analysis computer. It would not be a good experience for a user to arrive at a beamline after a long flight and learn that they needed to install a completely different version of Python. A larger reason for schema forwarding concerns the ability to attach more fully fledged and user-friendly user interfaces which use the same brokered connection as AutodiDAQt remote. Because plain, text-based data is passed across the connection exclusively, any remote client can easily speak the same protocol as AutodiDAQt receiver and display data from the experiment or act as a system to issue commands as desired.

3.8 Conclusions and Long-Term Prospects for AITL DAQ Systems

Compared to existing approaches to the data acquisition task, AutodiDAQt represents a compromise between data acquisition simplicity and holism that makes it the ideal platform for scientists who want software to get out of the way of their central task: to collect and understand data. By providing a robust approach for synchronous control and analysis via AutodiDAQt Receiver, this compromise is better seen as a strength which we believe will enable a new generation of experiments driven by real-time machine learning analysis, as well as traditional analysis, of acquired data during the experiment. While the constraints guiding AutodiDAQt's design are best understood through the lens of minimizing the engineering effort required by scientists its irreducible footprint—defining and implementing the hardware degrees of freedom—the positive externalities extend into the way scientists conduct their experiments. As scientists are allowed to make more parsimonious use of experimental resource and acquisition time, AutodiDAQt opens routes to novel experimental methodology using existing apparatus and heralds the opportunity of treating data collection as a living and malleable part of the analysis process.

4 Emergent Behavior and Localization in Electronic Superlattices

When Yuan Cao and coauthors announced the observation of correlated insulating states and separate superconducting domes on the hole and electron doped sides of a carefully engineered bilayer graphene system, the two-dimensional materials world was enthralled (*83*, *84*). Nature Springer's decision to publish the results in separate and simultaneous *Nature* papers timed to coincide with the talk at the APS March Meeting sealed the fate of the two-dimensional materials world for the next few years as much as did the structure of the bilayer graphene phase diagram. Sitting in the audience, the bilayer graphene phase diagram bore uncanny resemblance to the archetypal cuprate high T_c phase diagram. Maybe after decades of study an equally interesting but completely different material might yield answers to one of the oldest and most confounding problems in condensed matter physics?

Physicists had studied bilayer graphene for years, essentially since the moment graphene was first isolated by exfoliation in 2004 by Andre Geim and Konstantin Novoselov (*85*). How had a conspicuous feature like superconductivity gone unnoticed for nearly two decades? The answer involves practically infinite subtlety. Most of the work on bilayer graphene concerned Bernal stacked bilayers—the form you would naturally find two graphene sheets adopted if you peered inside a graphite crystal—where an uncomplicated band structure yields massive rather than massless excitations at the K and K' points but no exotic superconductivity (*86–88*). Physicists had also studied artificial bilayers, created by isolating separate monolayers before reassembling them or by growing them, in more limited studies (*89–91*). Assembling bilayers in this way permits creating graphene bilayers whose registry can be forced into configurations which cannot be obtained straightforwardly through isolating neighboring layers from pyrolytic carbon. Graphene's low energy physics arises from its p_z orbitals, which in the monolayer form massless Dirac cones but which permit interlayer hybridization in Bernal stacked bilayers. In twisted heterostructures, these bands participate in hybridization too: where the dispersions cross between cones, anticrossings corresponding to Bragg scattering between the cones via the interlayer hopping of these p_z orbitals create minigaps, minicones or replica cones, and renormalize the low energy physics and band velocity.

By strange coincidence, the way in which this hybridization occurs meant that if two graphene sheets were assembled within a particular angular range, then the lowest energy miniband resulting from the hybridization became nearly perfectly flat (*92*). Flat bands correspond to charge carriers which have negligible velocity leading to the interpretation that these electrons become trapped in the electronic corrugations left by the twisted partner layer. Slowing the electrons down permits the electron-electron Coulomb interaction to dominate the kinetic energy of the electrons and for correlated states to emerge. This demonstrates why correlated electronics were not seen in graphene bilayers previously and why they arose at the "magic angle": it is only in a discrete set of angular ranges that the miniband becomes sufficiently flat for this condition to arise. If the angle is larger, the Bragg gaps are far from the K point and the low energy band structure resembles an unrenormalized Dirac cone. At smaller angles, the local stacking order so

closely resembles either Bernal or in the non-Bernal AA stacking that the lattice just adopts these high symmetry orders, and large-scale correlations no longer occur.

Between this conceptual model for the flat bands, the rich phenomenology of insulating states and superconductivity, and graphene's reputation, it would be easy to assume that twisted bilayer graphene (t-2LG, "twisted two-layer graphene") is a coincidence and a wonder material for realizing and simulating correlations on a triangular lattice.

4.1 A general approach to flat band emergence in 2D materials

Twisted bilayer graphene is an outlier, not for being unique in the production of flat bands and moiré correlations, but because the mechanism producing the flat bands—hybridization between the Dirac cones of the two layers producing a perfectly flat lowest energy band—is unique. A more general approach for producing flat bands and correlations exists(*93*, *94*) which permits stability of correlations to higher temperature, and which permits the realization of flat band physics in broader material classes where the interaction of flat band electrons and native material properties can produce a rich electronic and optical phenomenology.

Most bands are not linearly dispersing. Physicists have only recently become interested in graphene and topological materials which support this kind of dispersion. More generally, the band edge of a material can be expanded quadratically at low order under the effective mass approximation

$$E_{VB}(\boldsymbol{k}_0+\boldsymbol{q}) \approx E_{VB}(\boldsymbol{k}_0) + \boldsymbol{q}\left(\nabla_{\boldsymbol{k}} E_{VB}(\boldsymbol{k})|_{\boldsymbol{k}=\boldsymbol{k}_0}\right) + \frac{1}{2!}\boldsymbol{q}^T\left(D^2_{\boldsymbol{k}} E_{VB}(\boldsymbol{k})\big|_{\boldsymbol{k}=\boldsymbol{k}_0}\right)\boldsymbol{q} \tag{4.1}$$

$$E_{VB}(\boldsymbol{k}_0+\boldsymbol{q}) \approx E_{VB}(\boldsymbol{k}_0) + \frac{1}{2!}\boldsymbol{q}^T\left(D^2_{\boldsymbol{k}} E_{VB}(\boldsymbol{k})\big|_{\boldsymbol{k}=\boldsymbol{k}_0}\right)\boldsymbol{q} \tag{4.2}$$

$$E_{VB}(\boldsymbol{k}_0+\boldsymbol{q}) - E_{VB}(\boldsymbol{k}_0) \approx \frac{\hbar^2 \boldsymbol{q}^T (M^*)^{-1}\boldsymbol{q}}{2} \tag{4.3}$$

where subscripts on the mass tensor indicate that the band curvature can vary between directions in momentum and where the expansion is carried out around a momentum point $\boldsymbol{k}_0$ symmetry equivalent to the valence band edge—frequently, a high symmetry location. In our case, we can restrict ourselves to considering an isotropic mass, the generality does not help us.

$$E_{VB}(\boldsymbol{k}_0+\boldsymbol{q}) - E_{VB}(\boldsymbol{k}_0) \approx \frac{\hbar^2 q^2}{2m^*}. \tag{4.4}$$

What happens if we now apply a weak but long-range periodic potential $V(\boldsymbol{r})$? Now, the situation should resemble very closely the process by which we construct a crystal's bands in the presence of the atomic potential. The quadratically dispersing band at low energy plays the part of a free electron, albeit with a different mass m^* which may even by hole-like, and the long range superlattice potential $V(\boldsymbol{r})$ plays the same role as the atomic potential. Seen in this light, at low energy the superlattice acts like a crystal with very long lattice constant built up out of the low energy excitations of the parent material's quasiparticles. What comes next is familiar: Bragg scattering by $V(\boldsymbol{r})$, which is peaked in certain places in the superlattice mini-Brillouin

zone, opens avoided crossings in the free-electron-like dispersion and creates the crystalline band structure. These new bands are what we call minibands: they respect the symmetries of the mini-Brillouin zone, or minizone and they describe the excitations of the system at low energy.

The lowest energy state as a total bandwidth which is limited roughly by

$$W < \frac{3}{2}\frac{\hbar^2 G_m^2}{2m^*} \propto \frac{G_m^2}{m^*} \propto \frac{1}{m^*\lambda^2}. \tag{4.5}$$

The last two expressions involving the moiré reciprocal lattice vector $\boldsymbol{G}_m$ and the moiré lattice constant λ are equivalent ways of looking at the bandwidth in terms either of reciprocal- or real-space. If we want a flat band, it is sufficient to make the minizone small, and it helps to work with a very massive band, unlike graphene, when we go to design superlattice flat bands.

What about correlations? As in t-2LG, if the bandwidth falls below the Coulomb interaction energy, then we can expect correlations. The Coulomb interaction scales according to

$$U = \frac{e^2}{4\pi\epsilon r} \propto \frac{|\boldsymbol{G}_m|}{\epsilon} \propto \frac{1}{\epsilon\lambda}. \tag{4.6}$$

This is the first of two linear-quadratic crossovers which are relevant in the moiré physics of two-dimensional materials. Because the interaction energy falls only linearly while the bandwidth falls quadratically, to engineer a situation where correlations can take place requires only increasing λ subject to constant moiré perturbation potential $V(\boldsymbol{r})$ so that $U \gg W$. Careful choice of materials with a large effective mass and less dielectric screening can also be beneficial, leaving aside material dependent factors which influence the strength of V_m.

4.2 Getting an Interaction Potential

This discussion makes it clear that to get flat bands in quadratically dispersing bands we "just" need to turn on a moiré potential with a sufficiently long moiré wavelength λ. What it still leaves mysterious is where we can possibly get such a thing. After all, graphene has a very strong tendency to interlayer hybridize even when mis-stacked: in a simple picture of the Slater-Koster tight binding model in t-2LG, it's clear that the local registry between layers changes periodically with λ which provides a strong, hybridization driven V_m, even if the flatness of the bands is not immediately apparent. For this reason, when early photoemission studies were conducted on rotationally misaligned graphene bilayers, large hybridization gaps were observed at the intersection of graphene (*91*).

In studying general processes, it still pays to examine specific exemplars. To build an understanding of how an appropriate moiré potential arises in the massive 2D flat band context, we will restrict our attention to superlattices of WS_2 and WSe_2. These two materials are part of a larger family of 2H-TMDCs—WS_2, WSe_2, MoS_2, $MoSe_2$, among others—possessing practically identical, semiconducting band structures. Although these materials were long studied in their bulk forms, their relevance only increased after the isolation of graphene and later TMDC monolayers, tightly bound excitons with rich phenomenology, and spin-orbit physics leading to well defined spin-orbital texture at the K and K' valleys. This incredible phenomenology

requires the existence of a direct optical gap at the K and K' points in monolayers. However, for all thicker 2H-TMDC systems from bilayer through the bulk limit, the optical gap is indirect, only making a crossover to a direct K/K' gap in the monolayer limit. How does this occur?

4.2.1 Orbital Structure of 2H-TMDCs WS_2 and WSe_2

Like graphene, WS_2 and WSe_2 have places in their band structure where hybridization can occur between more strongly out-of-plane orbitals. Whereas at the K point in graphene p_z orbitals are responsible for hybridization, the K point in the valence band of WS_2 and WSe_2 is composed of in-plane W d orbitals and chalcogen p_x, p_y orbitals. Due to the van der Walls gap, these orbitals do not hybridize much in the multilayer and bulk limits, leading to essentially identical band structures in these Brillouin zone regions. Following the valence band toward the Γ point the orbital composition changes smoothly so that at the Γ point the orbitals are $d_{3z^2-r^2}$ and p_z derived. The consequence is that these bands hybridize strongly in multilayers, leading to bonding and anti-bonding sub-bands at the Γ point. This same effect is responsible for the increase in bandwidth at Σ in the conduction band and the crossover from direct to indirect gap as the second layer and subsequent layers are added.

Whereas misaligned graphene heterostructures still hybridize significantly, with ARPES sequelae like the large interlayer hybridization gaps, optical measurements of TMDC heterobilayers are consistent with a picture closer to non-interaction than the strong hybridization seen in aligned homobilayers. In concordance with this picture, photoluminescence experiments see emission consistent with direct gap excitons at the intralayer exciton energies for each layer comprising heterostructures and twisted bilayers. The microscopic mechanics for this are simple. Imagine that you randomly stack two layers of TMDCs on top of one another and choose an atomic site X in the upper layer L. As you walk by lattice vectors to atomic sites $\{X'\}$ equivalent to X (for instance, from transition metal to transition metal sites in MX_2), the local environment seen near $\{X'\}$ changes quickly. Consider now the process of bringing the two layers in contact so that they lie a distance d from one another. When d is much larger than the lattice constant and the sheets are far apart, then the energy per unit cell of the layer L is a fixed constant E_0 independent of d. As the sheets are brought closer, van der Waals forces arising from zero-point fluctuations in the electric dipoles of the layers permit decreasing the energy by a factor proportional to d^6.

The layers settle at a nonzero van der Waals contact distance, comparable to the interlayer distance in 2H-MX_2, because it eventually becomes costly to bring the layers closer towards one another because of Coulomb interactions and Pauli exclusion. It's natural to ask whether this distance is the same, slightly smaller, or slightly larger than the interlayer distance in an aligned bilayer.

To answer this without resorting to complex theoretical modeling, we will assume that we can separate out the energy contributions of each atomic site X_i in the unit cell of layer L, and that the energy change from its contribution to E_0 in the isolated monolayer at interlayer distance d, angle θ, and spatial phase (x, y) is $\delta E(d, X_i, \theta, x, y)$. The arguments to δE are arranged to suit

the problem we are considering: we need to sum over X_i in the unit cell and integrate over the geometry (x, y) under fixed superlattice structure θ to determine the total energy change δE_{tot}

$$\delta E(\theta) = \min_d \sum_{X_i \in L} \int_{x,y \in S} \delta E(d, X_i, \theta, x, y). \tag{4.7}$$

Here, there is a subtlety which concerns the integral region S in the superlattice. If the superlattice is incommensurate, then the integral is conducted uniformly over the superlattice, while if it is commensurate then the integral can be peaked in a subregion or even a discrete set of points in the superlattice. This can, in principle, give a qualitatively different flavor to commensurate structures than to incommensurate ones, especially commensurate structures with relatively short superlattice wavelengths. Most randomly stacked layers are not close to a short superlattice commensurate structure though, so we can safely consider what happens in the purely incommensurate case.

In this situation, the integral is evenly weighted over $x, y \in S$ and as a result, the energy extrema set the behavior. Because the energy grows so rapidly below the van der Waals contact distance with decreasing d, only a small region in L needs to be large so that $\int_{x,y \in S} \delta E(d, X_i, x, y) > 0$. In practice, this means that we should expect the interlayer distance to be slightly larger than the interlayer distance for an aligned homobilayer. Randomly stacked layers cannot settle so that the upper layer can fit into the grooves presented by the corrugated surface of the lower layer: they must sit "on top" of the corrugations instead.

We now have a plausible mental model as to why heterobilayers and twisted homobilayers retain their monolayer-like characteristics and direct and edges. The van der Waals contact distance may be too large for significant hybridization to occur.

4.2.2 Long wavelength superlattices

In the prior analysis, we triggered the layers as rigid so that we only had to think about the least energetic contact distance for the layers. The layers will not be completely rigid if the introduction of strain permits the total lowering of energy of the system. How much energy can be saved this way? Recall, we concluded that the layer settled at the larger interlayer distance to avoid particularly costly local stackings. If strain is should be introduced, it will happen to maximize the size of the regions over which low energy stackings can be assumed while still avoiding costly registrations with the lower layer. The energetic balance of introducing strain into the system can be calculated by comparison of a penalty term quantifying the cost E_s to strain the layer L from its equilibrium lattice sites and a reward E_h quantifying the change in energy resulting from assuming more favorable local stacking orders. Quantitatively, these terms can be described in terms of their scaling with the supercell size, λ.

Strain will be introduced around the low energy stacking regions to increase their size at a cost proportional to the circumference of these regions. In the limit of larger λ, strained regions appear as nearly one-dimensional domain walls where the atomic registry must adjust from one low energy stacking order to another. Consequently, $E_s \sim \lambda$: the energy cost of straining the

system increases proportionally to λ because the length of the strained regions, which are quasi-one-dimensional subregions in the supercell, scale with the supercell lattice constant.

On the other hand, the size of the low energy stacking regions increase according λ^2. This is the second linear-quadratic crossover in the superlattice physics of two-dimensional materials. In the eventual limit of large λ, the energy saved by reconstructing the lattice to allow proportionally larger low energy stacking regions overwhelms the cost of the atomic reconstruction. This makes long wavelength superlattices qualitatively different than short wavelength ones and explains the origin for the supercell periodic potential modulation V_m which can create flat bands at the massive valence band edge.

So far, this does not explain how large V_m can be, nor the microscopic reason behind the variation in energies of the different local stacking orders. We discussed how the corrugation in the charge distribution of the underlayer provides a modulation and this is extremely relevant in setting the difference in energy between local stackings. The other significant factor we have already encountered: it is the orbital hybridization which lowers the total energy of the two layers and permits tighter stacking of the sheets. To determine an approximate size for V_m, we can consider the energy difference between the available high symmetry stackings which can be assumed in the supercell. For TMDCs because of their two-atomic basis, this can differ between near alignment ($\theta \approx 0°$) and anti-alignment ($\theta \approx 60°$).

4.3 Correlates and Consequences of General 2D Material Flat Bands

Extremely flat bands are not unique to the situation of magic-angle twisted bilayer graphene but arise practically by necessity from geometric consequences of the energy landscape of stacked bilayers. We built a mental model illustrating how relatively large V_m can arise from energetic differences in local stacking orders in the superlattice and how this can lead to strain, Bragg gaps, flat bands, and superlattice hybridization. The downstream consequences of the flat band physics in TMDC superlattices are not straightforward to predict, but we can anticipate that electron-electron or hole-hole correlations can dominate the total mini-bandwidth by a large factor. A large variety of phenomena, overlapping but distinct from the phenomena in twisted-bilayer graphene superlattices, were quickly observed.

In WS_2/WSe_2 superlattices, the most direct correlates of the moiré potential were the observation of multiple split intralayer exciton peaks arising out of the formation of K and K' point minibands. Indirect optical measurements which relied on photoluminescence detection of the in-plane transport to a monolayer region and separately on detection of the dielectric constant seen in an hBN separated monolayer sensor as modulated by conducting or insulating states in a nearby heterostructure were further able to demonstrate that a Mott insulating state existed at half filling of the superlattice. They further demonstrated not only same-site avoidance but next-site avoidance: generalized Wigner crystallization for partial occupation of the superlattice at a variety of fractional fillings. Later measurements spatially imaged the Wigner crystal states with STS and their homogeneity and stability with MIM. To date, because of the comparatively short superlattice wavelength λ and the moderate effective dielectric constant in TMDC superlattices,

these are some of the most robust correlation effects thus observed, as we might expect from the Coulomb interaction strength

$$U_n \approx \sqrt{\frac{n}{n_\lambda}}\frac{e^2}{4\pi\epsilon\lambda} \approx 180\frac{\epsilon}{\epsilon_0}\sqrt{\frac{n}{n_\lambda}}\ \text{meV}. \tag{4.8}$$

The term involving the filling approximates the distance between electrons in the superlattice normalized to the superlattice constant. Using a reasonable value for $\epsilon \approx 4\epsilon_0$ we find that the Coulomb potential in the Mott state between neighbors is comparable to 45 meV while it is close to 26 meV and 37 meV in the $1/3$ and $2/3$ filling Wigner crystal states. This is not a perfect description on two accounts. First, in more complex fillings distance between charge carriers is not well approximated. Second, the shape of the Coulomb potential is modulated as a function of the transport characteristics and the distance between carriers, though the Coulomb approximation is not far from the truth at large λ.

For most condensed matter physicists, superconductivity is the most exciting of the moiré flat band sequelae. Although transport measurements are much more difficult on TMDCs and TMDC superlattices large Schottky barriers at the contact lead to large contact resistance and inefficient current flow into TMDC devices, limited evidence on twisted double bilayers of WSe_2 appear to show many zero resistance domes at partial filling of the superlattice, much the same as in magic-angle bilayer graphene, magic-angle trilayer graphene, and other derivatives and indicative of possibly superconductivity (*95*). Comparably, zero resistance domes at low temperature flanking an insulating state has been observed in twisted double monolayer WSe_2 (*93*). Furthermore, aligned hBN-ABC trilayer graphene, one of the systems where moiré flat band phenomena were first proposed after twisted bilayer graphene, exhibits robust superconductivity. At a categorical level, hBN-ABC trilayer graphene fits much better into the massive moiré flat band context we have discussed with a small V_m originating from roughness corrugations via the hBN lattice incommensuration, although very recent observations of superconductivity even in ABC trilayer graphene without a moiré complicate this categorization somewhat.

4.4 Electron Localization and Hybridization in WS_2/WSe_2 Superlattices

Given the existence of correlated states and superlattice-induced excitonic effects in WS_2/WSe_2 and twisted TMDC superlattices, it becomes more valuable to determine whether the general picture we gave is born out in these concrete material platforms. Because superlattices have inherently less well defined real-space and momentum space structure, both local, spatially resolved electronic measurements and reciprocal space electronic measurements offer direct insight into the origins and manifestation of the electronic superlattices in TMDCs.

ARPES is one of several compelling tools which can provide insight into superlattice phenomenology. Valuably, photoemission can disentangle K and Γ point states, can peer into filled states at higher energy, can examine the issue of interlayer hybridization in heterostructures, and can directly measure the local band curvature or flatness. In considering photoemission experiments on TMDC superlattices, we determined it would be appropriate to

work with WS_2/WSe_2 superlattices over twisted TMDC alternatives for practical and scientific reasons. WS_2/WSe_2 is a lattice incommensurate moiré rather than twist-angle controlled moiré which means that the sample fabrication procedures are more robust and tolerant to twist angle inhomogeneity, Figure 4.1. Because WSe_2 has an approximately 4% longer lattice constant than WS_2 ($a_{WSe_2} \approx 0.330$ nm, $a_{WSe_2} \approx 0.315$ nm), WS_2/WSe_2 has a robust long wavelength superlattice within a degree or two of alignment. This in stark contrast to the magic angle conditions in graphene where the control of interlayer alignment to the scale of a few hundredths of a degree is essential to adequately tame the Dirac bands. The quadratic dispersion of the band edge also reduces the dependence of the superlattice physics to the precise geometric details, leading to a "flat band continuum" in twisted homobilayers (*93*).

These practical reasons also help underpin WS_2/WSe_2's role as one of the best studied moirés. Using TMDCs provides a clear scientific advantage too: studying the unmeasured momentum-resolved electronics makes it possible to connect a wide body of optical literature to the superlattice-induced electronic changes by measuring optical responses like the optical absorption contrast—mapping the dipole allowed exciton features—on the same samples used for ARPES.

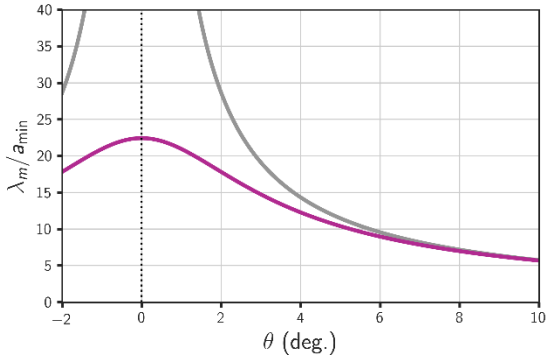


Figure 4.1 **Differences Between Lattice Incommensurate and Twisted Superlattices** Moiré superlattice wavelength near alignment for lattice incommensurate heterobilayers (purple) and twisted homobilayers (grey). For lattice incommensurate bilayers, variation in the superlattice wavelength is softened near alignment.

4.5 Preparation of Moiré WS_2/WSe_2 Samples for ARPES

We previously discussed the requirements and challenges of surface cleanliness for ARPES, and how these interact with the beam sizes available in current generation photoemission experiments. Here, we discuss how we prepared superlattices samples which satisfy the requirements of clean, exposed surfaces and large enough homogeneous regions for photoemission. These requirements are not more or less challenging than those of other very surface sensitive techniques like scanning tunneling spectroscopy but reflect a different set of challenges: STS makes do with a very small sample region with an atomically clean surface whereas ARPES requires a very large homogeneous region with high cleanliness.

While wet transfer of two-dimensional material flakes was the approach of choice at the birth of the field, dry transfer which minimizes exposure of the sensitive monolayers and thin films to chemicals and solvents used in wet transfer processes has emerged as the approach of choice for

superior cleanliness and more robust control of the in-plane and out-of-plane sample geometry. Under modern polymer dry transfer techniques, an encapsulating buffer layer can be picked up with a polymer stamp and used for the remainder of the transfer. This means that flakes and films constituting the active part of the sample never have to touch polymer, but instead only form van der Waals contact with neighboring layers. Trapped atmosphere and impurities between layers are forced into blisters whose density can be controlled by tuning parameters of the transfer process, and, in practice, samples of several layers can be formed with atomically clean interfaces over hundreds of square microns. In combination with inert atmosphere exfoliation of thin films and inert atmosphere transfer and encapsulation, the range of materials which can be studied in nano-electronic device configurations has expanded from graphene to air sensitive compounds like WTe_2, $NbSe_2$, CrI_3, and others. The fine control over fabrication geometry afforded by the dry transfer process has also made it possible to construct samples with desired interlayer twist angles. Using crystal habit of monolayers, it is possible to assemble heterostructures with an interlayer alignment precision of about 1°. By tearing monolayers and intentionally rotating the remaining portion of sample before resuming transfer, precisions below 0.1° which have enabled studies of magic-angle t2LG, t3LG, and twisted bilayer and double-bilayer TMDCs.

In optical and transport studies, the inclusion of an insulating encapsulation layer does not hinder experiments. Insulating hBN increases rigidity, keeps samples clean, prevents shorts to electrical contacts, and is optically transparent. However, for STS and ARPES, an encapsulating layer prevents access to the sample surface and prevents measurement entirely. In ARPES' case, a very thin encapsulating layer is not strictly verboten: one must only pay the fine of inelastic scattering by the encapsulating material reducing the photoelectron flux reaching the analyzer. Whether a buffer is present is more than just a measurement time concern. Without a buffer layer and using standard polymer transfer techniques, the measured sample surface is exposed to polymer during the transfer process which may leave residues preventing measurement or which substantially dope semiconducting samples.

ARPES also requires a good path to ground so that electrons removed from the sample by the photoemission process can be replenished before the arrival of the next x-ray pulse, or else in the steady state the sample potential will drift away from the vacuum potential, ruining measurements. Remarkably, hBN thinner than 20 nm or so will provide an adequate path to ground for photoemission measurements conducted with soft x-ray photons—perhaps because the x-ray pulse excites enough carriers in the hBN that charge can be replenished in combination with the comparable penetration depth of the x-rays ($2\pi\hbar c/70\ \text{eV} \approx 17\ \text{nm}$). This means either thin hBN or graphite can be used to create an atomically flat, clean, and rigid substrate for samples during measurement. Collective experience has shown that a flat substrate of this kind is essential to improve the photoemission quality from 2D materials because of the dielectric roughness of common substrates like p-Si and TiO_2. These constraints leave not too much room for imagination when it comes to sample formats. For our measurements we adopted vacuum exposed or graphene capped heterostructures on either graphite or hBN, as shown in Figure 4.2.

In the next section, we briefly discuss how we fabricated samples on this type, before turning to experimental results.

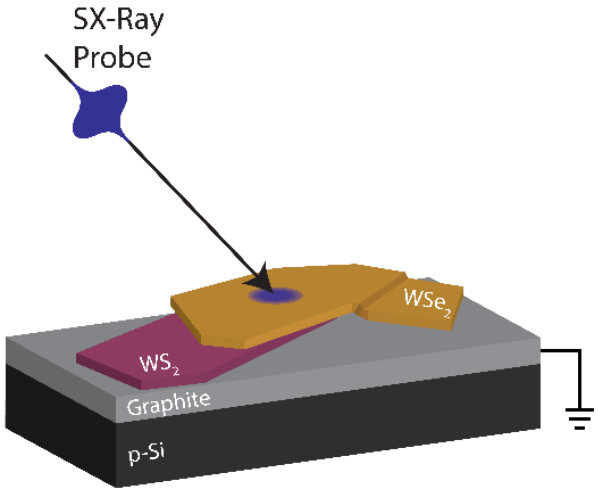


Figure 4.2 **A schematic diagram of a prepared ARPES WS_2/WSe_2 sample.** Vacuum exposed WS_2 and WSe_2 regions are available for monolayer measurements which permit sample quality assessment. An aligned heterostructure region a few times to a few tens of times larger than the x-ray probe beam sits on top of either graphite or hBN. Under graphite, the p-Si substrate forms a conducting path to ground required for photoemission measurements. Using hBN as a substrate without a separate grounding contact is acceptable in the place of graphite so long as it is thinner than 20 nm or so. Because the conductivity of TMDCs is very low, heterostructures should either have a top grounding contact of graphene, in electrostatically gated geometries, or should be grounded directly to p-Si with thin hBN or graphite.

4.5.1 Flip-and-place Polypropylene Carbonate Transfer

To prevent surface contamination while eliminating a cap, we adopted a technique where transferred polymer stamps were inverted and placed polymer side down onto the target substrate after separation from the remainder of the stamp (*59*). This differs substantially from standard transfer approaches where a liftoff process is used to separate the polymer layer from the sample. Using the flip-and-place approach, shown in Figure 4.3, the pickup order is also inverted so that the more sensitive TMDC layers are transferred last and do not directly contact the transfer polymer used. Polypropylene carbonate (PPC) stamps are stretched over patches of polydymethyl siloxane (PDMS) affixed to glass slides which form the base of the working transfer stack. During pickup, the PPC stamp is heated to approximately 45° C—above the glass transition temperature of PPC—to improve adherence of the PPC to 2D material flakes. Inverse transfer order benefits PPC transfer, which otherwise often requires pickup of a large piece of hBN or graphene to start transfer of monolayer flakes.

After transfer, the PPC layer which remains between the sample and the p-Si substrate must be removed. Unlike transfer processes using a polymer wash to dissolve residues of the transfer polymer which would cross contaminate the vacuum exposed surface, the technique used relied on sublimation of the thin PPC layer from below the sample by anneal in high vacuum at $150° - 250°$ C. This high vacuum anneal stabilizes the geometry of the sample which may settle as the comparatively very thick polymer layer sublimates from beneath the sample, but it does

not remove all the residues from the sample and adsorbed atmosphere which prevents photoemission measurements. For this reason, we clean samples with a second vacuum anneal in UHV at pressures around 1×10^{-9} torr and temperatures exceeding 350° C for 24 hours prior to ARPES measurement in a vacuum chamber attached via UHV to the measurement chamber.

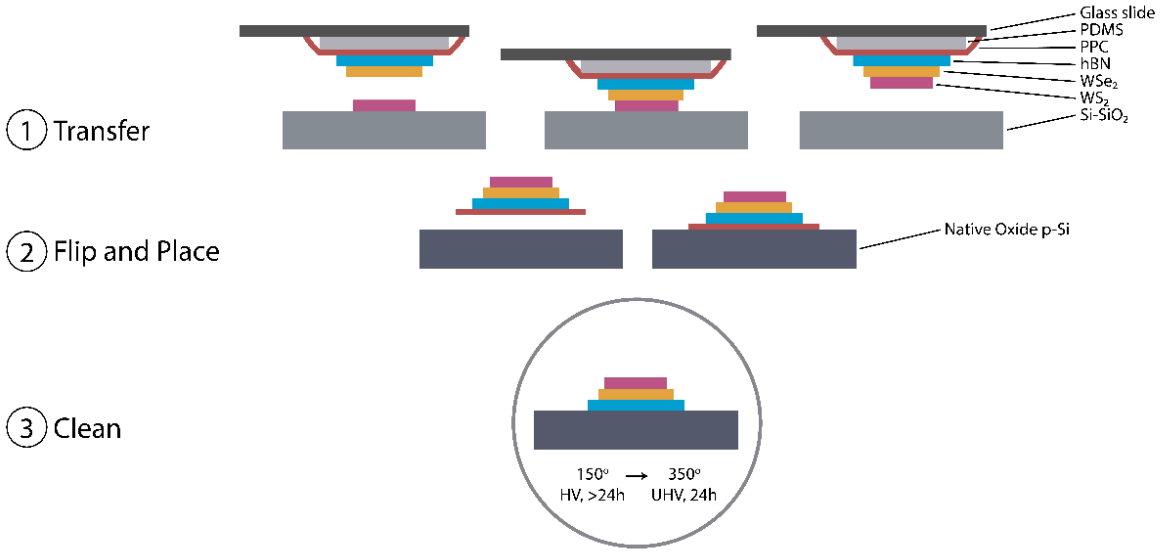


Figure 4.3 **Schematic diagram of the PPC flip-and-place transfer procedure.** Transfer proceeds by inverse order assembly to standard dry-transfer techniques, so that the vacuum exposed layer, depicted here as WS_2, is picked up last. After the final pickup step, the PPC film with the sample is carefully removed from the PDMS stamp with tweezers and laminated against a p-Si chip which will be the substrate for measurement. The sample, substrate, and PPC film are briefly annealed to 80° C, below the melting point of PPC, to ensure adequate lamination. Finally, the PPC is sublimated over 12h-48h from below the sample by a HV and subsequent UHV anneal at temperatures up to 250° C. Samples are annealed again prior to measurement with ARPES.

4.5.2 Polyethylene Terephtalate Transfer

Experiments were also conducted with samples assembled using dry transfer by polyethylene terephthalate (PET) (*96*). During PET transfer, the stage is heated to 60° C for each pickup step before depositing the entire stack on a p-Si substrate at 130° C. The transferred 2D material stack sticks to the substrate and disconnects from the PET stamp during the final placement step, although residues remain. As in PPC transfer, samples were cleaned by HV anneal and subsequent UHV anneal according to the same procedures.

4.5.3 Determination of the Local Interlayer Twist Angle after Fabrication

Even despite care in assembly and the insensitivity of TMDC moirés to small variations in the interlayer alignment, we need some way of determining the alignment during or after the experiment. A combination of factors gives a robust way of estimating the alignment to the required precision. Second harmonic generation spectroscopy (*8*) can be used both before and after sample assembly to determine the alignment coarsely. In situ, the angle can also be estimated to within a degree from the momentum mismatch of the valence band edge of the Γ point d_{z^2} bands. Most precisely, the orientation of diffraction peaks from a sharp band like graphene due to the moiré superstructure (diffraction by G_m) provides sub-degree accuracy to the

distribution of alignments under the beam during measurement. In the discussion of experiments conducted, we will use conservative estimates by saying that the wavelength of aligned superlattices exceeds 7 nm.

4.6 Band Structures of Semiconducting WS_2 and WSe_2

To understand superlattice induced changes in WS_2/WSe_2 moirés, we need to have a picture for the band structures in the parent materials: monolayer WS_2 and monolayer WSe_2. To provide this, we will use the same sample fabrication methodology and the same experimental probe as in the next section where we consider heterostructures. This consistency is valuable both for setting ground expectations for ARPES on TMDCs and two-dimensional materials, and because some of the phenomenology of TMDC multilayers will be useful in the interpretation of superlattices and heterostructures later.

This section is not purely introductory though: monolayer and multilayer TMDCs are interesting in their own right. Furthermore, comprehensive pictures of the electronics stemming from direct measurements of the electronic bands, as nanoARPES provides, are only just emerging. Although ARPES often suffers from a last mover advantage, the phenomenology of 2D materials is surprisingly rich for two reasons. First, ARPES has unique sensitivity to momentum resolution and binding energies away from the valence band edge (or Fermi surface, in metallic and semi-metallic TMDCs) and with this vantage many phenomena like band hybridization, spin-orbit effects, orbital composition, and sequelae of the local dielectric environment are apparent in the spectral function. Secondly, there are simply a huge variety of degrees of freedom in two dimensional materials ranging from doping variations, thickness, substrate material, epitaxy and alignment, heterostructure effects, strain, and moiré effects. Each technique brought to bear on two-dimensional material systems has revealed more depth to the phenomenology.

For this reason, while the full electronic band structures of the bulk- 2H-phases are well known from angle-resolved photoemission spectroscopy and from theoretical models, the band structures in the few-layer and monolayer limit have not been explored comprehensively in high-quality exfoliated samples. For many systems, including the canonical semiconductor WS_2, a complete understanding of the monolayer to bulk crossover is insufficiently understood due to a combination of the sources of variation we mentioned above, and the reliance in the early literature on films grown by chemical vapor deposition.

In this section of the thesis, we will discuss the complete two-dimensional band structures for high-quality exfoliated TMDC films in the one layer (1L), two layer (2L), and three layer (3L) cases acquired via nano-ARPES. As compared to previous studies on these materials, we provide a systematic study by thickness on a single sample, thereby isolating multilayer effects from other sources of variance in practical experiments. By global fitting for the K point trigonal warping and mass, we find a slightly larger renormalized K point mass than previously reported, and in WS_2 a slightly smaller Γ point mass. Our measurements are the first direct ones for the trigonal warping parameters at the K point, which we use to populate a comprehensive set of parameters for the band structure in 1L, 2L and 3L films. As a point of caution, we demonstrate that even following stringent sample preparation protocols cannot prevent sources of substantial

variation across the surface of a single sample, by observing that local doping variations may change as dramatically as $24\ \mathrm{meV}/\mu\mathrm{m}$. This final finding indicates that using small submicron beams is necessary for studying two dimensional materials even when they are morphologically uniform over many times the beam size and suggests that studies aiming to resolve electronic characteristics across some parameter like thickness or strain should aim to use a single sample to establish a reliable phenomenology.

4.6.1 Motivating Continued Study of TMDC Thin Films

Interest in obtaining(*97*, *98*) and characterizing films of transition metal dichalcogenides (TMDCs) in the few-layer limit grew in the wake of the graphene revolution(*99*) and after the prediction and discovery of categorically different optical behavior in this limit stemming from large scale changes in the band structure with a direct/indirect gap transition (*100–102*). Even after nanoscale angle resolved photoemission spectroscopy (nano-ARPES) provided band structures for monolayers of semiconducting TMDCs (*51*, *103*), material physicists are revisiting these materials again, in the wake of the superlattice revolution. Additional scrutiny has provided countless reasons to pursue higher quality characterization of these materials in the few and multilayer limit and new appreciation for physical phenomena native to monolayers and their substrate interaction. Considering only a few recent points of interest, the discovery of superconductivity in electron doped TMDCs and the suggestion of topological superconductivity via hole doping demonstrate that the 2H TMDC family can hardly be characterized as plain semiconductors (*104*, *105*).

The range of possible phenomena and their character is necessarily constrained by fine details of the electronic structure, notably the band edge masses, the energy offsets of bands for optical applications, and—in the case of K point superconductivity—the trigonal warping (*104*, *106*) which can suppress superconducting pairing at higher doping levels. While previous studies have extracted critical parameters like the valence band edge masses in monolayers and in the bulk(*38*, *107*) and have linked direct electronic measurements to optical response (*51*, *107*), a comprehensive accounting of the full valence band details must be pieced together from various reports with some essential features still unmeasured. Layered over this patchwork are a series of practical questions related to interpretation. Given recent observations of dramatic consequences for the band structure arising from superlattices (*58–60*), do the shorter wavelength superlattices arising from substrate superstructure influence the low energy electronic structure in addition to creating novel minibands (*108–110*)? Are morphologically homogeneous regions of sample also electronically equivalent and does this influence how we should interpret spatially resolved measurements of sample band structure? Providing robust electronic measurements requires experimental holism so that environmental factors, such as the superlattice interaction, can be held constant.

4.6.2 Nano ARPES Measurements of 1L WSe_2 and 1L-3L WS_2

The schematic layout of fabricated graphite TMDC heterostructure samples is shown in Figure 4.4a. Grounding for photoemission was achieved by transferring exfoliated TMDC samples and a conducting graphite substrate onto native oxide p-Si as previously described. In Figure 4.4b, an optical micrograph taken before transfer onto graphite to maximize monolayer-multilayer

contrast is shown with different thickness regions indicated. Subsequent photoemission data discussed are taken at the approximate locations shown in white circles. Figure 4.4c shows a core level photoemission intensity image of the WS_2 sample, exhibiting uniform structure and absence of bubbles from transfer over the sample extent. By comparison to the optical image in Figure 4.4b, we identify regions with comparable integrated x-ray photoemission (XPS) intensity corresponding to monolayer, few-layer or multi-layer, and bulk thicknesses. Core level fitting for the spin-orbit split W $4f_{7/2}$ and $4f_{5/2}$levels, shown for a representative location in Figure 4.4f, provide another measure of sample quality not accessible by observing an absence of bubbles and uniform photoemission yield. By conducting this fitting across the entire sample surface, in Figure 4.4d, we observe clear contrast between few-layer and bulk portions of sample which manifest as moderate doping variations.

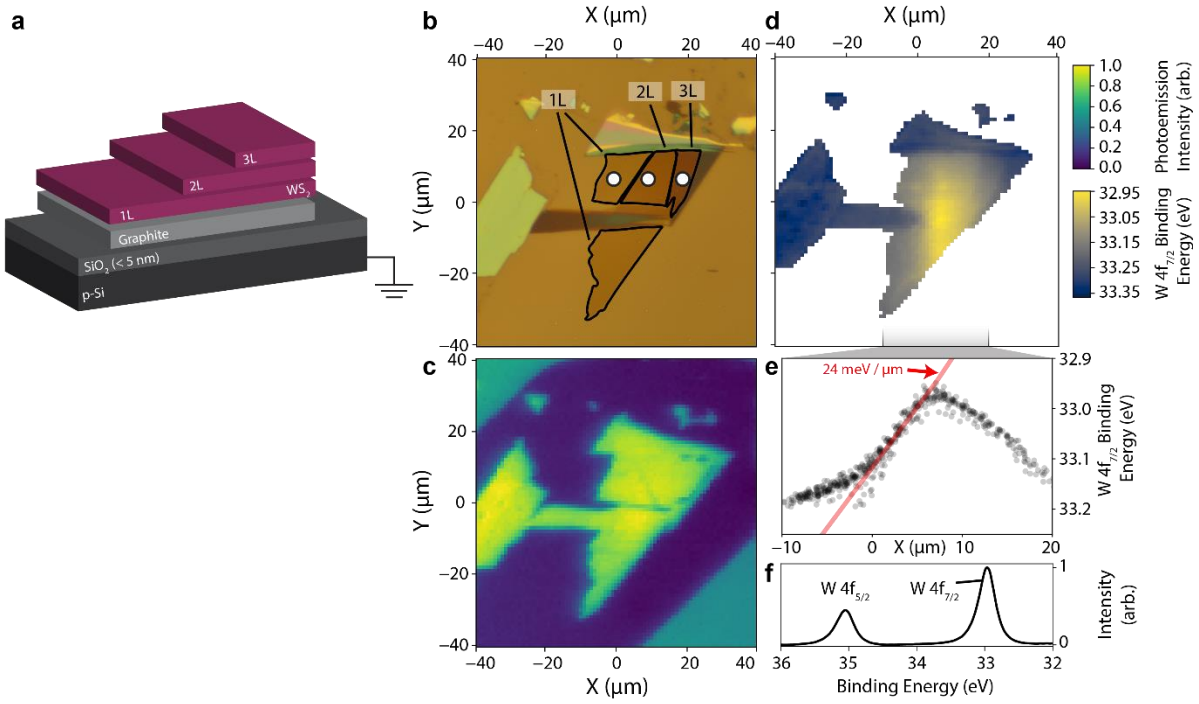


Figure 4.4 **Nano-ARPES of WS_2 multilayers and sample topography. [a]** The sample geometry adopted for angle-resolved photoemission spectroscopy measurements. To study the relative significance of intra-sample and inter-sample variance, photoemission studies were conducted on a multilayer with regions of varying thickness ranging from monolayer (1L) to trilayers (3L). Samples were grounded by placing them on atomically flat substrates of graphite on native oxide p-Si, as show. **[b]** Optical micrograph of the WS_2 sample before transfer onto graphite with sample thickness identified by micrograph contrast and corroborated by photoemission. **[c]** Total core level intensity after transfer, showing no multilayer/bulk contrast and revealing clean, bubble-free regions for sample measurement. **[d]** Global fitting for the binding energy of the tungsten $4f_{7/2}$ core level across the sample surface. A consistent, moderate doping effect is revealed across the sample surface. **[e]** Scatter of the core level locations between $X = 0\ \mu m$ and $Y = 12\ \mu m$ which reveals a consistent and moderate doping imprinted across the sample with varying X coordinate. The doping gradient is as large as $24\ mev/\mu m$ across the multilayer portions of the sample. **[f]** Representative core level spectrum at the tungsten $4f$ spin-orbit split doublet, indicating the core level peak which whose fits are shown in [d,e].

Variations in the $4f_{7/2}$ level binding energy is observed as a function of the X coordinate across the sample. We conjecture that this doping effect is due to slight polymer contamination from the transfer process rather than from defects induced by heating, because we do not observe that the doping level correlates with the distance from the edge of the sample or in other samples prepared under comparable vacuum annealing conditions. In Figure 4.4e, core level fitting results over the multilayer portion of sample are collapsed onto a single curve by projecting the Y value, giving an essentially universal doping profile across the measured X coordinate of the sample. The corresponding doping gradient is actually very large, as high as $24\ \mathrm{meV}/\mu\mathrm{m}$ in our samples. Sources of inhomogeneity in the apparent binding energy of features from photoemission may also arise from differences in local electrostatics and substrate dielectric constant (*111*). For CVD grown samples which have higher defect rates and correspondingly larger spatially inhomogeneous dopings (*112*), these resolution broadening effects place harsher constraints on resolution. Surprisingly, the doping gradients we observe indicate that doping variation presents challenges to precise electronic metrology on semiconductors via ARPES, even if uniform sample morphology and gentler preparation via exfoliation suggest otherwise.

For all photoemission measurements, a resolution of better than $1.5\ \mu\mathrm{m}$ was achieved using an x-ray capillary which results in 36 meV doping gradient broadening in the worst case, below other resolution contributions. Nonetheless, future study design should account for the difficulty in assuring electronically uniform conditions under currently available ARPES spatial resolutions and especially when working with beams wider than $1\ \mu\mathrm{m}$ which have been used in much of the existing literature.

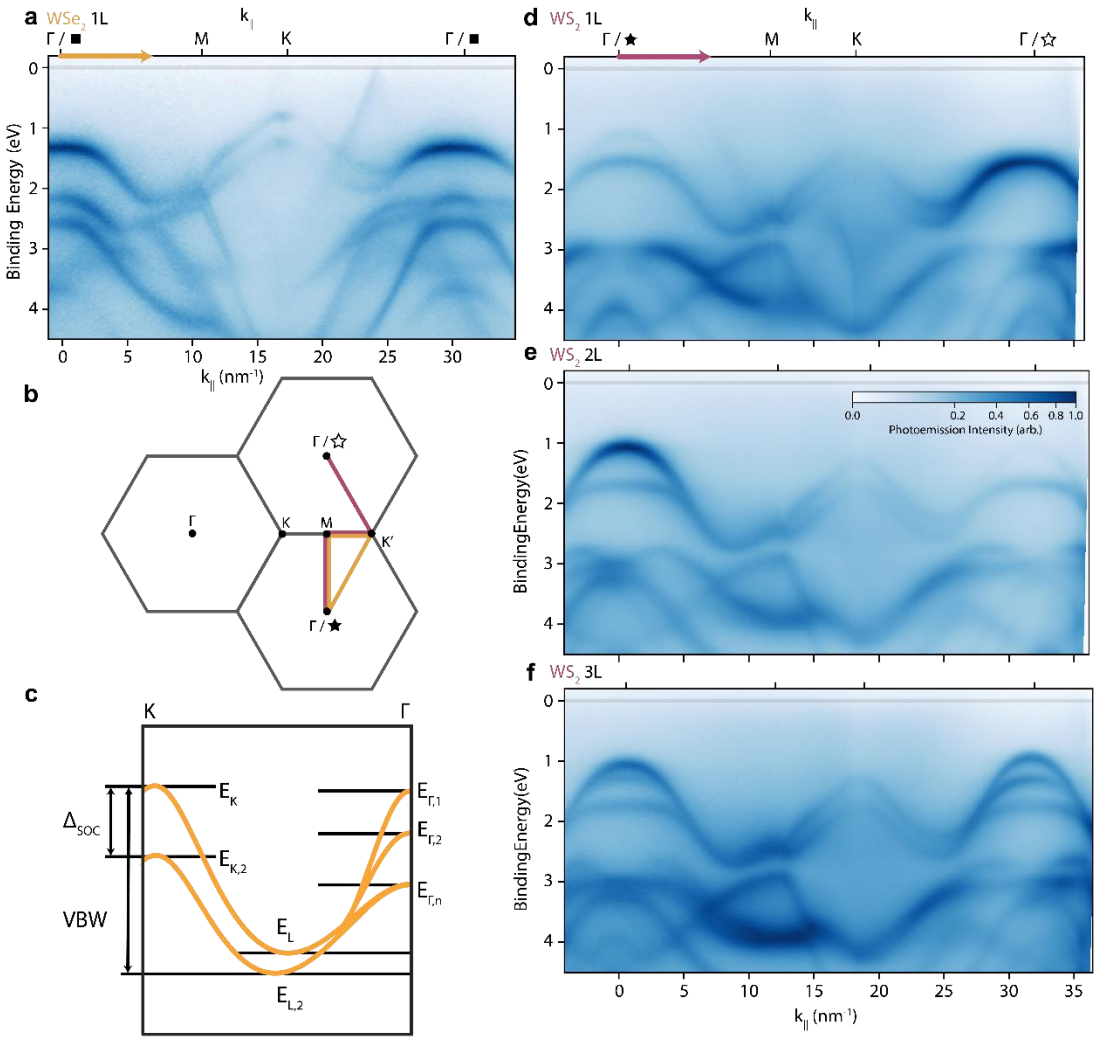


Figure 4.5 **Momentum resolved high-symmetry electronic structures of WS_2 and WSe_2. [a, d-f]** Combined high symmetry electronic structures for 1L WSe_2 [a], 1L WS_2 [d], 2L WS_2 [e] and 3L WS_2 [f]. Each plot shows the ARPES intensity measured over a combined path in momentum space covering the high symmetry directions in the band structure. Replica graphite cones due to substrate interaction are visible at $24\ nm^{-1}$ in [a] but are observed across the Fermi surface according to the superlattice resulting from large lattice incommensuration and interlayer twist between the TMDC layers and the substrate graphite. In the multilayer series [d-f], consistent, even unchanging, band structure is observed across the entire momentum window except for a few relatively narrow windows in momentum and energy. Principally, this is observed mostly at the Γ point for the valence band and next lowest energy band where interlayer hybridized sub-bands are made possible by an out of plane $d_{z^2-r^2}$ orbital composition. Continued evolution in the multilayer appears not to continue in the next-to-valence band and is observed only for the near Γ valence band manifold. ARPES matrix elements at the experimental geometry provide nearly exact cancellation of intensity across neighboring Brillouin zones. **[b]** Brillouin zone diagram and high symmetry paths for the band structures in [a,d-f]. The gold path shows the folded momentum path for WSe_2 in [a], while the purple path shows that for the multilayer sequence in WS_2 [d-f]. Filled and hollow stars identify different neighboring Brillouin zones while the zone on the left contains the zero-momentum point. **[c]** Schematic labeled band structure for the valence band in multilayer 2H-MX_2. Values extracted from photoemission results shown here are listed in correspondence to this diagram in Table 1.

Having ascertained a sufficiently uniform environment on the sample where representative WS_2 regions of different thicknesses are available, we turn to photoemission band structures and their evolution across thickness. Figure 4.5b and Figure 4.5d-f show the complete high symmetry band structures for 1L WSe_2, 1L WS_2, 2L WS_2, and 3L WS_2 respectively. A connected but nonlinear path traversing each high symmetry segment in the Brillouin zone is shown in Figure 4.5b and matching the lower axis labels in Figure 4.5a, and Figure 4.5d-f. The leftmost indicated zone in Figure 4.5b contains the zero-momentum point (normal photoelectron emission). Monolayer band structures (Figure 4.5a and Figure 4.5d) are in excellent qualitative agreement with previously produced results. Looking at the evolution in with layer thickness from 1L to 3L, we see that the band structure is essentially completely unchanged for WS_2 except for a few places in the Brillouin zone where substantial out of plane orbital composition permits interlayer hybridization to break the layer degeneracy. This occurs most dramatically at the Γ point where the valence band is split in a sequence of sub-bands which enumerate the layers. Interestingly, in the second Brillouin zone where measurements were conducted, per Figure 4.5b, not every Γ point state is visible in either zone due to a near total suppression of intensity due to ARPES matrix elements. We will return to a discuss of these low energy states when we perform a more detailed accounting of the electronic metrology later. To see any other evidence of changes in the multilayer sequence, we need to turn to the next-to-valence band at the Γ point, visible at 3 eV binding energy, or so. This band has substantial out of plane orbital composition across a wide momentum range and exhibits a split layer degeneracy across the entire ΓM momentum range. In Figure 4.5c, a schematic band diagram matching the multilayer valence band is labeled with relevant band parameters and sets notation for tabulated results across the presented monolayers and multilayer sequence (Table 1).

Sample	$E_{\Gamma,1}$	$E_{\Gamma,2}$	$E_{\Gamma,3}$	$E_{L,1}$	$E_{L,2}$	$E_{K,1}$	$E_{K,2}$	m_Γ^*/m_e	m_K^*/m_e	C_w^3	Δ_{SOC}
1L WSe_2	1.329	–	–	1.724	2.022	0.826	1.265	2.76 ± 0.1	0.53 ± 0.05	0.05 ± 0.01	0.439 ± 0.010
1L WS_2	1.610	–	–	2.306	2.589	1.386	1.801	2.01 ± 0.1	0.46 ± 0.05	0.05 ± 0.01	0.417 ± 0.005
2L WS_2	1.129	1.754	–	2.289	2.590	1.353	1.783	1.09 ± 0.1	0.53 ± 0.05	0.05 ± 0.01	0.436 ± 0.005
3L WS_2	1.010	1.369	1.824	2.298	2.583	1.345	1.863	0.73 ± 0.1	0.47 ± 0.05	–	0.446 ± 0.005

Table 1 **Band parameters and masses for WS_2 and WSe_2 multilayers.** Energies are given in eV units, mass renormalization ratios are unitless, and trigonal warping parameters are given in $eV \cdot nm^3$ units. Uncertainties on each estimate are discussed in the supplementary materials. Missing Γ point energies for thinner samples are due to fewer sub-bands in thinner samples.

Although ARPES has previously been used to perform characterization of single-layer (*51*, *103*, *111*), few-layer (*113*), and bulk (*114*) TMDCs under a variety of conditions, many details of their low energy band structures have not been quantitatively explored, despite practical relevance. Monolayers of 2H-TMDCs have brought the most interest because of their direct gap at the K point, but multilayers have been proposed as infrared sensors when doped so that the chemical potential falls across the lowest energy Γ point sub-band. The precise characteristics for these sensor applications relies on knowledge and control of the sub-band spacing, which makes understanding the interlayer coupling and hybridization at the Γ point of technological relevance.

At the same time, monolayer 2H-TMDCs have received immense interest as superconductivity has been achieved in electron-doped states at the K point of monolayer and few layer MoS_2 (*105*). Proposals for the hole-doped side of the spectrum posit possible topological

superconductivity whose pairing may be suppressed by the large native trigonal warping (*106, 115*) at the K point which distorts iso-energy contours at the K point (Figure 4.6a and Figure 4.6b-e), a feature TMDCs share with the archetype material graphene. At the same time, this trigonal warping has profound implications on nonlinear optical generation (*116*) in few layer films, and on the transport and spin Hall effect in monolayers (*117*). In TMDCs the simplest representation of the trigonal warping is formed by introducing a next-order cubic term to the dispersion relation (*106*). In this scheme, the massive quasiparticle dispersion relationship at the K point is modified in the presence of trigonal warping to include a cubic term in the K-relative momentum coordinate

$$E_K(\boldsymbol{q}) = \frac{\hbar^2 q^2}{2m_K^*} - C_w^3 |q|^3 \cos(3\phi) \tag{4.9}$$

In the dispersion relation above, $\boldsymbol{q}$ is measured relative to the K point for the TMDC layer and ϕ is the angle which the vector $\boldsymbol{q}$ makes relative to the ΓK direction. Despite its dramatic effect on the low energy structure at K and its relevance for hosting exotic states and in the optical properties of TMDCs, this band parameter remains unmeasured, despite having been commented on qualitatively in prior bulk measurements (*114*). Several reasons may explain the absence of this parameter in prior studies, despite its practical relevance, stemming from the difficulty in establishing a high-quality measure of this parameter. Because the K point intensity in tungsten based dichalcogenides is very weak due to the orbital matrix elements at soft X-ray energies used for ARPES, lineshape fitting is challenging. This problem is exacerbated by the vanishing intensity in the matrix elements past the K point in the ΓK direction. Most studies recover the K point mass using only the dispersion along the high symmetry ΓK direction, while the global trigonal warping structure can be inferred by considering all of the momentum points in a small disc around the K point.

This motivates using two-phase global fitting for the WS_2 dispersion around the K point where we first fit the constant momentum lineshape as a function of q for $|q| < 4\ \mathrm{nm}^{-1}$. These points are then filtered according to the quality of the fit and used to fit a global model for the dispersion as a function of the momentum offset q. This method permits using more EDCs away from the K point and away from any high symmetry direction to lend statistical support to the evaluation of C_w^3. This method permits recovering both the K point mass and trigonal warping parameter C_w^3 which provides excellent agreement across a large energy window (about 0.4 eV, see Supplementary Materials). Despite excellent agreement between the cubic model and the experimental dispersion inside a 0.4 eV window from the valence band edge, as can be observed in the overlaid model and experimental data for 1L WS_2 in Figure 4.6b-e. A step discontinuity is observed in the quality of results beyond, indicating a breakdown of this simplified low energy picture. Perhaps surprisingly, a larger effective mass is consistently observed for the valence band at the K point compared to previous values in the literature, which sit just about $0.4\ m_0$, when incorporating the warping. This discrepancy cannot be attributed to deficiencies in the quadratic approximation used in prior studies, because the mass is typically quoted in ARPES by fitting the half interval from K towards Γ, where the warping would contribute to produce a

larger mass rather than a smaller one. The very narrow momentum ranges which are typically used to extract these masses may indicate that this difference is due to unanticipated difficulty in reliably extracting masses over such small intervals. Another possibility, which we deem less likely than issues with current analysis approaches, is that there is more intrinsic variability in TMDC samples than has currently been assumed. Taken together, our new measurements of the trigonal warping parameter and differences in the reported mass suggest that more robust analysis incorporating local structure off the symmetry axis, as by global fitting, is appropriate to recover important band parameters in materials by ARPES.

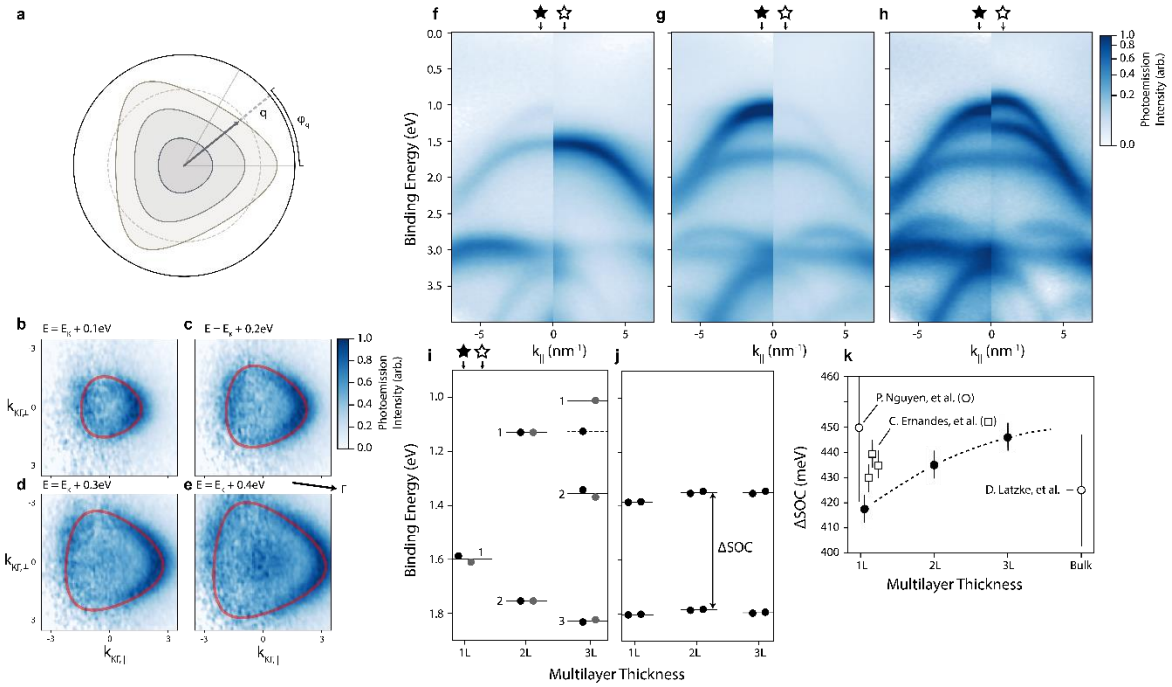


Figure 4.6 **Trigonal warping, band metrology, and multilayer band structure evolution in WS_2. [a]** Iso-energy contours for the model of trigonal warping given by Eq. 1. The wavevector displacement $\boldsymbol{q}$ from the high symmetry K point is shown with a black vector and decomposed into polar coordinates in Eq. 1 as illustrated in the diagram. **[b-e]** Experimental iso-energy contours in 1L WS_2 taken around the K point in 100 meV intervals. The top of the next spin-orbit split valence band is just visible at zero momentum offset in [e]. Red lines in each plot illustrate the results of global fitting for the dispersion using a mass and warping model (Eq. 2) and corresponding to the results in *Table 1*. Excellent agreement is found over a large energy range between experimental data and this two-component model. **[f-h]** Inset high-symmetry band structures around the Γ point for 1L, 2L, and 3L WS_2. Left and right halves of each plot come from neighboring Brillouin zones as indicated by filled and hollow stars in *Figure 4.5*c. In each panel, an affine background is subtracted by curve fitting to improve agreement in the photoemission signal between the two zones for comparison. In multilayers, a matrix-element controlled alternating ladder of states is observed at Γ, with essentially complete cancellation for the lowest energy states and imperfect cancellation for higher energy states (see also [i]). **[i]** Curve fitting results for the lowest energy Γ point states shown in [f-h]. Results coming from two different zones are illustrated with black and gray points. Higher energy interlayer states have photoemission weight in both measured Brillouin zones, as can be seen in the presence of both black and gray dots at 3L, level 3. **[j]** Curve fitting results for the K point spin-orbit split binding energies. A small shift in the binding energy of the K points is observed consistent with the rigid-band effect in *Figure 4.4*f. **[k]** The thickness resolved spin-orbit splitting from the curve fitting results in [j], together with values from the literature *(27, 114, 118)*. A consistent increase in the splitting is observed with increasing thickness.

At the Γ point, a variety of interesting effects can be observed as we consider the thickness dependence of our samples. As is well appreciated from theoretical modeling (*119–122*), orbital overlap and interlayer hybridization of the valence band at Γ (*106*, *123*) breaks layer degeneracy and leads to a ladder of states separated by only a few hundred meV. Across neighboring Brillouin zones, differences in the photoemission matrix elements due to the orbital symmetry of neighboring levels leads to essentially complete suppression of photoemission intensity at the lowest energy state in our experimental geometry (Figure 4.6g-3i). At higher energy, this effect is only approximate with substantial ARPES intensity visible in both neighbor zones for the lowest energy states. Extracted levels and binding energies are tabulated in Table 1 and represented schematically in Figure 4.6i and Figure 4.6j. A similar analysis of the K point binding energies is shown in Figure 4.6j. Minor shifts in absolute K point binding energies are in precise agreement with previously noted core level doping trends and corroborate the rigid band nature of this shift. As has been noted in some studies previously, the spin-orbit splitting in the K point states is suppressed in the monolayer as compared to multilayers (Figure 4.6k). Overall, the trend for Δ_{SOC} is not in excellent agreement with the bulk values.

4.6.3 Significance of ARPES Measurements on Multilayers

Our results establish a consistent and precise characterization of the band structures in the WS_2 family under controlled environment and substrate interaction parameters. By comparison to prior studies, we identify two sources of evidence—spatial doping variations and effective masses in moderate disagreement with the literature—indicating a preference for in situ comparative work if quantitative band parameters are to be extracted and compared between materials. For the first time, we more fully constrain theoretical models by providing high resolution measurements of the Γ point sub-band spacing and the trigonal warping at the K point. Through an inclusion of the trigonal warping and global fitting of ARPES data to a broader momentum and energy range around the K point, we find reason to assert a moderately larger effective mass in the valence band (hole mass) than has previously been measured in photoemission studies, with the trigonal warping parameters in excellent agreement with theoretical models to within experimental accuracy.

4.7 High Symmetry Spectra in Monolayers and Moirés

Having now taken time to see what the ARPES spectra of monolayer and multilayer TMDCs look like, and to evaluate some of the interlayer phenomenology which emerges as we transition into thicker films, we turn to our primary goal of studying superlattices.

Using nano-ARPES, we measure the electronic band structure in each of the parent monolayer materials, WS_2 and WSe_2, the band structure in a misaligned short wavelength superlattice, and the band structure in a very long wavelength moiré superlattice within 1° of crystallographic alignment in the heterostructure. The high symmetry ΓK direction gives us a very good picture of what is happening in the electronics of each of these systems. Before we delve into these results, we caution that the parent Brillouin zone high symmetry direction carries less meaning in each of the superlattices: the mini-Brillouin zone of the superlattice will be oriented at some entirely different direction which is a function of the misalignment and the presence of an interlayer

twist, especially for the intentionally misaligned samples, means that the direction presented can only be a high symmetry direction for one of the two layers (see Figure 4.7 caption).

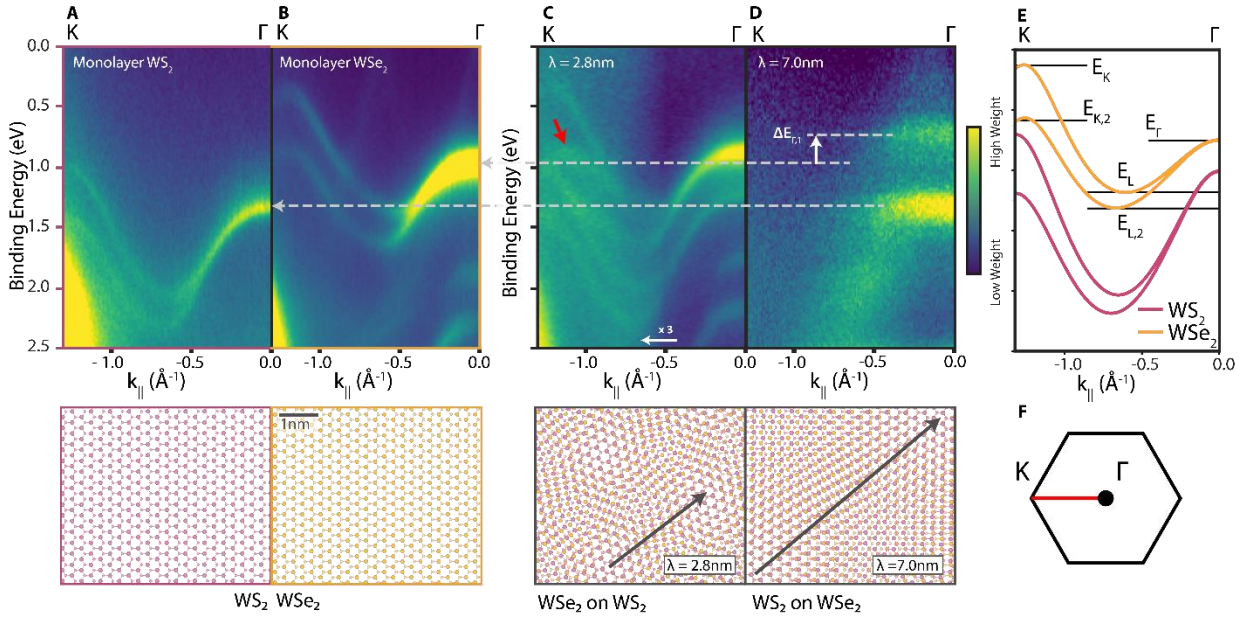


Figure 4.7 **High symmetry band structure for WS_2, WSe_2, WS_2/WSe_2, and moiré WS_2/WSe_2 [a]** WS_2 and **[b]** WSe_2 monolayer high symmetry spectra collected from nonoverlapping portions of the misaligned sample shown in **[c]**. **[d]** shows the aligned superlattice sample which has a moiré periodicity exceeding 7 nm. The real space atomic diagrams below each figure show the local lattice configuration in equal scale and the stacking order for the samples of [c] and [d]. [a] and [b] are in excellent agreement with previous ARPES spectra on semiconducting monolayer MX_2. In [c], we observe an inelastic mean free path weighted sum of the monolayer spectra from WS_2 and WSe_2 with a rigid band shift. However, in [d] the aligned superlattice samples have very different phenomenology characterized by flattening of the valence band and a much larger energy spacing between the WS_2 and WSe_2 originating bands. In [b] and [c] additional bands not attributable to the TMD layer can be seen. In [b] this is due to a diffraction-like replica of the underlayer graphite dispersion shifted by the moiré of the TMD and graphite. In [c], it is stray photoemission from a piece of WSe_2 which is rotated by 5° relative to the heterostructure. The beam size used in [d] and for all presented aligned heterostructure data was $1.5\mu m$ while a $10\mu m$ beam was used for [a-c]. Dashed lines are used to show energy levels of the valence band edge evolving in the heterostructure samples. The increase in the level spacing in the aligned sample is annotated as $\Delta E_{\Gamma,1}$. **[e]** Tight-binding calculation of the valence band structure for individual monolayers which show energy degenerate bands at Γ. The degenerate bands at Γ are split by the spin-orbit coupling in the materials at the K and K' points. **[f]** The direction of the cuts in [a-d] in the Brillouin zone of WS_2 (a), WSe_2 (b), WSe_2 (c) and both layers (d) respectively.

Figure 4.7a and b show the measured monolayer electronic structure in WS_2 and WSe_2 respectively, while Figure 4.7c shows the band structure for the 6° misaligned heterostructure and Figure 4.7d presents results collected from an aligned, long wavelength superlattice. In each case, results are presented along the ΓK direction for one of the constituent layers (in the case of Figure 4.7c which presents the only serious ambiguity, the WSe_2 layer).

First, we will consider the monolayer sample data in Figure 4.7a and b. The monolayer data are broadly in agreement with previous ARPES studies (*51*, *107*, *124*) and theoretical models (*119–121*, *125*). A schematic, TB based calculation of the lowest energy valence bands is shown in Figure 4.7e which permits a discussion of the core features. In semiconducting MX_2, locally

parabolic, hole-like bands at Γ disperse towards the zone corner. Towards the K and K' points, spin-orbit coupling breaks orbital degeneracy of the valence band. At the K and K' points, semiconducting MX_2 has an incredibly large spin orbit splitting, Δ_{SOC}, which in tungsten-based materials exceeds 0.4 eV. Bands at higher binding energy—not shown in Figure 4.7—have more complicated orbital character. The experimental linewidths in our monolayer samples are limited by the chosen spectrometer settings. Taken together with the excellent correspondence to theory, these narrow measurement linewidths indicate high quality of samples prepared according to the methods used in this study, which simplifies the interpretation of changes in heterostructure samples fabricated to the same standard.

The situation is very similar for the shorter superlattice wavelength sample ($\lambda = 2.8$ nm). The ARPES spectrum we record here (Figure 4.7c) is a shifted and weighted superposition of the spectral weight from the monolayer spectra we have already discussed. It is a weighted superposition because the electron inelastic mean free path suppresses photoemission intensity from the underlayer WS_2 relative to the vacuum exposed WSe_2 layer. By shifted, we refer to a rigid shift in energy towards the chemical potential by slightly less than 100 meV. This is consistent with an increased dielectric screening relative to monolayers on graphite (*111*).

The standout is the aligned, long wavelength superlattice sample, results of which are shown in Figure 4.7d. At alignment, the electronic structure is dramatically modified relative to the monolayers. We observe three main features

1. **(lineshape)** a substantial increase in the spectral linewidth of states and a decrease in energy symmetry around the peak line intensity throughout the entire momentum window
2. **(flattening)** a strikingly apparent flattening of the bands near the zone center
3. **(bonding-antibonding)** an increased binding energy offset, annotated in Figure 4.7 as $\Delta E_{\Gamma,1}$ between the two lowest energy bands in the heterostructure

Over the next few sections, we will build up an interpretation of these features, and contextualize them in terms of the evolution of interlayer coupling in 2D material superlattices. Before turning to that program, however, we first digress to consider whether issues with sample quality or properties of the measurement process could create spurious broadening of linewidth in long wavelength superlattices.

4.8 Intrinsic Linewidth Modification in Moirés, Ruling out Sample Quality and Measurement Effects

The observation of large linewidths and accompanying band structure changes is not enough to establish that our measurements show superlattice induced changes. After all, an increase in the linewidth is consistent with degraded sample quality and these issues plague x-ray ARPES measurements. Broadly, we can consider three possible classes of origins for the changes we observe: extrinsic effects, intrinsic effects, or final state effects and artifacts of the measurement. These are discussed below in turn.

Extrinsic effects – The fabrication process for our heterostructures is complicated and involves several steps which may reduce the quality of the photoemission data. As part of sample

fabrication, we must anneal samples, our materials may contact polymer despite use of the stack-and-flip procedure, and inhomogeneity from bubbles or local twist variation can all cause blurring of the photoemission signal and herald poor sample quality. We call on several strong pieces of evidence to rule out these effects. Fabrication methodology is not likely to be at play because monolayers fabricated via the same schemes exhibit sharp and intense (high yield) photoemission spectra. Because monolayer TMDC flakes are never perfectly matched for fabrication, exposed monolayer regions of WS_2 and WSe_2 are available on aligned samples during measurement and provide a control—much like a twin study—for damage which may have occurred to the heterostructure. Universally in monolayer regions neighboring aligned heterostructure portions of sample, we observe high quality photoemission. In the extreme limit, we can examine the edge of the sample where heterostructure and monolayer lie within the same 1.6 μm beam, as shown in Figure 4.8a. In this margin region, high diffuse heterostructure bands and sharp monolayer bands are both observed, which rules out sample quality problems. Our findings are also robust to the details of sample preparation, because we observe the same photoemission characteristics in samples prepared via PPC as in those prepared via PET.

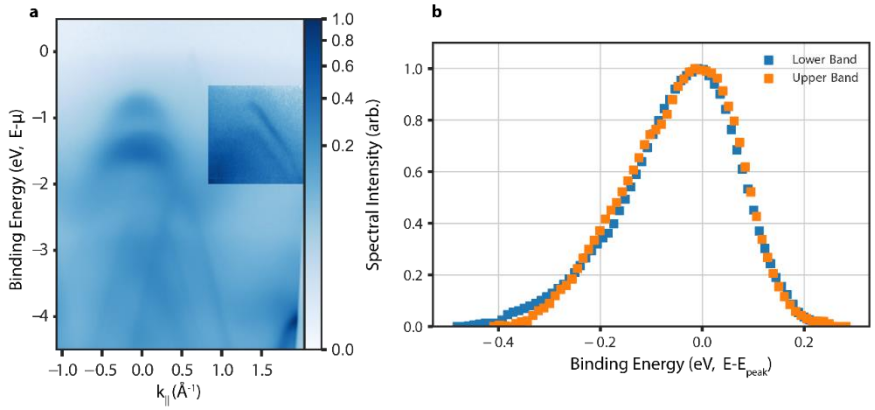


Figure 4.8 **Coexistence of Highly Dispersive and Sharp Spectra with Aligned Heterostructure ($\lambda = 7nm$) Bands. [a]** ARPES measurements at the monolayer border of an aligned superlattice sample. Within the same beam, both sharp monolayer samples and diffuse superlattice bands appear. The inset box showing the monolayer K point intensity has been rescaled nonlinearly and nonuniformly by histogram equalization to show that the K point for the monolayer hosts bands whose sharpness is limited by the spectrometer settings. **[b]** At the Γ point in superlattice samples, we observe strongly asymmetric lineshapes after subtracting a background to isolate the intensity from the band and shifting the energy so that the two peaks can be compared. The upper and lower band features refer to the WSe_2 and WS_2 originating bands in the heterostructure, shown, for instance, in *Figure 4.7*d. Nearly identical lineshapes are consistent with substantial variations in the binding energy across the supercell which come from modulations in the low energy stacking configuration throughout.

Twist inhomogeneity cannot exist for a monolayer, so direct comparison to monolayer photoemission cannot rule out this extrinsic cause. However, twist-angle inhomogeneity can be ruled out via the experiment geometry. If $\Delta\phi$ varied across the same, this would lead to a decreased resolution in the second Brillouin zone relative to the first, which we do not observe. We can even measure $\Delta\phi$ locally, by looking for replicas of the substrate photoemission which are shifted by G_m, the moiré reciprocal wavevector. This discussion will be left for a later section where we consider additional superlattice structures, but the conclusion is useful here: the angle

smearing in replica graphene photoemission in aligned samples restricts $\Delta\phi \approx 0.1°$. This is far too small to cause the effects we observe.

Intrinsic effects – The observed changes may be due to an actual change in the spectral function of superlattice WS_2/WSe_2. To be as measured as possible, we proclaim that the results are intrinsic on a diagnosis of exclusion. However, there is also some evidence we can consider which buttresses support that the changes are physical. STS measurements on $WSe_2/MoSe_2$ moirés, on WS_2/WSe_2 moirés measure a variation in the moiré potential depth of approximately 0.2 to 0.4 eV (*126*, *127*) depending on the chemical composition and superlattice wavelength. For moirés matching the characteristics of aligned WS_2/WSe_2, a value of 300 meV is appropriate and matches the observed degree of asymmetric broadening for our Γ point constant momentum spectra (Figure 4.8b). After subtracting a constant intrinsic linewidth of 130 meV measured from the monolayer K point, we obtain a broadening of 350 meV at Γ. These values are also broadly in consistency with theoretical calculations for homobilayer superlattices (*94*, *128*). Under this interpretation, the distribution of spectral weight from the Γ point also permits us to identify the portion of the supercell contributing the greatest spectral weight—unsurprisingly, most weight comes from states at lower binding energy.

Final state effects – The photoemission process is complicated because the final state diffraction can create replicas of the main band. In the presence of a very short $|G_m|$, diffraction replicas can conspire to blur the photoelectron spectrum. Ultimately, diffraction replicas are difficult to separate from minibands—which are anticipated to occur—because the miniband energy scale lies below the energy resolution for the current measurements. There is a simple argument to be made to rule out diffraction at Γ, however. Because diffraction will be elastic, diffraction peaks cannot create an increase in the linewidth for a band with very high effective mass, as at Γ, because the band curvature and small diffraction replica spacing in momentum do not permit shifting intensity in energy. The band effective mass is lower bounded by 2 at Γ, while we can use $2\ \mathrm{nm}^{-1}$ as an overestimate for $|G_m|$, to approximate the energy smearing from diffraction as

$$\delta E_{\mathrm{diffraction}} \approx \frac{\hbar^2 |G_m|^2}{2m^*} = 38\ \mathrm{meV}\cdot\mathrm{nm}^2 \frac{|G_m \cdot \mathrm{nm}|^2}{\frac{m^*}{m_0}} < 76\ \mathrm{meV}. \qquad (4.10)$$

Using more realistic numbers only makes smearing due to diffraction less quantitatively appealing.

Additionally, we note that we observe the same phenomenological characteristics in several aligned samples prepared with slightly different characteristics at two different ARPES facilities—at beamline 7.0.0.2 of ALS's MAESTRO and at beamline ANTARES at SOLEIL. The consistency across samples, together with evidence presented showing discontinuous change in the apparent linewidth on the same sample moving from monolayer to heterostructure regions firmly establishes that in long wavelength superlattices of WS_2 and WSe_2, the underlying spectral function changes according to the description in the prior section.

4.9 Flattened Electronic States in Long Wavelength WS_2/WSe_2 Moirés

In this section, we discuss the **(flattening)** of the electronic states which we noted in the primitive Brillouin zone high symmetry ARPES cut of the long wavelength superlattices. In Figure 4.9a-d (momentum directions indicated by Figure 4.9j), we take a closer look at the high symmetry directions in the two heterostructure samples by performing curve fitting and constant momentum peak analysis of the states at higher binding energy. In Figure 4.9c-d, we show constant momentum curves along the momenta labeled by vertical tick marks on the x-axis of Figure 4.9a-b. The flattening of the states in the aligned heterostructure sample is even clearer from this perspective. For the $\lambda = 2.8$ nm superlattice sample in Figure 4.9d, two sharp, well-separated and dispersing peaks are tracked with colored circles, corresponding to the modified valence bands from the WSe_2 (gold) and WS_2 layers (pink). The peak locations for each of the WS_2 and WSe_2 bands clearly follow the monolayer valence band dispersion down to the valence band minimum at EDC 1.

By contrast, we observe three sequences of peaks for the $\lambda = 7.0$ nm sample in Figure 4.9a and c. All three peak sequences appear to be highly localized—non- or nearly non-dispersive—bands. The two lowest energy peak sequences are the same features we recognized in Figure 4.7d. Surprisingly, it is very clear from the EDC analysis of the aligned heterostructure data that the three lowest energy bands never overlap in energy. The highest binding energy peak sequence would seem to be the "tail" of the WS_2 originating band as it approaches the band bottom (in the vicinity of EDC 1). However, the highest energy peak sequence remains energetically separated by around 450 meV across the entire momentum range over which it is visible. What creates this gap? We cannot be definitive, but the miniband structure should create replicas of bands throughout the Brillouin zone. The presence of these replicas shifts weight away from the main band: when the band is dispersive, as it is transiting from Γ to the bottom of the band, this shift appears to suppress the intensity of the band. This does not appear to happen at Γ because the large intrinsic effective mass places the miniband spectral weight more or less on top of the original band.

Having seen that the Γ point states are flattened markedly in the 7.0 nm heterostructure, we turn to considering the K point, which is of the greatest experimental interest because it is anticipated to be the lowest energy region in the valence band. Because of the monolayer stacking order which suppresses intensity from WSe_2, we cannot observe the lowest energy states at K. Nonetheless, a consistent phenomenology can be seen. In Figure 4.9e and f, similar spectra to Figure 4.9a and b are shown cutting at a perpendicular direction across the K point. We also find that the lowest energy bands are flattened for each of the spin-orbit split branches at the K point, but only for the 7.0 nm sample. Although the flattening recalls Γ, the spectra are much weaker and more diffuse at K. Partially, this is attributable to matrix elements for tungsten which already make accurate K point measurements challenging. The formation of minibands from parent bands with larger effective mass ($> 0.4m_0$) only makes the situation worse, as we discussed in the context of an apparent "gap" in the dispersion towards the band bottom leaving Γ. The much weaker intensity means it is only possible to extract the dispersion locally for the more intense heterostructure band which lies at higher energy. For each of the 2.8 nm and 7.0 nm samples, a

dispersion recovered by curve fitting EDCs is plotted as red circles over Figure 4.9e and f. We compare the results of both curve fitting analyses to a monolayer tight binding model, after shifting energy for mutual agreement at K, in Figure 4.9g. Continuing the trend of agreement between completely decoupled monolayer spectra and the results in short wavelength samples, there is an excellent correspondence between the $\lambda = 2.8$ nm sample and the ML TB calculation.

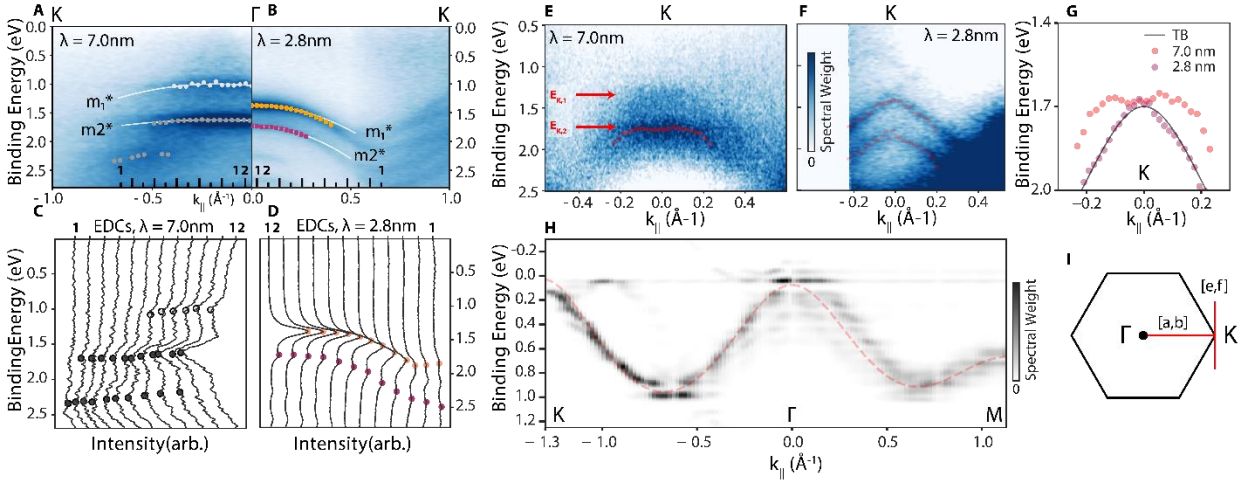


Figure 4.9 **Superlattice Band Structure in Long Wavelength WS_2/WSe_2 [a,b]** ΓK dispersions for the $\lambda = 7.0nm$ [a] and $\lambda = 2.8nm$ sample [b] The location of peaks from EDC curve fitting are shown as overlaid dots. A quadratic—effective mass—fit for the peak centers is shown with a white line and associated label for each band. Quadratic mass fits were conducted in a $2\ nm^{-1}$ range in either direction of Γ. **[c,d]** Energy distribution curves taken along the match vertical notches on the x-axis of [a,b]. The $\lambda = 2.8nm$ sample EDCs [d] show sharp dispersing peaks, while the EDCs taken from the $\lambda = 7.0nm$ sample [c] are non-dispersive to the band bottom, near EDC 1. In [c,d] peak locations were manually assigned to the intensity maxima based on peak shape over a larger range than the curve fitting in [a,b]. **[e,f]** Spectra taken at K perpendicular to the ΓK direction for the $7.0nm$ [e] and $2.8nm$ [f] samples. As in the ΓK direction data, the $7.0nm$ sample has flattened low energy bands at K. The peak positions from EDC fitting for the higher energy SOC split band are shown as red circles for the 7.0 nm sample. For the 2.8 nm sample, overlaid purple circles trace the dispersion of both SOC split bands. For clarify, the image plots of [e,f] are shown with an affine background subtracted. **[g]** A comparison of the energies from curve fitting results to a WS_2 TB model. All data have been shifted to agree at K for mutual comparison. Red and purple scatter sequences are from [e,f]. From the dispersion in the $\lambda = 7.0nm$ sample, we estimate the width of K point as $\sim 3nm^{-1}$.**[h]** The spectral function for a three-band TB model of WS_2 with a superlattice periodic on-site energy after unfolding to the Brillouin zone of WS_2. The essential characteristic, a flattening at the valence band edge, can be reproduced qualitatively even with this simple model. The dashed red curve traces the bare valence band dispersion from the original TB model. **[i]** A Brillouin zone for WS_2 showing the cut directions for the data in [a-b,e-f].

The results for the 7.0 nm are simply not comparable, on the other hand. If the dispersion is not entirely flat at K for the 7.0 nm sample, it would seem to be because of an electron-like concavity, rather than the expected, hole-like parabolic dispersion. In fact, if we look closely at the EDCs near the K point, there is a shoulder below the primary peak sequence which appears very weakly, Figure 4.10. The spacing of 100 to 150 meV is large, but reasonably consistent with a miniband origin for the shoulder intensity. Using the mBZ width and effective mass, the anticipated energy spacing $\frac{\hbar^2 k^2}{2m^*} \approx 95$ meV for the minibands is close to the observed shoulder spacing and the intralayer moiré exciton splitting. This interpretation provides the cleanest explanation for the observed concavity if we assume it is physical. Higher resolution studies

would be invaluable to better determine the low energy electronic ordering of the K point states and to directly observe the miniband phenomenology.

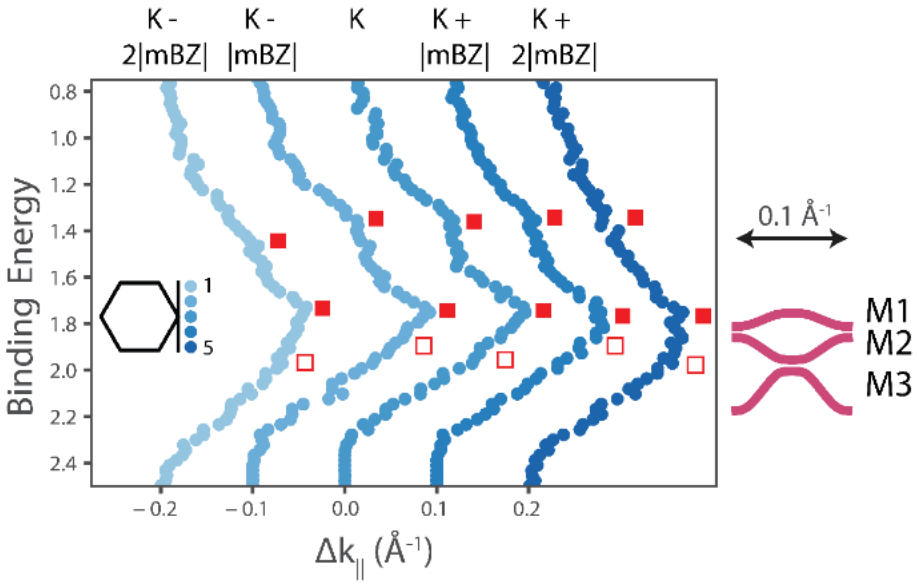


Figure 4.10 **Possible Minibands at the K Point in a 7.0 *nm* Sample** By examining closely the EDCs near the K point, we find a clear shoulder (hollow peak sequence) neighboring the primary peak sequences (filled peak sequences). These shoulders disperse slightly and may arise from miniband van Hove singularities in the mBZ. The location in momenta for the EDCs is shown with a reference Brillouin zone (inset). The schematic on the right shows coarsely how miniband folding can create shoulders for the K point states in superlattices.

Theoretical models are ideal tools to conceptualize physical phenomena and to limit their underlying causes. Even in situations where constructing full physical models is either too technically challenging, or it is too expensive in time or compute, simplified models can distill phenomena to understandability. To examine the underlying mechanisms that might be at play in the formation of these flattened states at the band edges in long wavelength superlattices, we constructed a supercell unfolded spectral function from a three-band TB model of WS_2 following the construction of Liu et al (*119*).Whereas low energy effective models are excellent surrogates for understanding experiments which are restricted to the low energy physics of a system—like transport and some optical measurements—making sense of the broad changes we observe in the ARPES measurements of WS_2/WSe_2 requires calculating, at least coarsely, what happens to the spectral function.

To accomplish this, we constructed superlattices of varying sizes from the primitive-cell TB model with the aid of *pybinding*, a software for tight-binding calculations (*129*). After direct diagonalization in the supercell, the eigenfunctions of the superlattice Hamiltonian with perturbation were unfolded onto the primitive cell Brillouin zone according to the prescription of Nishi, Matsushita, and Oshiyama (*130*). A brief discussion of the details follows before we consider what interpretations we can pull from such a simple model.

4.9.1 Obtaining Supercell Spectral Functions by Unfolding

To unfold supercell eigenfunctions to the primitive cell, we rewrite the spectral function for the primitive cell in terms of the overlap matrix elements between band index (I, i) labeled superlattice wavefunctions Ψ_I, the primitive cell wavefunctions ψ_i and the momentum projection function $R(\boldsymbol{k}) \to \boldsymbol{k}_{SC}$ which reduces the momentum by G_m so that it lies in the mini Brillouin zone. Concretely, Nishi, Matsushita, and Oshiyama shows that the spectral function can be written as

$$A(\boldsymbol{k}, \epsilon) = \sum_{R(\boldsymbol{k}),I,i} |\langle \psi_i(\boldsymbol{k}) | \Psi_I(R(\boldsymbol{k}))\rangle|^2 \delta\big(\epsilon - \epsilon_{R(\boldsymbol{k})I}\big). \tag{4.11}$$

To make progress, we can expand the matrix elements in the sum above by Fourier transforming into the light-binding basis (*130*). After this Fourier transformation, the spectral function depends only on quantities which we can evaluate by diagonalizing the supercell TB Hamiltonian and bookkeeping the spectral weight contributions in the larger primitive zone.

$$A(\boldsymbol{k}, \epsilon) = \sum_{R(\boldsymbol{k}),I,\sigma,s,\boldsymbol{x},\boldsymbol{x}'} L^2 e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} C^{\sigma s \boldsymbol{x}*}_{IR(\boldsymbol{k})} C^{\sigma s \boldsymbol{x}'}_{IR(\boldsymbol{k})} \delta\big(\epsilon - \epsilon_{R(\boldsymbol{k})I}\big). \tag{4.12}$$

In the equation above, adapted from Nishi, Matsushita, and Oshiyama, the tensor contraction of the wavefunctions is carried out between identical site (s) and orbital (σ) labels, over all lattice points $(\boldsymbol{x}, \boldsymbol{x}')$ of the primitive lattice in the first unit cell of the superlattice. L^2 is a normalization factor giving the relative volume of the supercell to the primitive cell so that spectral weight is the same as in the primitive cell when $V_m \to 0$ and $L > 1$. Our calculations were performed on superlattices with $L = 8,12,16$ corresponding to 192, 432, and 768 orbitals respectively. The final supercell size, $L = 16$, is an appropriate model for moiré WS_2/WSe_2.

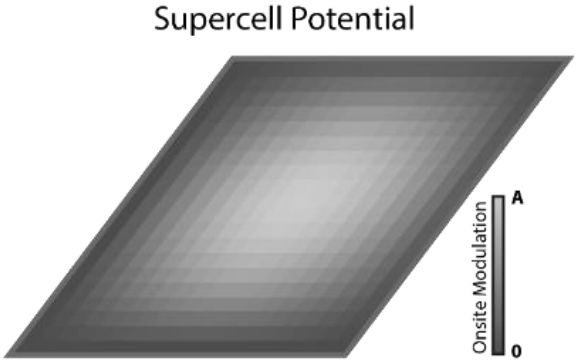


Figure 4.11 **The moiré potential used for supercell unfolded TB calculations.** The moiré potential is taken to be commensurate periodic. Rather than modeling the heterostructure interaction which would provide realistic moiré effects, for this toy model we introduce a varying onsite V_m with maximum perturbation **A** introduced to the tungsten *d*-bands at the Γ point.

To avoid too much complexity in the model while still capturing localization, we eschewed a realistic interlayer interaction between WS_2 and WSe_2 in favor a superlattice periodic on-site

perturbation due to the presence of the WSe_2 layer, shown in Figure 4.11. For commensurate supercells labeled by site indices $0 \leq (m,n) \leq L$, the algebraic form of the potential is

$$V_m(m,n) = A \sin\left(\frac{\pi m}{L}\right) \sin\left(\frac{\pi n}{L}\right). \quad (4.13)$$

We have already shown the shape of the potential used, but where is it introduced? Deciding how to modify the supercell TB Hamiltonian requires making some physical assumptions, but the end results are similar. For the current discussion, we added a spatially periodic on-site potential to the d_{z^2} orbital, while leaving the terms involving the remaining orbitals and hoppings unmodified. Although this is not realistic in the face of strain which will also modulate the hopping terms, it represents a first order approximation meant to capture the flattening at the band edge. Later, we will also use it to examine the interplay between commensuration and the ARPES observables, and even to extract information about the spatial structure of the supercell which lie below the resolution of our ARPES measurements. Having decided how to modify the Hamiltonian, the exact functional form of the superlattice potential does not change the qualitative results. However, because the orbital character changes as the band disperses, the valence band edge at K remains delocalized unless a modulation is added also to the in-plane p and d-orbitals, or a similar superlattice periodic perturbation is made to the hopping terms, which would be appropriate to model strain in the supercell.

4.9.2 Applying Supercell Unfolding to Understand Moiré WS_2/WSe_2

This model permits an examination of which of the observed spectral features can be assigned to structural effects, which are modeled effectively without recourse to the partner layer in the heterostructure through modifications to the on-site and hopping terms. The bands were unfolded to produce the spectral function on the first Brillouin zone of WS_2 and are shown along the high symmetry directions in Figure 4.9h. Upon the introduction of only a moderate on-site perturbation averaging 100 meV across the cell—or with a comparable perturbation to the hopping terms—our model reproduces essentially flat bands at the top and bottom of the valence band in agreement with our ARPES results and over a comparably large momentum range. The on-site perturbation of 100 meV was chosen conservatively smaller than the linewidth broadening to demonstrate that localization can occur even with a moderate perturbation. As we have already discussed, the moiré potential depth is likely substantially deeper than this. The degree of flattening at each of the Γ and K points can be independently tuned by the degree of modulation of the on-site energy for the relevant orbitals. While this model captures the flattening of the valence bands throughout the Brillouin zone, it does not and cannot reproduce other large-scale changes to the spectral function including the increased band separation at Γ.

This construction for the flattening also leaves open the question of the underlying cause. There is strong evidence from STS (*131*) and from HAADF-STEM (*132*) that the relaxed supercell is highly strained so that it can begin to minimize its energy in the supercell corrugations. In this toy model, there's room for localization at the Γ point to arise out of strain, via modification of the hopping terms, or out of an on-site potential which can be used to inhomogeneous hybridization. In the next section, we set to demonstrate that inhomogeneous hybridization

accompanies the strain and lattice reconstruction which has already been observed, by considering the "bonding-antibonding" splitting we observed at Γ.

4.10 Incommensurate Superlattice Hybridization at the Γ Point of Moirés

We identified three core features of the superlattice ΓK spectra in Figure 4.7d: increased linewidth, flattening, and a peak shift at Γ. We have now seen explanations for two of these features. The increased linewidth comes from supercell averaging over the moiré potential, which is large enough to produce the observed linewidth broadening. Flattening occurs due to some combination of inhomogeneous strain or inhomogeneous hybridization. Recall that when we discussed the WS_2 multilayer ARPES data presented at the very beginning of this chapter, the introduction of more layers of MX_2 created a ladder of bonding-antibonding states (*123*) which arise due to interlayer hybridization of the Γ point bands. In tungsten based TMDCs, these Γ point bands have local, out-of-plane $d_{3z^2-r^2}$ and p_z orbital character (*122*, *123*), which means that they more strongly interlayer hybridize than the in-plane orbitals which make up the rest of the valence band (in the conduction band, it is the K point bands which are composed of out-of-plane, p_z orbitals).

The increase in the energy separation between the two lowest energy states at Γ, $\Delta E_{\Gamma,1}$, is too large to be a miniband separation. The increase of around 300 meV and total separation of around 800 meV is much closer to the energy scale for the interlayer hybridized Γ point sub-bands in either WS_2 or WSe_2. Additionally, the relative intensity of the bands is much more closely matched in the $\lambda = 7.0$ nm sample. This strongly suggests, leaving aside the question of ARPES matrix elements and slightly different experimental geometries which can complicate the comparison, that although there is a preference for receiving electrons at the detector for the higher binding energy "WS_2 originating" band which arises due to the inelastic mean free path through the top WS_2 layer, this preference is not as strong as would be expected if all the electrons from the lower binding energy band actually came from the WSe_2 layer. This is strong evidence that much of the photoemission signal from the lowest energy band actually comes from the WS_2 layer, which can occur only in the case where there is substantial hybridization between the WS_2 and WSe_2 d-bands. This should not surprise us either, because the energy difference between different low energy stacking orders is largely dominated by hybridization between layers. Because this hybridization depends on the local stacking registry, it is reasonable to presume a role for inhomogeneous hybridization in flattening the Γ point bands even as inhomogeneous strain must flatten the band edge at the K point.

4.11 A Phenomenology for MX_2/M'X'_2 Moirés

Our ARPES measurements have allowed us to make some deep insights about the electronic modifications which accompany long wavelength WS_2/WSe_2 superlattices. We have seen that the electronic bands are localized not only at the K point, as was anticipated: they are also localized at Γ. ARPES ability to measure the photoemission linewidth agrees with STS measurements showing deep moiré potentials, perhaps exceeding 300 meV in our samples. Finally, we can provide a direct measurement indicating a strong role for incommensurate hybridization of the superlattice d-bands. How transferable are these results and concepts?

4.11.1 Constructing a Bridge to the Optical Literature

Because optical spectroscopists have been studying TMDCs, excitons, and heterostructures for a decade, optical probes were first on the scene in studying TMDC moirés (*133–135*). Although moiré excitons provided the first experimental observables of superlattice physics, a veritable zoo of phenomena has followed (*136*, *137*). This rich existing literature provides an invaluable opportunity to bridge our electronic measurements to optical ones. It would not be surprising if the optics and electronics were part of a consistent phenomenology, but it is valuable to demonstrate this experimentally—superlattice physics has proved to be full of surprises.

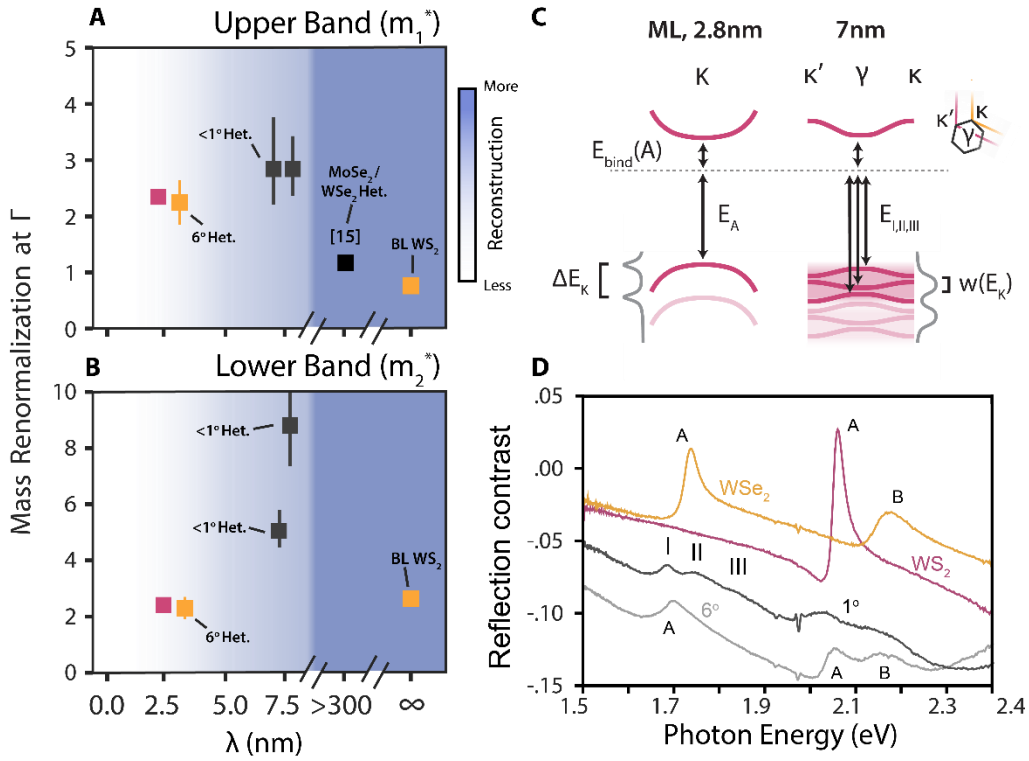


Figure 4.12 **The Evolution of Electron Localization and Excitons in Heterostructures [a,b]** The effective mass across a family of bilayers arranged by the superlattice constant. Native bilayers are taken to be a $\theta \to 0, \lambda \to \infty$ limit while we include monolayers at $\lambda = 0$ for comparison. Moderate values of λ show strong interaction, as indicated by a renormalization of the band effective mass or localization which differentiate this regime from both extremal limits, $\lambda \ll 1\ nm$ and $\lambda \gg 1\ nm$. A sufficiently slowly modulated local alignment is required to incentivize the superlattice to adopt low energy stackings and create the moiré phenomena. Too large, and commensurate domains dominate by area and the measured mass is again small. **[c]** In a moiré superlattice, the miniband formation permits for additional peaks in the exciton spectrum from miniband transitions, albeit at lower oscillator strength. In a plain excitonic material, the exciton binding energy, optical gap, and electronic gap are related to one another through energy conservation alone. In the case of finite resolution ARPES from a superlattice, minibands manifest in a broadening of the linewidth $w(E_k)$ which should match the exciton peak splitting. **[d]** Reflection contrast spectroscopy for each sample. For samples more closely aligned than 1° (dark grey curve), we observe moiré exciton splitting of the intralayer peak which is absent at larger angles (6°, light grey curve) in agreement with established literature. The energy separation in the moiré exciton peaks is 150 meV—which makes it comparable to $w(E_k)$, see *Figure 4.9*e.

Establishing this connection was an explicit part of our study design, which we accomplished by performing reflectance contrast spectroscopy and ARPES on the same samples. Although the sample requirements of ARPES—principally, the conducting substrate and exposed surface—are nonideal from the perspective of the local field for reflectance contrast measurements, the moiré exciton signature is very clear and unambiguous.

What are our expectations? Flattening and an increase in the linewidth, Figure 4.7a and Figure 4.9a,c,e, of bands in the $\lambda = 7.0$ nm sample is consistent with the emergence of superlattice minibands at the K-point band edge. Neglecting complications introduced by superlattice selection rules (*138*) and variations in the exciton binding energy which should be smaller than variations in the local electronic gap (*135, 139*), miniband formation provides a consistency between optical and ARPES results. Electronic miniband formation, or equivalently, superlattice periodic variation in the electronic gap, drive a superlattice varying optical bandgap which results in a splitting of the exciton binding energies into sequences of peaks—moiré excitons (*133–135*). This general picture is summarized schematically in Figure 4.12c for each of the $\lambda = 2.8$ nm and $\lambda = 7.0$ nm cases.

Indeed, by performing absorption contrast spectroscopy measurements on the same samples we used for ARPES—Figure 4.12d—we observed that the splitting of the exciton lines, attributed to the moiré, exactly accompanies the valence band flattening we observe in ARPES in each of our aligned samples, linking these phenomena as part of a consistent complex of superlattice phenomenology. Because of their very small size, direct observation of these K point minibands by current ARPES experiments is difficult and may remain so for the near future. One observable consequence, however, is that the unfolded minibands will increase the linewidth of the band at the K point and as shoulders in the EDC at K. This effect is annotated in Fig. 4C as an increase in $w(E)$. Our results directly demonstrate this widening and show a series of peak shoulders consistent with miniband formation at K as we discussed when considering band flattening at the K point. Because ARPES on undoped samples cannot resolve the single particle gap, we must concede at this juncture that the linewidth increase at the K point in ARPES only provides half of the picture: the conduction band minimum is unmeasured but changes there contribute to changes in the exciton binding energy. Nonetheless, the order of magnitude correspondence between these phenomena is consistent with an origin for the moiré exciton peak structure which comes from supercell periodic modulations in the single particle gap. More importantly, phenomenological agreement between ARPES and reflectance contrast spectroscopy firmly places our ARPES results in the broader context of TMDC moiré physics as we had hoped.

4.11.2 Extending Results across λ and across Material Families

In the previous section, we saw that our results are transferable across experimental probes, by linking changes in the electronics with the formation of moiré excitons. We might hope that our results are transferable also to different material families in the broader ecosystem of $MX_2/M'X'_2$ superlattices, and that we can predict and describe what ARPES might observe about the electronic structure in these related families. This program requires a quantitative way of comparing effects across samples and across material families. The simplest way to

accommodate this requirement uses the local band curvature effective mass, m^*, which is defined through a quadratic approximation to the band at a fixed momentum $\boldsymbol{k}$

$$E(\boldsymbol{k}) = \frac{\hbar^2 k^2}{2m^*} \tag{4.14}$$

as a proxy for the localization of electrons in the valence band and as a measure of the impact or strength of superlattice induced effects.

In Figure 4.12a and b, we plot the effective masses as a function of the superlattice wavelength for both the "upper" (lower binding energy) and "lower" (higher binding energy) bands at Γ. In each case, the effective masses m^* are directly extracted from the dispersion by curve-fitting for the peak locations $E(\boldsymbol{k})$ at fixed $\boldsymbol{k}$ and fitting the resultant peaks for the local band curvature using a parabolic band approximation. We already have seen that the upper and lower bands have strongly renormalized mass in $\lambda = 7.0$ nm WS_2/WSe_2 superlattices. The valence band or upper band (corresponding to m_1^* in Figure 4.9a and b) is more weakly renormalized than the next lowest energy band. In the latter case, masses are lower bounded by $\approx 5m_0$ and $\approx 8m_0$ in two different $\lambda = 7.0$ nm samples for which we have ARPES data, while for the valence band the value is at least $3m_0$. In the case of $\lambda = 7.0$ nm samples, the masses presented place practical lower but not upper bounds on the degree of valence band localization because a large enough momentum range (> 0.25 Å^{-1}) to observe band dispersion is required in order to form a mass estimate—momentum ranges smaller than ≈ 0.2 Å^{-1}, which are frequently used in assessing the effective mass of TMDCs in ARPES, are consistent with completely flat valence band edges.

The renormalization for $\lambda = 7.0$ nm samples may also include contributions due to the interlayer hybridization at the Γ point. In multilayers, we already have seen that the lowest energy band has a much smaller mass which saturates to a bulk effective mass close to the bare mass in WS_2 and WSe_2. The masses—to the extent which ARPES can measure them—are modified in competition between superlattice induced localization and hybridization induced tendency toward lower mass. The opposite situation occurs in the antibonding partner band with the superlattice and hybridization conspiring both to increase the effective mass.

Now we turn to considering other superlattice wavelengths and other TMDC superlattices. Although it is not possible to construct superlattices with $\lambda > 8$ nm in WS_2/WSe_2 due to the 4.4% lattice incommensuration (Figure 4.1), it is useful to treat $MX_2/M'X'_2$ superlattices in a uniform way by examining the relationship between λ across chemical families and in using the effective mass as a correlate of localization. Enough examples are available in the literature on other heterostructure systems (notably $MoSe_2/WSe_2$ (*107*)) that WS_2/WSe_2, which represents a middle ground for the superlattice wavelength, can be contextualized in terms of peers with both shorter and longer wavelengths. For this purpose, we also adopt native homobilayers as analogues to twisted homobilayers as $\lambda \to \infty, \theta \to 0$, and monolayers as analogues to the case where $\lambda \to 0, V_m \to 0$. The results are plotted together with our experimental values in Figure 4.12a,b.

Both extremes have much smaller effective mass, but the most interesting points to consider are our own $\lambda = 2.8$ nm heterostructure and the much longer wavelength ($\lambda \approx 100$ nm) aligned $MoSe_2/WSe_2$ sample from Wilson et al (*107*). These two results flank the aligned WS_2/WSe_2 superlattices and each looks much more like the very long wavelength and very short wavelength limits than our $\lambda = 7.0$ nm samples.

To really understand this, we need to consider what the ARPES process measures, given that the ARPES beam is much larger than the supercell in all cases considered, together with the spatial distribution of electronic behaviors across the supercell. In the limit of very large λ, the superlattice is allowed to reconstruct into commensurate or nearly commensurate domains of constant local stacking separated by narrow domain walls (*94*, *132*). This is the recapitulation of the discussion we used to lead into this chapter by considering what might furnish a moiré potential in the case of twisted lattice incommensurate TMDCs, but here it comes to fruition in terms of experimental results across results in STEM, ARPES, STS, and optics.

In Figure 4.13, we show the emergence of reconstruction and inhomogeneous hybridization causes and increase and subsequent decrease in the effective mass measured by ARPES. To keep things simple, we imagine a heterostructure system consisting of two layers MX_2 (top) and $M'X'_2$ (bottom), and we will take the perspective of the local environment seen by the MX_2 layer. We will track the distribution of different local electronic environments (local registry or stacking order) in the supercell as λ is changed. The local environment changes rapidly for $\lambda \ll 1$, but as we discussed at the beginning of the chapter this has to force the MX_2 layer to sit further away from its partner $M'X'_2$. In this case, the local environment is approximately the vacuum environment seen by an isolated monolayer of MX_2. This explains why we place a monolayer of WS_2 as the first exemplar along this curve and why we list the areal fraction of the "A" symmetry at 1 on the left-hand side.

As we increase λ, it becomes energetically favorable to introduce some strain so that a portion of the lattice can adopt one or more low energy configurations "B" at patches in the supercell. This requires moving X and M atoms from their monolayer equilibrium positions. This is the moiré regime occupied by WS_2/WSe_2. In this portion of the diagram, many or most of the atoms are in domain walls between the low energy configurations and sit at a continuum of local environments interpolating between the low energy configurations. ARPES emits from these with as much preference as from the lowest moiré minima, which smears the bands even as flat minibands emerge. If we further increase λ, the low energy patches are allowed to expand. Eventually, most of the atoms are in one of a discrete set of commensurately reconstructed low energy configurations (right hand side of the plot). As discussed in Weston et al. (*132*), the areal fraction adopted in the commensurate limit depends on the relative energy of the different commensurate orders and of the domain walls where the registry much switch to another order. Because ARPES measures only the spatial average over the supercell—it is "momentum direct" as opposed to being "spatially direct" like scanning probe microscopies—ARPES measures an average of all the commensurate structures in the supercell, weighted by their areal fraction in the supercell. This explains why both the $\lambda \to 0$, and $\lambda \to \infty$ limits produce sharp, narrow

linewidth bands with effective masses reasonable for semiconducting TMDCs and only the intermediate, moiré, regime produces nondispersive bands and flattened electronic states.

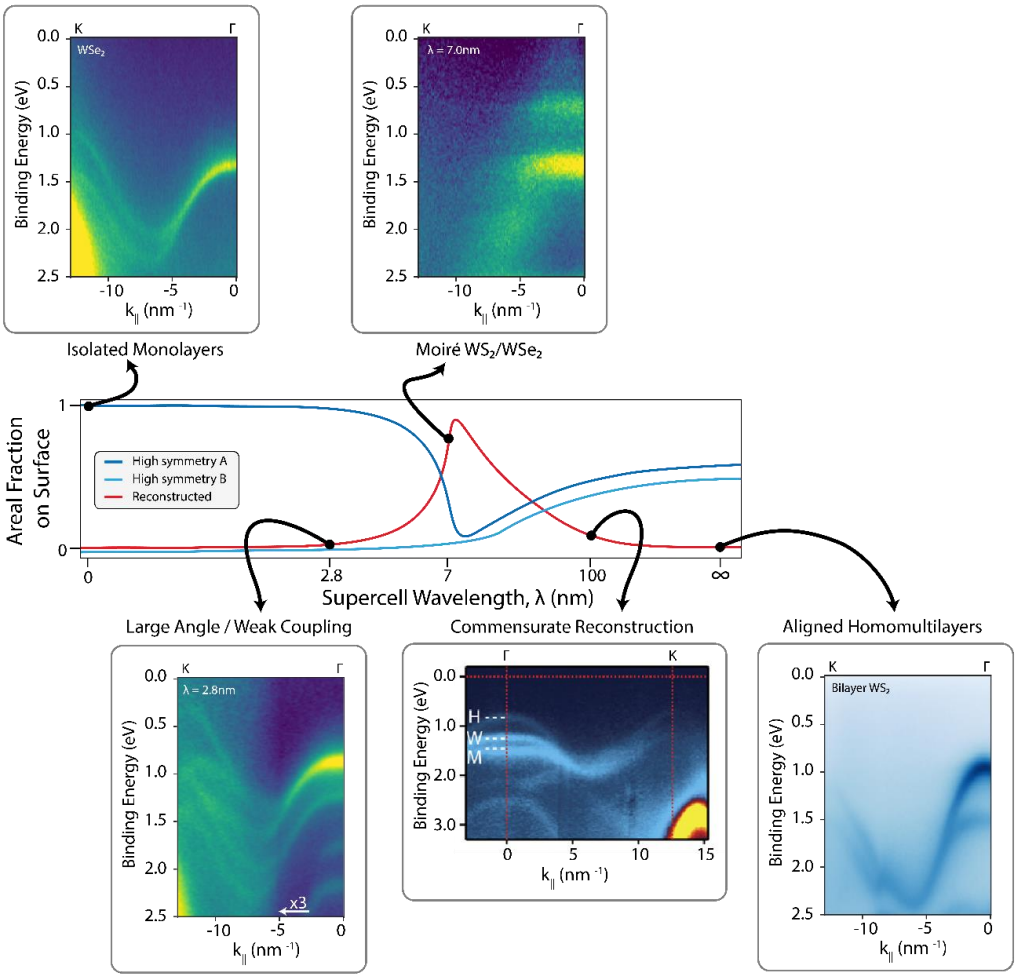


Figure 4.13 **How Sample Physics and ARPES Observables Change as a Function of Superlattice Wavelength.** The central line plots track the fraction of the sample surface which is composed of high symmetry sample regions or strained and moiré reconstructed region of sample. We take the perspective of one of the two layers in the heterostructure, so that we can consider a monolayer as the left limit of one of the high symmetry orders. On the far left, the layers are decoupled and the system has the symmetry of an isolated monolayer. As the superlattice wavelength increases, it becomes energetically favorable to distort the layer under consideration so that some of the supercell can sit at lower energy in the corrugations defined by the other layer in the superlattice. When λ is moderate, only a small region gets to adopt a high-symmetry configuration where it is in registry with its partner layer. Accommodating this small region requires introducing some superlattice periodic strain throughout the supercell. If we further increase λ, the low energy patch is allowed to grow according to λ^2, the area occupied by the patch. The energetic price to do this is to transition from one coordination to another at the border of the high symmetry patch, which scales like λ. This crudely explains why superlattice phenomena turn on suddenly when the $\propto \lambda^2$ term outcompetes the $\propto \lambda$ energy barrier. This description is a simplification because there are several stacking coordinations in the supercell and we consider only two. Nevertheless, it provides a simple framework in which to consider how the band structure, the real space picture, and the ARPES observables change throughout these limits. Because ARPES measures areal averages, it is sensitive to the distribution of the A, B, and transitional regions. In either $\lambda \to 0$ or the $\lambda \to \infty$ limits, the superlattice has high symmetry across the entirety of the supercell, which means that ARPES measures the weighted average of the discrete, high symmetry configurations. ARPES cuts from model $MX_2/M'X'_2$ systems are arranged by λ and demonstrate this evolution.

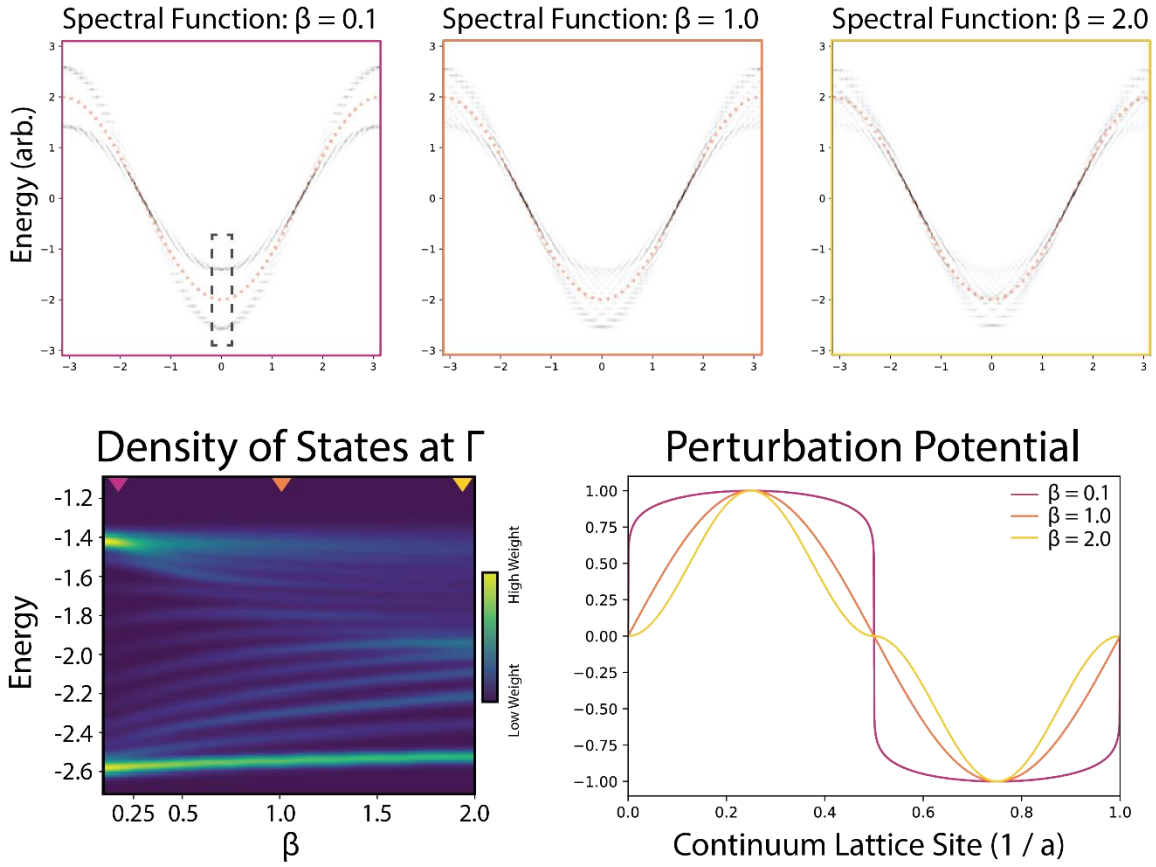


Figure 4.14 **Convergence to Unrenormalized Bands.** The recovery of a discrete set of sharp energy bands can be demonstrated in a 1D monoatomic chain TB model with an onsite moiré potential. The moiré-like regime is characterized by a continuously varying potential, $\beta \approx 1$, while the reconstructed limit is played by the $\beta = 0.1$ curve. Potential shapes are shown on the right hand side line plot, while spectral functions for each of $\beta \in [0.1, 1.0, 2.0]$ are shown in the top sequence. The image plot on the bottom left plots the momentum integrated spectral function in the dashed region—the zone center—of the top left plot. As β is decreased, minibands converge to a pair of energy levels separated by the maximum excursion of V_m. Mid-spectrum minibands lend spectral weight back to the main bands as the domain walls between high and low energy regions in the supercell narrow.

We can watch this process play out by using the same spectral function we used to model the flattening of the valence band edge, applied to a one-dimensional monoatomic chain with the tight binding Hamiltonian

$$H = \sum_{i \in S} t a_{i+1}^{\dagger} a_i + t a_i^{\dagger} a_{i+1} + V_i a_i^{\dagger} a_i \tag{4.15}$$

To model the effects of commensurate reconstruction, we will let $V_m(\boldsymbol{x})$ evolve from smoothly varying in the supercell towards a function which is nearly piecewise constant over the different patches of high symmetry stacking. At the domain walls where the stacking order must change, the value of $V_m(\boldsymbol{x})$ changes rapidly. For the discrete, TB model we replace the continuous moiré

potential $V_m(\boldsymbol{x})$ with the discrete potential V_i evaluated at the lattice positions $V_i \equiv V_m(x_i)$. To model the intermediate regime where the moiré potential is strong but slowly varying over the unit cell and to model the reconstructed regime, we adopt a form for the potential

$$V_i = V_0 \left|\sin\left(\frac{d}{\lambda} 2\pi i\right)\right|^{\beta} \operatorname{sign}\left(\sin\left(\frac{d}{\lambda} 2\pi i\right)\right). \tag{4.16}$$

The continuous limit of this potential is plotted in Figure 4.14. When $\beta = 1$, this potential is a sine wave across the moiré cell. Decreasing β pushes the shape of the moiré potential so that it more closely resembles two separated domains which have relative energies ± 1 with a smooth change between them, where as increasing β does the same but places a distinct zero energy level between them at the nodes of the sign function. The evolution of the system as β is decreased from $\beta = 1$ towards $\beta = 0$ models what happens as we increase the value of λ in moiré systems of 2D materials.

Using the same approach as for the two-dimensional Hamiltonian for moiré modulated WS_2, we can solve this Hamiltonian and unfold the eigenvectors onto the Brillouin zone of the parent system where $V_i = 0$. The results of these calculations are reported in Figure 4.14. When $\beta = 1$, we can see that the energies are essentially modified continuously so that the density of states at the band edge is smeared between t and $t + V_0$. At high binding energy, the density of states is smeared between $-t$ and $-t - V_0$. At momentum values in between, the spectral function is smeared according to the creation of minibands between two sinusoidal envelopes. These envelopes represent wavefunctions which are either in phase with the moiré potential, so that the potential raises the energy of the band, or out of phase with the potential so that the opposite occurs.

Given these general features of the spectral function, we can form a good estimate of the behavior by looking at the momentum resolved density of states at $\boldsymbol{k} = 0$. Varying $\beta \in [0.1,2]$ provides insight into the evolution from the moiré limit to the commensurate reconstruction limit at $\beta \ll 1$. Two striking features emerge from Γ point integral spectral function in Figure 4.14. The first is that several of the minibands are driven towards the lower or upper energy limits $-t - V_0$ and $-t$. The second effect is that minibands which do not correspond to energy extrema at Γ continuous transfer spectral weight towards the extrema energy bands as β decreases. This neatly matches intuition. When β is small, sitting in the trough or on the peak of the potential we would have the impression that V_i is not changing. The band structure is well approximated in these regions by replacing the moiré potential with a constant approximation and dispersive bands characteristic of the monoatomic chain reappear again.

By contextualizing WS_2/WSe_2 among other TMDC heterostructures measured by ARPES, we can see that the $(1 \ll \lambda \,/\, \mathrm{nm} \ll 100)$ regime occupied by aligned WS_2/WSe_2 is unique among the candidate materials which have been measured by ARPES in maximizing electron localization over available candidates. Future experiments should clarify the differences between twisted homobilayers which can be continuously tuned across the moiré regime and attempt to

resolve better the changes which occur to the very lowest energy electrons for both twisted and lattice incommensurate superlattices.

4.12 Accessing Intra-moiré Physics with Large Beams

Through the prior discussion and our experiments, we have seen that nanoARPES provides powerful direct access to the electronic structure in 2D material superlattices. Can it compete with spatially direct probes like STS in offering a description of what is happening inside a single moiré unit cell? Barring a revolution in the technique, ARPES beams will not soon reach below 10 nm. Because of sample aging issues and practical matters of geometry, measuring directly from a patch within a single unit cell may never happen with ARPES. Nevertheless, ARPES can be used to answer questions about the spatial structure of states in the supercell because moiré superlattices occupy a challenging intermediate regime where neither a real-space nor momentum-space picture of the electronics is strictly better.

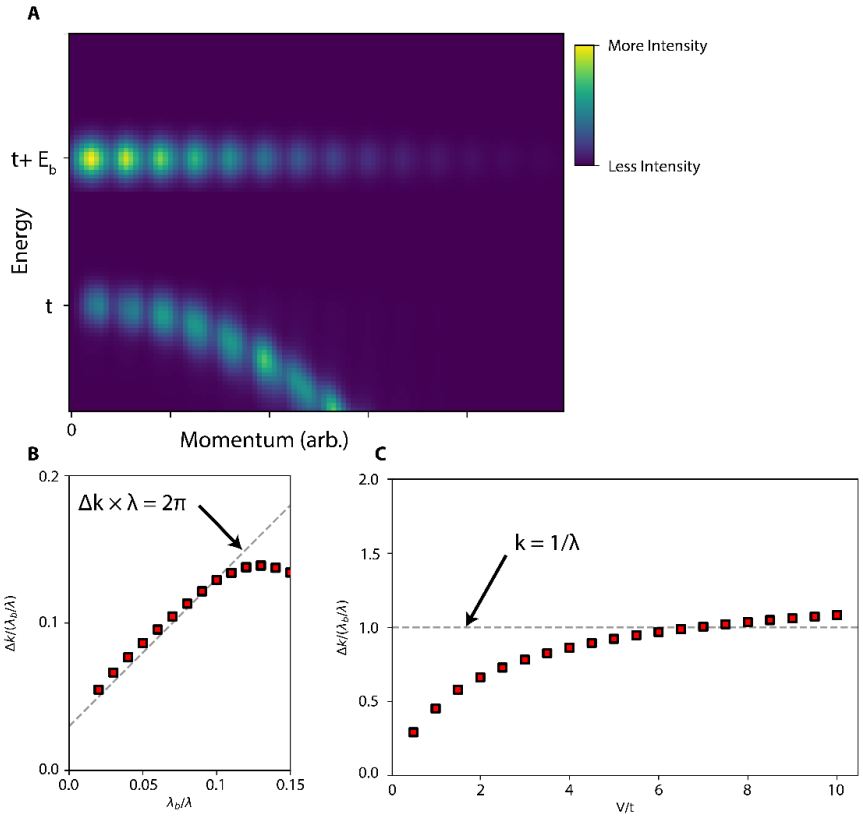


Figure 4.15 **Band Unfolded 1D Chain [A]** An example spectral function obtained from a monoatomic 1D chain with onsite superlattice potential. In the presence of a sufficiently strong superlattice potential V, the valence band edge develops bound states resembling trapped quantum well states (flat dispersion at $t + E_b$). The width in momentum of these states encodes detail about the size of the well. **[B]** Using a square well of width λ_b and depth V/λ_b, we can observe that the width of the well is in excellent correspondence with the momentum broadening seen in the spectral function for the localized state at the band edge. The fitted gray line has been shifted upwards slightly because the peak fitting procedure used slightly overestimates the width of the spectral function peak. **[C]** Fixing λ_b and varying V, we see that the momentum broadening saturates so long as the well is deep enough to support bound states. Because the fitting procedure slightly overestimates the broadening, the width saturates to a value slightly larger than expected from the inverse well width.

When we presented the data from the K point in aligned WS_2/WSe_2 in Figure 4.9e, we noted the flattened top of the K point band edge. At the time, this flattening was just part of a broader qualitative picture about the spectral function measured by ARPES. However, we can use the details about the width of this state to estimate how narrowly confined the electrons are in the moiré potential minimum. The width of the low energy state is between 3 and 4 nm^{-1}. From ARPES on twisted bilayer graphene moirés (*59, 60*), dimensional crossover in nanowires (*140*), and from ARPES on aromatic hydrocarbons (*141*), width photoemission features in localized systems are related inversely to spatial confinement of the system. From the measured width in momentum, we can estimate the localization of electrons in the supercell with LDOS contributing to the K point band edge by calculating via the uncertainty principle $\Delta\lambda\Delta k = 2\pi$ to be 1.6-2.1 nm. Because the supercell λ exceeds 7 nm, our observations indicate that the lowest energy electrons at the band edge are highly localized around a few lowest energy patches in the superlattice. The recovered values for the localization of the band edge states is in extremely good correspondence with the value of around 2 nm found by STS for the WSe_2 valence miniband in WS_2/WSe_2 (*131*).

The application of the uncertainty principle in this argument may appear slightly flippant, but it can be justified by again appealing to the spectral function unfolding calculations we have used as a guide throughout our discussion. It can also be seen as an extraordinarily simplified form of ARPES orbital tomography. To demonstrate that we can interpret the localization of the very lowest energy states—equivalently the shape of the deepest well in the moiré supercell—via the momentum broadening observed at the band edge, we consider an unfolded spectral function for a 1D chain on λ sites. The moiré potential applied is a square well with depth V which is zero outside of a width λ_b. The hopping strength on the chain is t. As the depth of the well increases, it becomes favorable for the wavefunction to localize at the well and a series of bound states form. The number of bound states supported by the well depends on its depth relative to t. Always, we will consider the lowest energy state. This situation is shown in Figure 4.15. The width of the partially confined state is measured by curve fitting the peak in an interval $0.02t$ around $t + E_b$. From Figure 4.15, we observe that the width of the peak is relatively insensitive to the well depth so long as it supports at least one bound state ($V/t \approx 1$ in Figure 4.15c). Furthermore, $\Delta k \lambda_b \approx 2\pi$ so long as $\frac{\lambda_s}{\lambda} \ll 1$ (equivalent to $V/t > 1$ here). This picture roughly follows the intuition we can borrow from the finite square well potential in the one-dimensional Schrödinger equation treatment. From STS in the literature (*127, 131*) we know that the moiré potential in WS_2/WSe_2 is in this strongly localized limit, with at least two confined states observed at the band edge. This is corroborated by the presence of weak shoulders in the photoemission signal at K which may correspond to minibands near the energy resolution of our experiments.

Some questions about extremely small structures can clearly be answered despite using a spatially averaging probe like ARPES. In fact, what we have done is not so different than what happens in STS when assigning edges in the measured local density of states to band edges at different locations in the Brillouin zone. ARPES is a tool to measure momentum structures while STS is perfectly adapted to mapping the real space charge distribution and density of states.

Because we understand and can model the probe interaction with the supercell variations, each of ARPES and STS can answer questions outside their bailiwick. The Γ point presents a large van Hove singularity and consequently STS can identify that the Γ point states are flattened (*131*). Because the V_m modulates the binding energies across the moiré cell, ARPES can measure the distribution of its values (Figure 4.8b). These two spectroscopies can appeal to energetic variations to explain what is naturally answered through the reciprocal technique. Rather than signifying that one or the other technique is sufficient to create holistic understanding in moirés, this redundancy is necessary because it permits associating results which are not observable in the partner. On the one hand, the dipole matrix element and spin-resolved photoemission may provide insight into orbital and spin-order in superlattices. On the other, STS is already proving invaluable to study the spatial structure of correlated electron states in partial filling of the superlattice by directly imaging Wigner crystallization.

4.13 Multiple Superlattices

Although we have focused on the effects of the TMDC superlattice, clean interfaces also exist between the heterostructure and the substrate. For samples which include a capping layer of graphene, the overlayer provides yet another interface which can enrich the physics. The most direct consequence of the substrate and capping layer interfaces is that they present the opportunity for coexisting superlattices that further tune material behavior. In this section we will very briefly touch on minibands and replicas of the photoemission signal in ARPES, which is made more relevant by the longstanding interest in cataloguing these phenomena.

In a sample with a graphene cap, we observe two coexisting moirés between the WS_2/WSe_2 and between a graphene cap and the upper WS_2 layer. Remarkably, the graphene cap is imprinted with the effects of both superlattices. The longer wavelength TMDC moiré causes a cascading set of replicas of the graphene Dirac point on the Fermi surface, aligned to and consistent with the mBZ spacing for the TMDC moiré (Figure 4.16a,c).

Reciprocal effects of the graphene-TMDC superlattice are apparent on the TMDC superlattice as well. We observe minibands of the superlattice hybridized Γ point with the spacing of the graphene/WS_2 superlattice (Figure 4.16b) and hybridization gaps arising in the graphene Dirac cone as it crosses the nearly flat valence bands of the heterostructure which can be seen in Figure 4.17. A “map” of the Brillouin zone orienting all these superlattice effects is shown in Figure 4.18. Taken together, these imply a richer set of superlattice phenomena than is typically recognized and call into question the degree to which we can consider the WS_2/WSe_2 superlattice a system in isolation of its local environment. Multilayer superlattice engineering is an interesting new frontier indicated by these observations, which may offer a greater degree of control in engineering material properties.

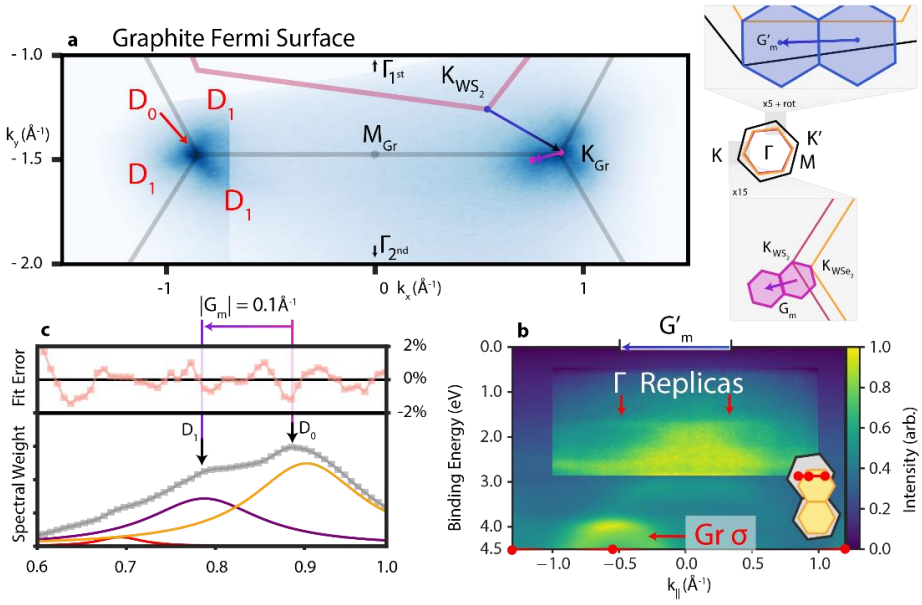


Figure 4.16 **Simultaneous Superlattices in WS_2/WSe_2/Gr: [a]** Replica Dirac cones of the capping graphene layer. A threefold symmetric set of replicas (D_1) is observed around the original Dirac point (D_0). **[b]** Minibands/replicas (indicated by red arrows) of the hybridized WS_2/WSe_2 band. The symmetry assignment to the WS_2/graphene superlattice can be observed by noting that one of the minibands coordinates over the Γ point of the graphene band. An inset (corresponding to Fig. S8) shows the location in momentum for the cut shown. A relative enhancement of the miniband contrast is obtained by performing adaptive histogram equalization of the spectrum intensity in the brighter inset region (inset discontinuous color region). By comparison to the sample geometry in Figure S8, we assign these bands to interaction with the graphene-TMDC superlattice. **[c]** Curve fitting the locations of the graphene replicas shows good correspondence with TMDC miniBZ in this sample.

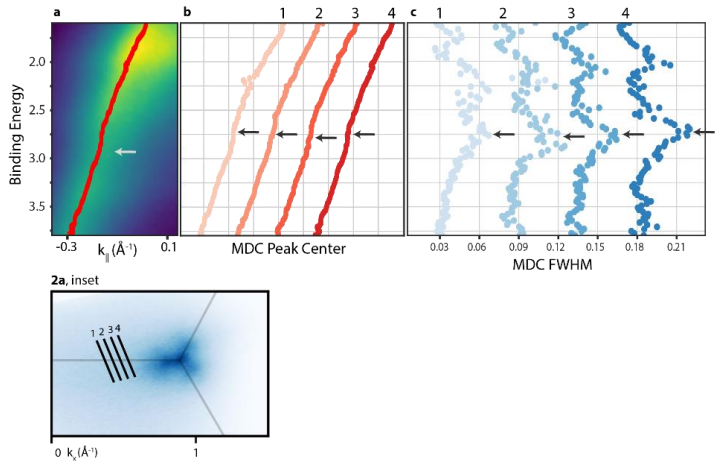


Figure 4.17 **Anticrossings of Graphene and Graphene Replica Bands with Moiré WS_2/WSe_2** [a] An example of the graphene and replica dispersion which shows anti-crossings and dispersion discontinuities at the locations of flattened heterostructure band edge states. **[b]** Fitted dispersions along the locations indicated on the Fermi surface in [2a, inset] which show at least one sequence of consistent breaks in the graphene dispersion marked by black arrows. Dispersions are recovered by constant energy curve (MDC) fitting across the window shown in [a]. **[c]** The constant energy curve (MDC) peak widths, shifted so that the trend as a function of energy can be observed. At the location of the dispersion beak in [b] the linewidth increases substantially corroborating the existence of an anticrossing in the graphene band. **[2a inset]** Reference Fermi surface showing the location of the analysis, away from the Dirac cone along a high symmetry direction in the graphene Brillouin zone.

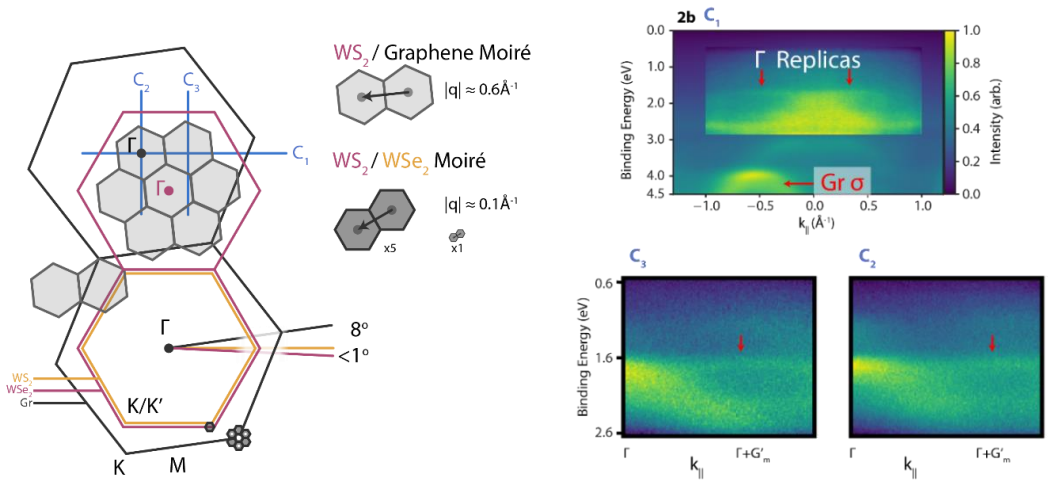


Figure 4.18 **Momentum Structure of the Superlattices and Alternate View of the TMDC Minibands. [left]** Annotated Brillouin zone showing the two different superlattices present in the system: dark grey TMDC moiré, light grey WS_2/graphene moiré. **[right]** Perpendicular cuts through the *Γ* point miniband features to show the correspondence with the primary band energy spacing. Both replicas in C_2 and C_3 have comparable intensity.

4.14 Concluding on Moiré Superlattices

By this point, we have seen a very large collection of phenomena related to the detailed band structures of TMDCs and superlattices of TMDCs. Using combined optical absorption spectroscopy and ARPES, we report the first direct, momentum resolved measurements of the electronic properties of WS_2/WSe_2 superlattices across both the weakly- and strongly modified regimes. The most salient feature we observed were flattening or localization, bonding-antibonding or hybridization splitting between the states at Γ, and moiré potential modulated linewidth increases throughout the band structure.

Significantly, we were able to discuss all of these results in a unified framework, making use of an emerging understanding of how TMDC superlattices behave in the short and long wavelength limits drawing on STEM, STS, optical spectroscopy, and our new ARPES results. Our results are the first momentum-direct measurements of moiré band reconstruction in TMDC heterostructures and demonstrate a consistent electronic and optical phenomenology. Through this discussion we were able to establish links between the observable signatures of different experimental probes and to generalize our results to material instantiations of $MX_2/M'X'_2$ superlattices which have not yet been measured with photoemission. We also saw that in moiré materials which have complex momentum structure and real-space structure, ARPES and scanning probe microscopies like STS and MIM play mutually complementary roles permitting a more holistic understanding of superlattice physics.

As a novel contribution, our results point to the importance of interlayer physics and hybridization in the setting the low energy band structure and as potential key ingredients to drive the emergence of correlated phases in superlattices. Unraveling the independence or interdependence between localization and hybridization effects should be an essential next step toward truly engineering correlations in TMDC moirés.

5 Reinterpreting ARPES Data and Standard Analysis Approaches

We appear to have arrived at the results we presented in the last chapter on TMDCs and superlattices of TMDCs by conventional data analysis approaches. At the core of our analysis on WS_2/WSe_2, we present curve fitting data, ARPES dispersions which have been routinely converted to momentum, qualitative comparisons drawn between peer materials, and analysis of the band curvature. There are very few indications in the proceeding chapter that we will fulfill the original promise we made to discuss how more varied analysis tools and machine learning can be used to advance the interpretation of ARPES data. As far as tools go in photoemission, we've seen hammers and saws, not layout squares and planes.

In truth, there are a few unusual differences which are apparent in how we treated multilayer TMDCs and superlattices. In the discussion on the trigonal warping on WS_2, we performed a global fitting for the trigonal warping dispersion around the K point to obtain a robust estimate for the warping parameter C_w^3. In several places in the discussion on superlattices, for instance in presenting replica bands of the WS_2/WSe_2 Γ point, we used histogram equalization to improve the contrast for very weak bands—with an excess spectral intensity of just a few parts per thousand—on top of strong inhomogeneous backgrounds. These are just the visible marks of having used more unconventional analysis tools. As when a carpenter constructs a table, pegged joints and wood cuts advertise the use of hammers and saws, but this does mean that they were the exclusive tools employed by the carpenter who must plan the table's assembly and ensure a flat and polished finish after construction. The situation is similar in photoemission, and probably in all crafts.

In ARPES, the routinization of complex analyses and the introduction of genuinely new tools excels in the first steps of analysis, where the scientist is trying to grapple with the structure of their data. When tools are applied at this stage, they are not often visible in the final analyses which make their way into published papers, even if their use was essential. The primary tool ARPES scientists use at this stage is their eye—young students will frequently be chastised by PIs to "look at the raw data"—but the dimensionality of ARPES data is rapidly growing and young students risk floundering or even drowning in it, as we described at the thesis' outset. NanoARPES measurements can resolve the spectral function by six angles, outgoing photoelectron kinetic energy, incoming x-ray energy, sample temperature, the two spatial dimensions on the sample surface, the electron filling in the sample, chemical doping, and more. While a typical ARPES experiment used to record data which varied only over two or three angles, or over one angle and temperature, recording over many of these variables sparsely during a measurement is now common. Better tools are required during the exploratory phase of analysis to help augment human eyes and tell them which portions of the raw data deserve direct scrutiny.

In Chapter 2 and Chapter 3, we already saw how these new tools can be considered part of a holistic approach providing utility across the ARPES experiment lifecycle from planning, collection, and analysis. Those two chapters describe necessary infrastructural components to this vision, but do not fulfill on pointing to ways in which machine learning and novel analysis

approaches help with the goal of extracting physics from data. As a general topic, this is far too broad, so in this chapter we will look at a few concrete manifestations which can illustrate the program and serve as a stepping off point for future studies.

5.1 Machine Learning Enabled Exploratory Data Analysis in ARPES

Dimensionality reduction techniques like principal component analysis, factor analysis, and matrix factorization excel at summarizing high dimensional data, like the kind recorded by ARPES. After projection to a lower dimensional space, these decompositions and others like them permit scientists to examine more of their data at the same time. Primarily, these techniques make use of assumptions about the underlying generative structure of data—as in factor analysis—or merely take the statistical standpoint that variations in the data should be retained—because they are explanatory, as in PCA—after projection. This approach is very general, so it is difficult to provide a comprehensive list of places where it can be leveraged in ARPES. Generality, together with poor support for machine learning approaches in traditional analysis ecosystems like Igor Pro, mean that most ARPES scientists do not turn to these tools at any point during exploratory data analysis. Nonetheless, providing support and demonstrating the applicability of tools like this was a design goal for PyARPES and I've leaned on these facilities across many projects.

In this section, we will tackle both issues by example: showing that dimensionality reduction is a powerful tool for understanding high dimensional ARPES data very quickly and showing that PyARPES makes it extremely simple to apply using interactive tools. As a case study and so as not to stray too far from topics at hand, we will use nanoARPES measurements of WS_2 multilayers and superlattices.

The most common high-dimensional dataset produced during a nanoARPES experiment are three- and four-dimensional scanning spectral maps which perform XPS or a valence ARPES cut at every point in a contiguous region across your sample. An example of a dataset like this can be seen in Figure 5.1 where we show how this applies to building a correspondence between an optical micrograph of a sample and the electronics recorded during this initial, surveying phase of nanoARPES experiments. At a variety of locations on the sample, we take a quick, surrogate measurement for the information we hope to record. Usually, this is an XPS spectrum over some levels which tell us about the chemical composition of our sample. The advantage of this approach is that by performing high efficiency surrogate measurements, we learn about the spatial structure of a complex sample quickly, and can decide where to prioritize more comprehensive measurements. Sometimes, but not always, this amounts to merely building a correspondence between the optical micrograph and the photoemission map—as in the correspondence between the left and bottom right portions of Figure 5.1—so that we can use prior knowledge about our sample to situate future measurements. However, there is typically a dose of uncertainty even in samples we believe we understand well. For exfoliated 2D materials and heterostructures, this may come in the form of residual polymer on the sample surface from transfer, tears which develop during UHV annealing, or blisters consisting of trapped atmosphere and impurities between layers of the sample which develop during the transfer process. For these

reasons and others, it remains necessary during real experiments to make decisions driven by surveys of the electronics.

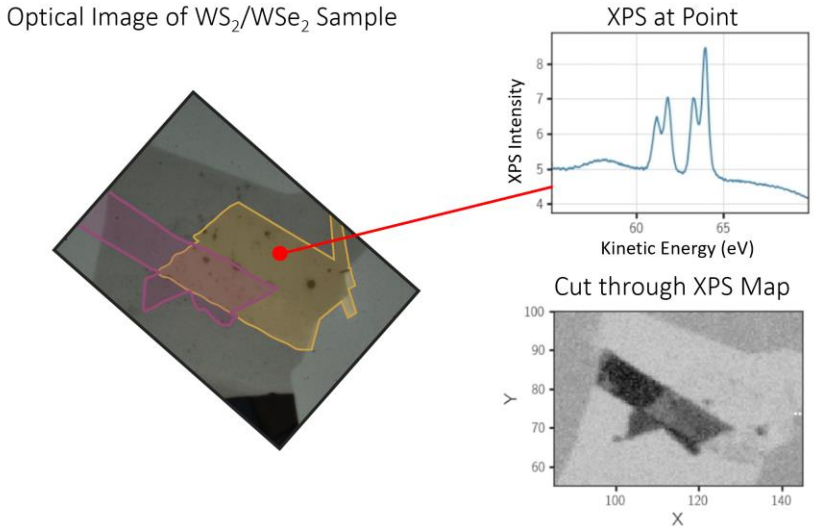


Figure 5.1 **A three-dimensional volumetric dataset from nanoXPS across the surface of a WS_2/WSe_2 moiré sample. [left]** On the left is an optical micrograph of an aligned (moiré) WS_2/WSe_2 superlattice. The purple region shows the location of the WS_2 monolayer while the gold outlined region shows the location of the WSe_2 monolayer. **[top right]** At each point in the optical micrograph, an XPS curve across the W $4f$ core level doublet was taken. As an example, a representative XPS curve is shown at the indicated point on the micrograph. **[bottom right]** By integrating the three-dimensional data volume along some portion of the energy spectrum, for instance between 60 eV and 61 eV, we can preferentially select the photoemission intensity corresponding to different regions on the sample, as shown in the cut through the volumetric dataset in the bottom right.

To borrow terminology from the broader discipline of data science, the process by which we use these surveys to make decisions is a form of exploratory data analysis. We discussed the role of this kind of analysis in iterative experimental decision making in the chapter on AutodiDAQt. In that section we took a more totalizing and convicted approach to the matter, but the discussion was at the level of the philosophy of experiment design. Here, I want to give a concrete example of how ARPES scientists can use the exploratory data analysis tools in PyARPES to drive faster insights into high-dimensional experimental data.

Dimensionality reduction techniques like PCA are incredibly simple from a statistical perspective. Despite this, they are used rarely in photoemission which gives them an arcane flavor. In other fields ranging as far afield as business analytics, the application of dimensionality reduction did not catch on until the availability of visualization tools which increased their accessibility to non-expert users. As a visually oriented experimental discipline, ARPES is in the same position of waiting for visualization tools before adopting more complex approaches to data reduction. This is especially true for data analysis approaches which threaten the direct interpretability of the ARPES spectral intensity.

PyARPES hosts several interactive exploratory data analysis tools for high-dimensional data which solve this visualization problem. Figure 5.2 shows a utility which takes data transformed under an arbitrary dimensionality reduction technique like principal component analysis, factor

analysis, non-negative matrix factorization, or *t*-distributed stochastic neighbor embeddings and presents a view where clusters of similar input points can be explored either by the dimensions which the dimensionality reduction summarized, or the retained dimensions. By applying PCA along the spectra dimensions of the volumetric nanoXPS dataset we might obtain in surveying WS_2/WSe_2, we form a digest of the energy spectrum S in terms of a small number of representative spectra e_{ik}. The decomposition summarizes spectra $s_k(x_j, y_j)$ at different locations (x_j, y_j) on the sample surface by representing them as a compact sum over only a few representatives

$$s_k(x_j, y_j) \approx \sum_i c_{ji} e_{ik}. \tag{5.1}$$

Often, just a few representative spectra are sufficient to approximately describe the variations in the data by summarizing the spectrum at location (x_j, y_j) in terms of a few coefficients c_{ji}. These coefficients can easily be scattered so that the experimenter can identify similar regions on their sample. The top right subplots in Figure 5.2a and b show this kind of view of the data. This representation is only so useful if it is purely static, however. In Figure 5.2a, we show how the exploratory tool in PyARPES permits selecting a group of measurements—the red circle on the scatter plot—which recomputes and presents the representative XPS spectrum (Figure 5.2a top right) and the location of the set of individual measurements (red outlined region, Figure 5.2a bottom left). This workflow lets the experimenter rapidly identify clusters of electronic behaviors across the sample surface. By corresponding the cluster locations back to the spectra e_{ik} or by viewing the averaged spectrum for the cluster, the physical meaning of these clusters can also be determined. For example, the selected cluster in Figure 5.2a shows overlapping W $4f_{5/2}$ and $4f_{7/2}$ peaks. These peaks are shifted and have different intensity ratios because of the decreased dielectric constant in the heterostructure from screening of the hBN substrate by the underlayer of WS_2. Additionally, the tool permits directly selecting contiguous regions in the representative spectrum which changes the intensity used in the spatial map, as is shown in Figure 5.2b. Because we are applying generic exploratory data analysis tools, the utility in PyARPES does not make requirements or assumptions about which dimensions should be reduced (here, spectral dimensions) and which dimensions should be mapped (here, spatial dimensions). For datasets resolved by other degrees of freedom, like temperature, the same analysis and tool can be applied transparently to visualize variations along whichever dimensions the scientist wants to explore.

This gives the impression machine learning's purpose in ARPES analysis is to automate or routinize very simple analysis. This runs counter to the lay interpretation of applied scientists who hear about machine learning and envision the computer providing new insights or answering unguided high-level queries about their data. The reality, although closer to the sober presentation above, lies somewhere in between.

To see that machine learning really can provide new insights, we will consider one more dimensionality reduction application on the same dataset before looking at how tight coupling

between machine learning and ARPES experiments lets scientists get around fundamental limits in their hardware, at least in constrained settings.

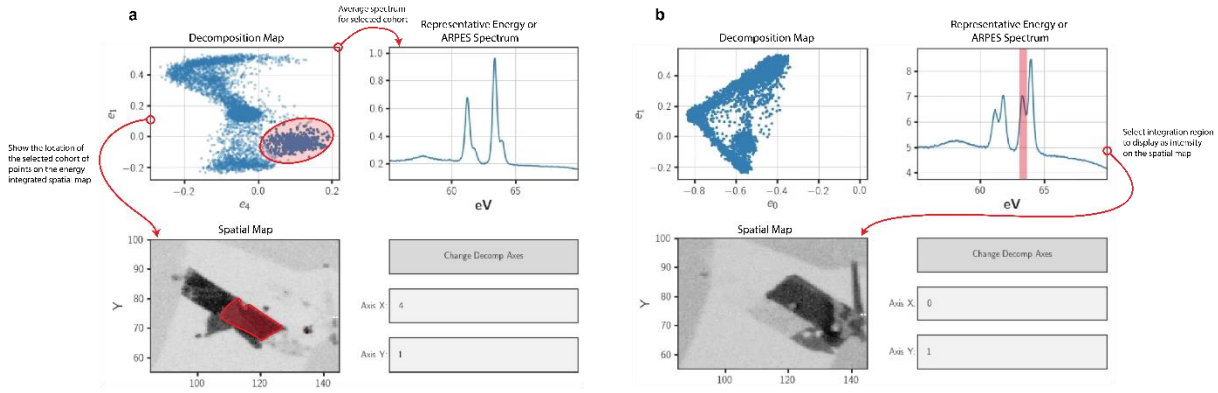


Figure 5.2 **Example interactions with an ML based exploratory data analysis tool. [a]** By selecting a group of points which belong to a cluster in the embedded dimension in the scatter plot, the tool presents a representative energy spectrum which consists of the average over the selected cohort. At the same time, it highlights the spatial distribution of individual curves in the map which shows that the related cluster all belong to the heterostructure region of the sample where the WS_2 and WSe_2 overlap in a moiré superlattice. After identifying a particular cohort, we can select the collection of individual curves for further analysis outside the interactive exploratory analysis tool, or the defining features of the cluster can be analyzed by examining the embedding dimensions $\boldsymbol{e}_4$ and $\boldsymbol{e}_1$. **[b]** By selecting a region from the energy spectrum, we can change the intensity which is plotted in the spatial map, permitting an exploration of the data based on its spectral characteristics as well as its characteristics in the embedded dimensions. **[not shown]** Any pair of the embedding dimensions can be chosen for the scatter plot in the top left using the selection controls in the bottom right quadrant of each tool view.

5.2 Noticing Variations in a Diffuse Heterostructure Band with Dimensionality Reduction

This kind of dimensionality reduction can also make it possible to spot spectral features which are extremely difficult to notice directly in the raw data, but which warrant additional analysis and consideration on the part of the experimenter. Although this still falls under the purview of exploratory data analysis, it represents a very different kind of workflow than we already discussed, because it can spur the application of more rigorous analysis and lead to novel finished analysis products.

This is especially likely to happen in situations where the experiments are or are very nearly signal-to-noise (SNR) constrained. As experiments resolve by more dimensions during an experiment and produce higher dimensional data, data either becomes sparser or more statistically undersampled because the electron flux from the sample is limited by space charge and other fundamental measurement constraints. Some improvements in spectrometers, as in two angle-resolving time of flight spectrometers, will improve this situation, but not by enough to compensate for the increased ambition of a generation of ARPES scientists used to nanoscale measurements.

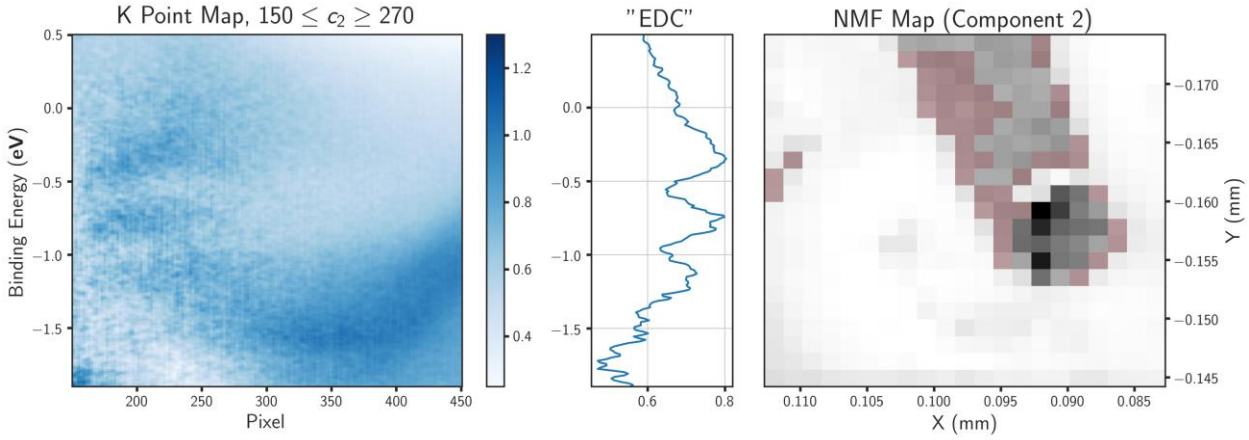


Figure 5.3 **Identifying Diffuse, Nearly Flat Bands in an Inhomogeneous Heterostructure.** When ARPES experiments are conducted on inhomogeneous samples, making sense of the resulting measurements is complicated because individual measurements may not be statistically robust, but measurements conducted at neighboring points on the sample surface may have very different electronics. This situation precludes direct neighbor binning which may otherwise be used to improve the signal to noise ratio in an ARPES spectrum necessary to resolve a faint feature. Dimensionality reduction can be used to identify electronic "neighbors" which are nonetheless separated in the measurement dimensions (e.g. spatially separated). **[left]** $K\Gamma$ direction averaged cut, after background spectral component subtraction, from an inhomogeneous heterostructure. Despite individual spectra being too diffuse, a collection of similar spectra shows several nearly nondispersive levels at the K point which motivates further scrutiny. **[center]** A constant momentum slice through the cohort averaged ARPES spectrum in [left]. **[right]** A masked spatial image showing the individual measurement points (red) among all measurements on the substrate (lighter areas) and heterostructure (darker areas). We can see that the identified cohort consists primarily of the heterostructure boundary, possibly indicating contamination of the signal from neighboring monolayer regions and resulting in the presence of a more dispersive band at a 0.3 eV binding energy.

Common approaches to mitigate these problems include binning data along one or several of the measurement directions under the assumption that the variation in the spectral characteristics is slow enough that when viewing this process as the application of a smoothing kernel, the increase in the electron arrivals in the smoothed window more than make up for the spectral inhomogeneity in the window. How likely is this assumption to hold? From experience, it is not uncommon to have spatially inhomogeneous samples where directly binning does not improve the effective SNR, because points which are neighbors in the measurement dimensions are not "electronic neighbors" or "spectral neighbors". Nevertheless, dimensionality reduction permits identifying a cohort of similar spectra from a large sequence of measurements which are alike, and therefore are electronic neighbors.

This approach requires the careful attention of the experimenter however, because algorithmic approaches to selecting a subset of the measurement set for integration and further analysis may bias the result and the selected cohort may not carry any semantics which make it desirable for analysis. The utility, and the dangers, of this approach are again best presented by a real example from the superlattice WS_2/WSe_2 project of the previous chapter. Despite efforts to control the fabrication process, electronic inhomogeneity is nonetheless present in some heterostructure samples. These inhomogeneities, higher dimensionality of ARPES cuts over XPS spectra, lower valence photoelectron yield compared to core levels, and very low K point intensity in tungsten

based TMDCs makes interpreting 4D (E_b, ϕ, x, y) spectral maps to explore the K point in the heterostructure extremely challenging. Initial efforts conducted by direct examination of the raw data seemed to suggest that it was impossible to make any direct claims about the K point in these materials. However, applying both a contrast enhancement approach by subtracting local backgrounds and performing a decomposition along the spectral dimension, as in the prior example, we can see the emergence of several bands at the K point (left of Figure 5.3) upon drawing from a complex subset of the sample surface shown in the right subplot of Figure 5.3. Perhaps strangely after the discussion of the previous chapter where two peaks at the K point were observed, we see three peaks in the constant momentum curve taken near the high symmetry location (central plot in Figure 5.3).

By considering the spatial structure of the sample and neighboring regions, we determined that the presence of this additional peak is due to signal contamination. This is perhaps unsurprising given that the cohort selected for analysis by the decomposition appears to consist more or less of points at the border of the heterostructure region. Nevertheless, this analysis justified further and more careful effort into observing the spectral features and flat bands states at the K point—an effort which concluded with the inclusion of a proper ARPES spectrum in our superlattice WS_2/WSe_2 paper.

5.3 Using Machine Learning to Extend the Capabilities of ARPES Experiments

In the previous section, we explored how even simple dimensionality reduction approaches can be used to rapidly explore ARPES data and facilitate traditional analysis. In machine learning applications, we often hope for more. Can machine learning be used to treat data in ways entirely outside of traditional ARPES analysis approaches? Usually, we are interested in a slightly more profound question. Can we use machine learning to extend the capabilities of our instruments, and accomplish previously impossible experiments? In the remainder of this chapter, we will investigate one very narrow way in which machine learning can be used to stretch the limits of current ARPES experiments, subject to the sample and experimental conditions. First though, we will situate why approaches like this are necessary, despite decades of progress in instrumentation.

Angle-resolved photoemission spectroscopy (ARPES) has earned a central role in the characterization of quantum materials (*7*, *8*) and as a tool in feedback with theoretical calculations and growth techniques to refine or tailor new material properties. The utility of the technique stems from its ability to resolve both many-body effects and to directly measure the band structure, which sets the essential material properties for energy and computational applications. While early ARPES experiments were applicable primarily to macroscopic single crystals, current and emerging ARPES experiments have brought the technique nearly to the nanoscale (*14*, *39*, *44*, *142–144*), through a combination of efforts to reduce beam sizes at synchrotron light sources, and in the ARPES instrumentation to turn existing spectrometers into scanning microscopy platforms. This approach may nevertheless reach a practical limit. Even if instrumentation advancements continue to drive spatial resolutions lower, samples of interest, especially those where the community stands most to gain from insight into nanoscale inhomogeneity, do not tolerate the intense x-ray doses that accompany further shrinking x-ray

probes (*44*, *145*). Analysis techniques that make the most of currently available probes and which are transferrable to higher resolution experiments will be an essential component to study these more fragile systems and push the boundaries of experiments.

As introduced, the general program of extracting more information from currently available apparatus is underdefined. To make progress in a narrower problem space, we will consider only the case where we perform scanning spectromicroscopy on samples with a finite number of distinct electronic orders on the surface. To motivate practical examples, these discrete orders could be due to domain patterns (*146–149*) on the surface of the sample, different surface terminations (*150*, *151*) arising from multiple cleavage planes, or bi-, tri-, and multilayer contamination in thin films grown in situ. In this case, material scientists want to access the electronic structures of the individual domains or terminations uncontaminated by those of their peer orders, but they cannot do so if the characteristic length scale λ of the domain or uniformly terminated area is smaller than the beam resolution σ. In this setting, using only slight adjustments to experimental protocols permits the recovery of full ARPES spectra from the constituent domains via decomposition of results obtained with large beams. We accomplish this by leveraging statistical fluctuations in the measurement for an effective improvement in resolution.

This phrasing of the problem places the recovery of ARPES spectra from mixed signals squarely in the context of spectral unmixing: the general problem of recovering unknown signals mixed in fractional abundance given mixed observations. Spectral unmixing (*152–157*) has found applications in a variety of remote sensing (*157*, *158*), spectral imaging (*159*), and other domains, but properly describes a problem class rather than a set of solutions. Focusing on any one domain, we find a wide collection of algorithmic techniques used to solve a specific phrasing of the unmixing problem: in some domains the number of spectral components or endmembers is unknown, in others there is available structure in the spectra to be recovered or in the abundances or distributions of how the spectra are mixed in observation. The recovery of ARPES spectra shares features with both the statistical algorithms approaches and the geometric approaches to unmixing (*152*). This is because in the case of a discrete set of patterned domains, the measured signal follows a linear mixing scheme with nonnegative signals and nonnegative mixing proportions. In ARPES, signals are nonnegative because they represent angle resolved photoemitted currents which are nonnegative and detected as discretized counts S on an analyzer. However, the photoemission process can be thought of as occurring independently at every location under the beam and parameterized by some configuration vector $\boldsymbol{\theta}$, so that the analyzer measures a spatial average $\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}}$. ARPES interprets the photocurrent in relation to the single particle spectral function modulated by matrix elements, the experimental resolution, nonidealities of the surface, and other effects, and we can think of $\boldsymbol{\theta}$ as capturing the dependence of the measured spectrum on these factors.

Many of the factors comprising $\boldsymbol{\theta}$, like surface disorder, are continuously variable in their effect. In some cases $\boldsymbol{\theta}$ is essentially finite dimensional, as in the case where we measure ARPES from symmetry equivalent electronic domains below their transition temperature (*144*, *146*, *160*).

Because we perform measurements over finite regions of sample, our measurement $\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}}$ is best seen as a spatial average against the intensity profile of the probe beam I_{beam}

$$\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}} \equiv \int_{\boldsymbol{x}} dxdy\, I_{\text{beam}}(x,y) S\big(\boldsymbol{\theta}(x,y)\big). \tag{5.2}$$

Here $S\big(\boldsymbol{\theta}(x,y)\big)$ carries all the information relating to the light-sample interaction and the spatially resolved response of the system. We are considering the case where there are a discrete set of behaviors under the beam, and so we can identify the response which would be received by measuring on a perfectly uniform region of sample as $S(\hat{\theta}_i)$. Written this way and defining $\alpha_i(x,y)$ is an indicator function for behavior or domain type i

$$\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}} = \int_{\boldsymbol{x}} dxdy\, I_{\text{beam}}(x,y) \sum_i \alpha_i\,(x,y) S\big(\hat{\theta}_i\big) \tag{5.3}$$

$$\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}} = \sum_i S\big(\hat{\theta}_i\big) \int_{\boldsymbol{x}} dxdy\, I_{\text{beam}}(x,y)\alpha_i(x,y) \tag{5.4}$$

$$\langle S(\boldsymbol{\theta})\rangle_{\boldsymbol{x}} = \sum_i S\big(\hat{\theta}_i\big)\, \langle \alpha_i\rangle_{\boldsymbol{x}}\,. \tag{5.5}$$

Each measurement we conduct is a weighted average of the unknown spectra $S\big(\hat{\theta}_i\big)$. No single measurement will allow us to recover a component $S\big(\hat{\theta}_i\big)$. However, by performing many measurements S_j we can recover the components $S\big(\hat{\theta}_i\big)$ so long as the measurements sample enough variance in α_{ij}. Writing the matrix of measured spectra as M, the matrix of unknown spectra $S\big(\hat{\theta}_i\big)$ as S, and the matrix of unknown sampling fractions A, we can use non-negative matrix factorization to find

$$\min_{S\geq 0, A\geq 0} \|M - SA\|_F + \lambda J(M,S,A). \tag{5.6}$$

The inclusion of a term λJ regularizing the solution to encode prior knowledge about the structure of ARPES spectral data will be discussed towards the end of this paper. When $\lambda = 0$, we have the standard formulation of linear nonnegative matrix factorization (NMF). This approach, culminating with the NMF objective is very natural because we have given a generative description of the nanoARPES signal with only the approximation that $\boldsymbol{\theta} \approx \theta_i$.

In condensed matter physics, nonnegative matrix factorization has shown some promise in reducing the dimensionality of data stemming from diffraction experiments on continuously varying material growths (*161*) and occasionally in the interpretation of low-dimensional spectroscopies (*162*). These applications to condensed matter physics treat the spectral unmixing process as a dimensionality reduction tool which increases the ease with which data can be analyzed, rather than as a technique by which otherwise unmeasurable spectral features can be recovered. Having situated the problem, we will consider applying regularized nonnegative

matrix factorization throughout the remainder of this chapter. First, we will consider the potential limits of performance in simulation for domain structures, which will motivate a connection between NMF analysis and making use of statistical variability in the measurement process. This is an interesting and novel interpretation because ARPES usually tries to integrate variance out during measurement to obtain a smooth representative spectrum. Next, we will apply nonnegative matrix factorization to real μXPS data in a challenging context where the domain structure is smaller than the beam. Here, NMF analysis of μXPS data provides recovery of ground truth spectra despite the large intrinsic beam size. Finally, we turn to regularization approaches and how NMF can be used to unmix entire two-dimensional sparsely sampled ARPES spectra. In this final section, we provide a guide for reasonable application of NMF in ARPES, discuss conditions for interpretation, and propose a set of regularization approaches which should be considered in future applications.

5.3.1 Measurement Variance, Domain Structure, and One-Dimensional Simulations

In this first section, we will establish some connections between measurement variance and spectral unmixing using simulations and drawing on our intuition for the measurement process. We will establish spectral unmixing first by considering anticipated performance using simulations for photoemission from different domain structures, as seen in Figure 5.4a. Although real domains are highly variable in shape and size due to pinning and local energetics, we will restrict our attention to checkerboards and Ising model simulations moderately below critical temperature. These two systems, although not directly realistic, are advantageous for providing a rigorously controllable paradigm in which to conduct simulated photoemission experiments, and because they represent a pessimistic limit (checkerboard structure) and optimistic limit (Ising domains) for real experiments. Throughout the discussion, we will characterize our beam size σ, shown in Figure 5.4b, in unitless terms where possible. For checkerboards, this means we will use σ/L (Figure 5.4c) as the beam size.

The parameter σ/L functions to represent the capability of the hardware to resolve the local structure on the sample. If $\sigma/L \ll 1$, as we usually assume we when perform scanning microscopy in ARPES, then it is straightforward to resolve the domain structure and to obtain the "pure phase" spectra. On the other hand, if $\sigma/L \approx 1$, every spectrum will be highly contaminated with the minority abundance spectrum, which prevents measurement of the underlying pure phase spectra without analytical processing. Phrased in the language of the introductory section on NMF, when $\sigma/L \approx 1$, then $\alpha_{1,i} \approx \alpha_{2,i} \approx 0.5$ for all measurements m_i. If we define a polarization measure

$$p_i \equiv \left| \frac{\alpha_{1,i} - \alpha_{2,i}}{\alpha_{1,i} + \alpha_{2,i}} \right| \tag{5.7}$$

and consider taking many measurements m_i at locations $\boldsymbol{x}_i$, we can examine the connection between the two beam size limits. Figure 5.4d shows the expected polarization $\langle p \rangle_{\boldsymbol{x}}$ (blue scatter) as well as the variance in polarization (shaded band) as a function of the normalized beam size. As the beam size increases, the expected polarization under the beam shrinks rapidly and approaches zero near $\sigma/L = 0.5$. At the same time, the variance increases initially, as the beams

become more contaminated by minority phase spectrum before the variance shrinks again. This is because in the large beam limit, every beam sees equal proportions of the two components in each sample, so $p_i \approx 0$ for each measurement.

For Ising domains (Figure 5.1e), the situation is similar, but a few key details are changed. For one, the polarization never shrinks to zero because there are domains of different sizes and some instantiations of a measurement will happen to sample slightly more of one domain than another. This also means that the variance in the polarization never quenches to zero either. The critical difference between the Ising and checkerboard situations is in fact that at equal expected polarization, the variance for a more disordered domain is substantially larger, as can be seen in Figure 5.4f.

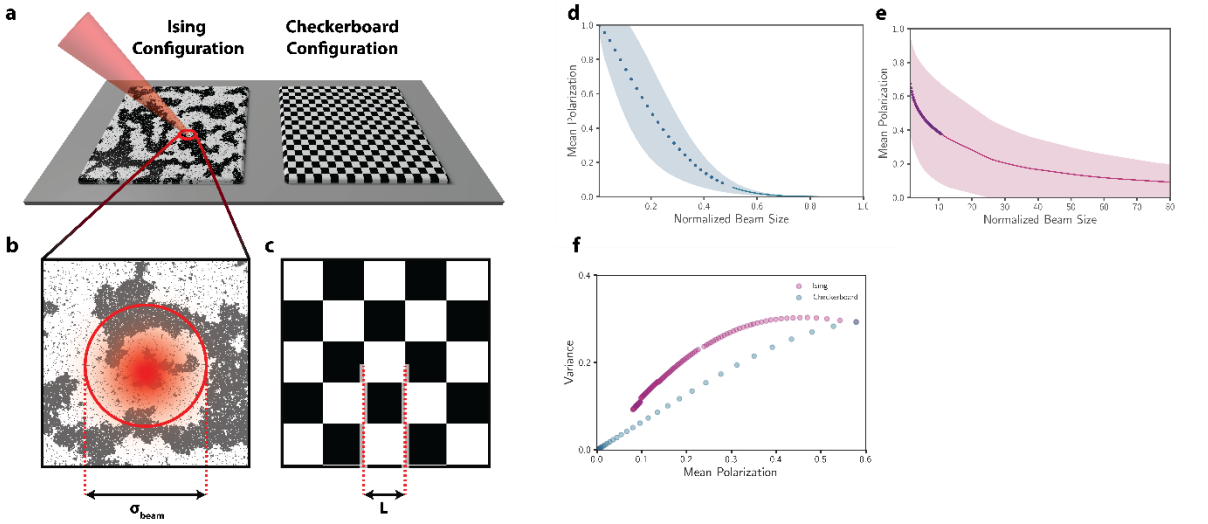


Figure 5.4 **Nano-ARPES of Discretely Structured Materials [a]** Checkerboard and Ising sample configurations presented in modeling results and **[b,c]** microscopic views of the beam interaction on these samples with their parameters. **[d,e]** Comparison of the polarization observed under checkerboard [d] and Ising type domain structures [e] for a variety of simulated samples and beam interactions. The envelope provides one standard deviation of the polarization observed around the mean value for repeated simulated experiments. A normalized beam size cannot be provided in the case of the Ising type domain without resort for the correlation length. **[f]** Comparison of the variance in observed domain polarization for an ensemble of modeled sample-beam interactions on Ising and checkerboard structures, providing a size agnostic quantization of the difficulty in recovering the underlying spectra from samples with two domains. The checkerboard structure can be seen as a pessimistic limit where the variance is very small at a given mean observed polarization.

How does this connect back to the measurement process? Usually when we conduct ARPES measurements, we integrate the angle and energy resolved photocurrent for several seconds or minutes to obtain an exposure on the detector which is sufficiently smooth to accommodate some target analysis, like self-energy extraction. If there is disorder under our beam, the measurement we obtain, m_0, samples the coverage vector α_0 just once. In some sense, we are integrating out the structure in $\alpha(x, y)$ which would have been informative about the ground truth spectra had we incorporated more measurements α_i. To use unmixing, we want a representative sample of α_i over the whole distribution which we could measure across the sample. This requires a different

methodology at experiment time, instead of settling on a "locally optimal" spot to measure and collecting until we have integrated statistical variations in the electron arrivals out of the spectrum, we should take many short exposures across a region of the sample. This makes the approach compatible with many more ARPES instruments than where spatially resolved measurements have been applied. Apparatus which are not designed for nanoARPES, and even some which are, may experience thermal contraction lasting several hours and which prevents acquiring spectra at a single, fixed location on the sample. So long as we don't also demand to incorporate the spatial structure of the measurements, spatial drift in the sample location does not prevent acquiring many undersampled spectra across the sample surface. Of course, this cannot be taken to an extreme: we need enough electrons in each measurement m_i to establish statistical connections between the spectra, and this means each individual measurement must sit at a fixed location $\boldsymbol{x}$ long enough to establish this. This is vague, because the degree of sampling required is a function both of distribution of coverages $\alpha \sim \mathrm{A}$ and of how similar the component spectral $S(\hat{\theta}_i)$ are to one another. If $\mathrm{Var}(\mathrm{A})$ is small, or if $S(\hat{\theta}_i)$ and $S(\hat{\theta}_j)$ are very similar to one another, longer acquisitions are required to distinguish them robustly.

Figure 5.5 shows the results of applying this methodology to the question of recovering the pure phase spectra from noisy mixed measurements on a checkerboard. The most challenging situation to consider for ARPES is whether any signal recovery technique respects narrow band separations, required to examine low energy physics near band crossings and gaps. For simplicity in these early simulations, we will consider experiments drawn from a one-dimensional spectroscopy, like μXPS, on a two-domain system where the pure phase or ground truth spectra are offset Gaussian peaks, shown in dashed lines in Figure 5.5b. In Figure 5.5a, a representative EDC drawn using a beam with width $\sigma = 0.4L$ is shown. As ca be seen, the spectrum is highly mixed and a clear peak structure is not identifiable. This spectrum defines the single-frame acquisition statistics, and is comparable to acquiring at a core level for less than a few hundred milliseconds. In Figure 5.5b, we assemble a virtual experiment of 400 frames taken at random offsets in the sample to obtain the matrix M and using off-the-shelf nonnegative matrix factorization obtain the component spectra shown as red and blue scatter sequences. This demonstrates that even straightforward applications of unmixing to μXPS data can be promising, as the recovered peaks are in excellent agreement with the ground truth data despite the noisy and mixed inputs, but it provides little clarity as to whether we can expect reasonable results as a function of relevant system parameters like σ/L of the energy splitting between the peaks.

The remainder of Figure 5.5 explores these questions in detail. In Figure 5.5c, we look at the effect of NMF performance (blue scatter sequence) for spectrum unmixing as a function of the beam size. As an alternative to the measurement and analysis protocol used for the NMF decomposition, we contrast the expectation results (purple scatter sequence) of merely picking a random spot on the sample—as might otherwise be done by an experimenter using quality measures including spectral intensity or apparent peak width. In order to succinctly provide a quality measure for spectra obtained by each method, we use the recovered peak full width half maxima (grey curves, right hand side) and the coefficient agreement with the ground truth spectra g_i, defined by solving the least squares regression

$$G\mathrm{B} \approx s_i \tag{5.8}$$

for coefficients B. Because there are two components, this amounts to determining β_1, β_2 minimizing the reconstruction error $\|s_i - g_1\beta_1 - g_2\beta_2\|$. We define the coefficient agreement γ in terms of the regression coefficients β as

$$\bar{\beta} = \frac{\max(|\beta_1|, |\beta_2|)}{\sqrt{\beta_1^2 + \beta_2^2}}, \tag{5.9}$$

$$\gamma = \frac{\bar{\beta} - \frac{1}{\sqrt{2}}}{1 - \frac{1}{\sqrt{2}}}. \tag{5.10}$$

This performs an assignment of each spectrum onto the closer ground truth spectrum and normalizes the quality measure so that $\gamma(g_i) = 1$ and $\gamma(g_1 + g_2) = 0$. This allows us to interpret γ in terms of the quality of the unmixing. From Figure 5.5c, we see that the unmixed components have much higher agreement with the ground truth spectra than is obtained by choosing a spot to measure and performing no unmixing. In particular, the NMF components agree well with ground truth spectra over a large range beam size, suggesting that the recovered components can be trusted to provide a robust estimate of the pure phase spectra even when the beam size is a significant fraction of the typical domain size.

We can furnish another way of examining whether the NMF components are a robust estimate of the pure phase spectra by looking at the quality of the unmixed components for fixed beam size but as a function of the separation between the Gaussian peak locations for the pure phases normalized by their peak width. These results are shown in Figure 5.5d for a constant beam size of $\sigma/L = 0.4$. For very small peak separation Δ compared to the peak broadening σ_E, the two peaks cannot be recovered. Although our quality metric is designed such that $\gamma(g_1 + g_2) = 0$, in the this limit the linear regression is degenerate. This degeneracy causes the coefficient agreement to converge to a finite nonzero value in the small splitting limit. As the splitting increase, the NMF components rapidly saturate in quality, although the peak width increases moderately until the peaks are well separated ($\Delta/\sigma_\mathrm{E} > 3$). Here, nonnegative matrix factorization offers a substantial improvement in both quality metrics for a large range of peak splitting.

These simulations provide evidence that unmixing can be applied to recover ARPES and XPS spectra from mixed measurements, even in the limit where the beam size is very close to the typical domain size. Before we turn to a test application, we consider why performance drops off as σ/L grows past 0.3. From Figure 5.4c, $\langle p \rangle_{\boldsymbol{x}}$ decreases very rapidly as soon as the beam size begins to increase and is already very small by the time $\sigma/L = 0.3$. However, the variance in the polarization is only begins to decrease at around this point. This makes qualitative sense as well, NMF and related decomposition techniques are closely related to the variance in the data: even when spectra are contaminated, covariance in intensity in different regions of the spectrum indicate the existence of multiple underlying spectra in the two-component generative

description. Beyond $\sigma/L = 0.3$, $\mathrm{Var}_x(p)$ decreases rapidly, as does the quality of the recovered spectra.

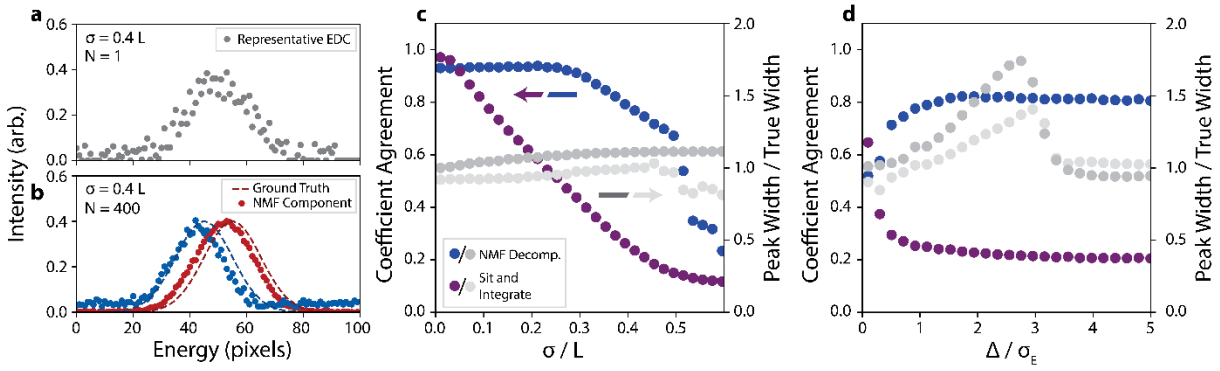


Figure 5.5 **Simulated μXPS results [a]** A simulated EDC drawn from a two-domain system where two Gaussian peaks are separated by an energy comparable to their linewidth. A single noisy exposure from a random location on the sample is shown here using a beam comparable to the size of a typical simulated domain $\sigma = 0.4\ L$. In this limit, every spectrum is at least moderately contaminated with signal from neighboring domains. **[b]** The dashed red and blue lines in the plot show the "true" single-domain spectra, while the blue and red scatter show the nonnegative matrix factorization components determined using $N = 400$ equivalently drawn random samples to [a]. Despite highly contaminated measurements and overlapping peaks, the nonnegative matrix factorized spectra are in excellent agreement with the true spectra. **[c]** Unmixing performance as a function of the normalized beam size σ/L. Nonnegative matrix factorization is compared with the expectation of merely picking a spot on the sample surface to acquire data using a figure of merit consisting of projecting the recovered (NMF) or measured (sit and integrate) spectra against the ground truth spectra by least squares regression. As a separate quality measure, the normalized FWHM of the peaks is shown on the right axis. **[d]** Unmixing performance as a function of the peak width normalized peak separation Δ/σ_E near the applicability limit of NMF ($\sigma = 0.4L$). At very small peak separation, both NMF and choosing a random spot to measure cannot distinguish between the identical peaks so the quality measure converges to $1/\sqrt{2}$.

What does this mean for real applications? We said that a checkerboard structure was a pessimistic limit for domain structures, because it is essentially perfectly ordered. By contrast, the Ising simulations we previously discussed at larger variance at equivalent polarization. This suggests that when applying spectral unmixing to ARPES spectra coming from real domain structures, we should expect better performance than indicated by the discussion of the results in Figure 5.5. In the next section, we will establish an actual connection between the simulations we have discussed, and unmixing of μXPS data.

5.3.2 Applying NMF Unmixing to NanoXPS Data from an Artificial Domain Pattern

In the previous section, we simulated results using Ising and checkerboard spatial patterns in a 1D spectroscopy using domains consisting of narrowly separated peaks to model the improvement in resolution which we can obtain if our task is to recover the underlying spectra. In this section, we test the resulting protocol on an artificial domain pattern of gold deposited on native oxide-Si by measuring the mixed core level and valence features with nano-XPS measurements.

Using experiments on artificial domain patterns is advantageous for several reasons. For one, some experimental validation is necessary. Numerical experiments provide valuable insight into

how our proposed approach works for nanoARPES spatial resolution improvement, but they do not provide a total and realistic model because they are built on an incomplete model for the physical photoemission process. Real materials are not perfectly partitioned into a small handful of discrete orders. Various continuous factors—ranging from chemical disorder on the surface or in the bulk, and x-ray induced beam damage to vacuum electrostatics—break this description, even for samples with physical domains. There are furthermore inherent sources of noise arising from the detection process, from space charge and temperature broadening, and from photoelectron scattering which are not well captured even by sophisticated models of the detector response. Consequently, it bears out applying the unmixing process to real experimental data on a tightly controlled sample. Using an artificial pattern is also advantageous because our knowledge of the test pattern's structure and the XPS and valence spectra of Au and SiO_2 permits straightforward comparison of model results to ground truth.

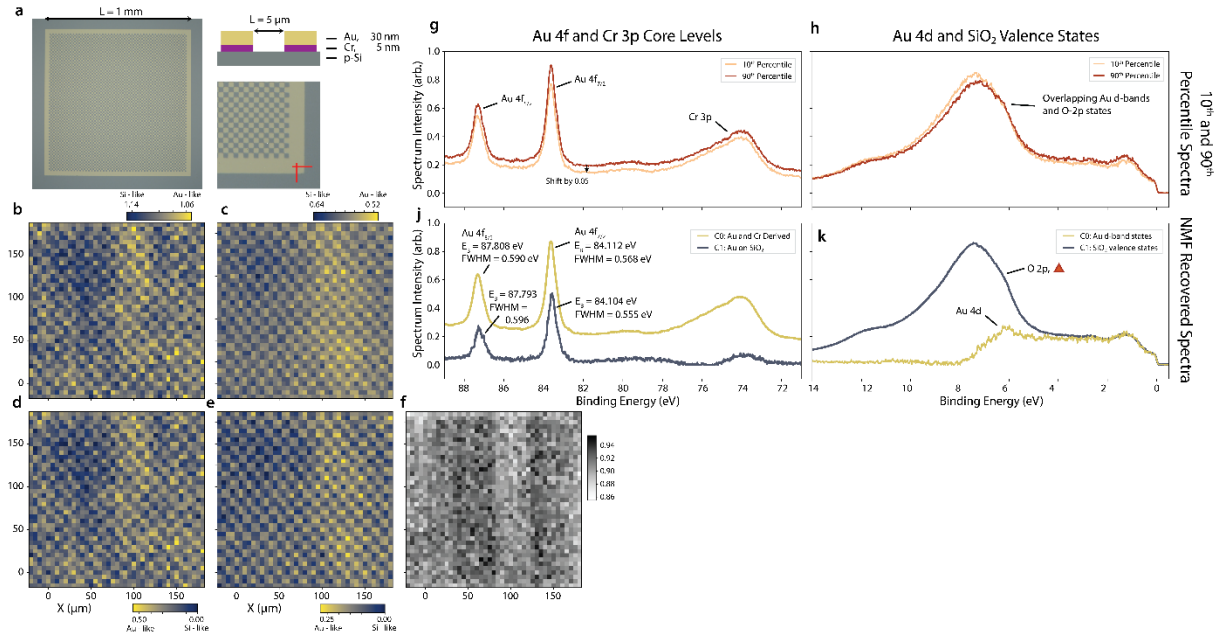


Figure 5.6 **Testing Spectral Unmixing with 1D μXPS [a]** Optical micrograph of Au test pattern on native oxide p-Si. A checkerboard of gold islands with a chromium stick layer was patterned across a 1 mm^2 region with a pitch of 10 μm. The red lines depicted in the inset show the locations where resolution checks (see Supplementary Materials) were performed. **[b-f]** [b-e] Coverage maps for isolated component spectra and original spectral contrast integrated across the Cr 3p peak [f]. **[b]** Spatial map of the Au and Cr derived spectral component in the core level energy region (see [j]). **[d]** Spatial map of the Au on SiO_2-like spectral component in the core level region. **[c,e]** Valence level regions corresponding to the equivalent spatial maps in [b,d]. **[f]** Energy integrated map for 1 eV bandwidth across the Cr 3p peak. Unlike in the identified component maps, the underlying checkerboard structure visible in [b-e] cannot be seen. **[g-k]** Source spectra and uncovered components. **[g]** Sample spectra at the 10th and 90th percentile of the Au-like component abundances. Percentiles are taken according to identified component abundances to isolate spectra represent the variance in the original dataset. The orange curve has been shifted downward for legibility. **[h]** Identical to [g] but in the valence region showing O 2p states. **[j]** Uncovered component spectra in the core level region with annotated peaks, chemistry, and parameters. **[k]** Discovered component spectra in the valence region.

In Figure 5.6a, an optical micrograph shows a fabricated test sample used for μXPS studies. Gold islands were deposited with a 10 μm pitch on the surface of a grounded p-Si chip using a chromium stick layer. Samples were cleaned before photoemission experiments by annealing at 210°C for six hours prior to measurement. Extremely short exposure XPS spectra were sampled in 5 μm intervals across a $200\ \mu m \times 200\ \mu$m area in the interior of the test pattern in the binding energy window of 90 eV to 70 eV and in the valence region from 15 eV to the Fermi level. These islands were smaller than the available beam resolution for any measurement conducted. Consequently, only highly mixed spectra comprising signal from bare p-Si and gold and chromium coated islands could be obtained. Figure 5.6g and Figure 5.6h show the degree of variation which could be observed at the 10^{th} and 90^{th} percentiles in the μXPS spectra obtained from the checkerboard interior of the sample. Even at these extreme quantiles, the variations in the spectra are subtle and do not permit distinguishing between states in the spectrum or in establishing the surface coverage of the different regions on the sample. Despite highly overlapping spectra in both the core level and valence regions, the application of nonnegative matrix factorization to the recovery of the Au-like and Si/SiO_2-like regions clearly picks out two spectra which together explain nearly all the variation recorded across the sample. For the purposes of showing variation in the raw spectra in Figure 5.6g and Figure 5.6h, spectra were ordered according to cosine similarity with the first, gold-like, NMF extracted component spectrum.

In Figure 5.6j, the recovered core level spectra are shown and assigned to predominantly gold island composition (gold curves in Figure 5.6j and Figure 5.6k) or $p\text{-}Si/SiO_2$ composition (blue grey curves in Figure 5.6j and Figure 5.6k). The spin-orbit split Au $4f_{7/2}$ and $4f_{5/2}$ states around 84.1 and 87.8 eV excellently match the binding energies and energy splitting from the literature. Additionally, the gold-like spectrum shows a wide peak around 74 eV attributable to exposed Cr stick layer after plasma cleaning of test patterns.

Interestingly, while the Cr $3p$ peak is not visible in the second isolated component, weak Au peaks are still present in this second component. Slight gold contamination of the p-Si regions due to plasma etching and subsequent anneal steps are consistent with the presence of these remnant Au $4f$ peaks. It is also possible that these peaks are due to fringing at the border between positive and negative regions in the checkerboard. Additionally, these weaker peaks are narrower, slightly shifted towards the Fermi level, and display a higher intensity ratio of the Au $4f_{7/2}$ to $4f_{5/2}$ peak intensities. Each of these effects is consistent with a small amount of gold migrated onto the p-Si and SiO_2 regions neighboring the gold islands. Negative core level binding shifts and peak narrowing are frequently observed because of size effects and surface effects in small sized particles. Although we cannot rule out measurement effects, no straightforward aspect of the XPS experiment can produce as an artifact both the negative binding energy shift and the different peak intensity ratios, although the latter can be explained by nonlinearities in the photoelectron detection process.

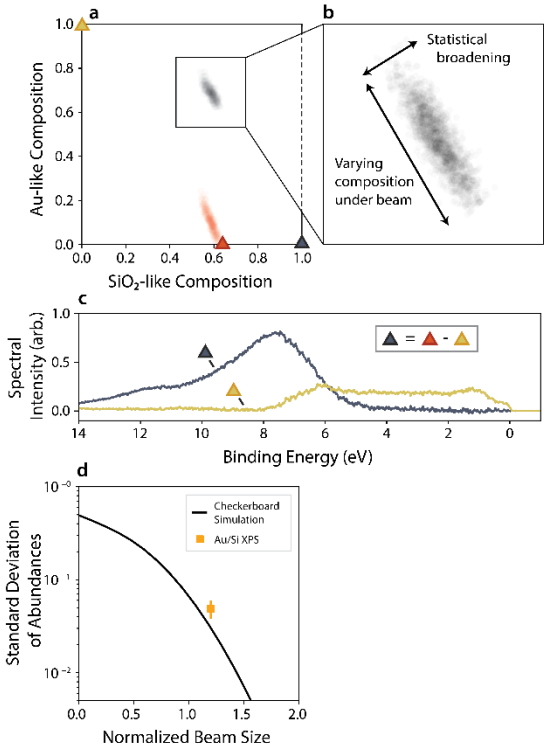


Figure 5.7 **Theoretical Comparison for Au μXPS Experiment [a]** Distribution of individual valence region μXPS spectra along the decomposition identified by the nonnegative matrix analysis in Fig 3. The red points are the raw coordinates while the black points are decomposed along axes which correspond to the difference of the underlying SiO_2/Au mixed and Au-like spectra. This change of basis is represented schematically by identifying component spectra with triangle labels in [a] and plotted in [c]. **[b]** Inset of [a] for the distribution of spectra after change of basis. The dataset is principally variant along a direction which interpolates between the Au-like and SiO_2 like spectrum. Variance in the opposite direction is due to spatially variant photoelectron yield, statistical broadening, and possibly two component model underfitting. **[c]** Spectra after change of basis to pure phase spectra. The red triangle curve corresponding to the mixed phase is labeled in Figure 3k. **[d]** Comparison of the experimental distribution of spectra to the simulated model of Figure 2. The black curve gives the standard deviation of the distribution of abundances as a function of the normalized beam size σ/L. The measured abundances extracted by NMF are in excellent agreement with expectations from simulation given the underlying sample geometry. The measured variance ($\sigma = 0.049 \pm 0.01$) slightly exceeds the simulation variance. Observed variations in the photoelectron yield across the sample slightly increases this variance, while departures from Gaussian beam shape on the sample would decrease it.

Turning now to the recovered valence states we can observe the same general phenomena. While the original spectra (Figure 5.6h) are too highly mixed to permit identification of spectral features, the unmixed spectra (Figure 5.6k) are clearly identifiable as the Au d band and surface states spanning roughly the range from the Fermi level to 8 eV binding energy (gold curve), and the SiO_2 valence states (blue grey curve) comprising O $2p$ derived states with a pair of characteristic peaks at 8 and 11 eV. Because of highly matrix element dependent effects for the valence states, we should be careful to assign otherwise overlapping features to a particular chemistry without additional guidance. Thankfully, a measurement of the SiO_2/Si valence states is available in the literature at a photon energy of 150 eV, matching the current set of experiments. As in the recovered core level spectrum, while the gold-derived spectrum appears

to have no contamination from SiO_2/Si valence states, the Au d-band is clearly imprinted on the Si-like recovered spectrum as well, matching expectations of some Au migration to the Si regions during sample fabrication and cleaning.

In addition to recovering the pure spectra, the factorization produces coverage maps which readily demonstrate the checkerboard geometry of the sample (Figure 5.6b-e) despite the integrated spectrum across the Cr-derived peak in Figure 5.6f showing essentially no spatial contrast across the sample surface. In Figure 5.6b and Figure 5.6d, the core level abundances are shown for the gold-like curve (Figure 5.6b) and the silicon-like curve (Figure 5.6d) with colorscale inverted so that the spatial maps can be interpreted to depict the location of gold-islands on the sample. Figure 5.6c and Figure 5.6e show the same for the decomposition performed on the valence region. We see excellent agreement between these recovered abundance maps and the real domain structure of the system, which suggests that the recovered NMF components above are not spurious statistical artifacts or attributable to some other intrinsic source of variation.

How closely do these experimental results match the simulated picture in the prior discussion? If they are very close, then our results of the prior section would indicate broader applicability to experiment. The most straightforward way to establish this connection is to look at the distribution of recovered abundance values, which is described in Figure 5.7. Figure 5.7a depicts a scatter plot of the recovered abundances in coordinate space defined by the SiO_2 like (x-axis) and Au like (y-axis) NMF components. To account for the possible gold contamination of the SiO_2 surface, we perform a coordinate change by subtracting out the Au-like composition, as in shown in Figure 5.7c. This transforms the original abundance point cloud, in red, to the grey point cloud. In Figure 5.7b, we notice that the distribution of the abundances is highly non-isotropic. Broadening along the minor axis is attributable to the statistical broadening from the short frame acquisition times while the major axis is due to variation in the polarization p for Au-like and SiO_2-like regions underneath the beam. By extracting the major axis length of this nearly Gaussian point cloud, we can compare the distribution of abundances as a function of beam sizes to the measured distribution of abundances and the spatial resolution of the experiment. In practice, the spatial resolution of the experiment was established by line profiling the Au step edge at the border of the sample as a function of the goniometer rotation angle θ and the entrance and exit slit parameters on the beamline (as is shown in Figure 5.8). By simulating the abundance variability according to the methods of Figure 5.4 with the known beam parameters and sample domain pitch, we can situate the recovered abundances on the curve in Figure 5.7d. In general, the observed abundances are in very good agreement with expectations for a checkerboard structure. A slight increase in the measured abundance variance is likely attributable to imperfections in lithography process and variations in the photoemission yield across the sample surface.

In this application to μXPS, nonnegative matrix factorization is both able to separate signals cleanly and to reveal unexpected spatial variation which would otherwise remain invisible. In the task of signal recovery, the experimentally obtained results are essentially in perfect agreement with numerical simulations (figure 4d). Meanwhile, the recovered spectra permit the

identification of the Au-negative regions with a small amount of gold due to test pattern preparation techniques.

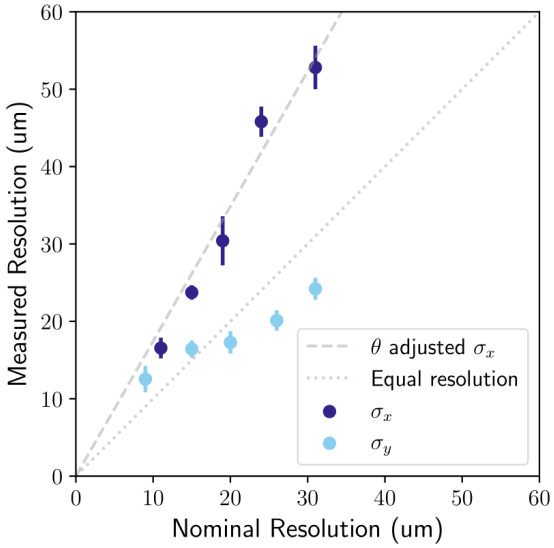


Figure 5.8 **Experimental Resolution Calibration for μXPS Studies.** The exit slit sizes at the MAESTRO μARPES beamline were adjusted across the nominal resolution range of $10\mu m$ to $30\mu m$ to select a beam size comparable to the pitch of the artificial domain sample. Because of the geometry, the projection of the beam on the sample is non-isotropic. After calibration a $16\mu m$ by $12\mu m$ beam was used for the μXPS experiments discussed.

5.3.3 Exploring Regularization Functions for ARPES NMF Applications

In the two prior sections, we demonstrated that off the shelf spectral unmixing approaches should be considered to extend the domain of applicability of nanoARPES to domain structured samples which are too small to be measured with current hardware. The technique we used, nonnegative matrix factorization, has been applied to similar affect in other disciplines, and yet we have used no aspects of the photoemission process leaving aside the generative description of our measurements. This is a missed opportunity, because ARPES data has internal structure which permits greater facility in unmixing data. Furthermore, regularization is also necessary to recover meaningful results in the limit where the spectra are always highly mixed, this is because nonnegative matrix factorization has degenerate solutions in this case, even leaving aside the nonconvexity of the problem which result in many local minima for the objective. As an example, if we decide that our data is well described by

$$m_i \approx \left(\frac{1}{2} + \alpha_i\right) S_1 + \left(\frac{1}{2} - \alpha_i\right) S_2, \tag{5.11}$$

then reparameterization according to $S'_1 = S_1 - \beta S_2$, $S'_2 = S_2 - \beta S_1$ gives a decomposition of the form

$$m_i \approx \left(\frac{1}{2(1-\beta)} + \frac{\alpha_i}{1+\beta}\right) S'_1 + \left(\frac{1}{2(1-\beta)} - \frac{\alpha_i}{1+\beta}\right) S'_2, \tag{5.12}$$

which is still a valid decomposition so long as $S_1', S_1' \geq 0$, together with the new coefficients. If the constraints on α_i are slack, we should expect some region where reparameterizations of this form give the same reconstruction error. This degeneracy needs breaking.

How should we use the structure of the ARPES spectrum? First, an ARPES spectrum is highly structured in that it is typically composed of several continuous bands with continuously varying intensity overlaid on top of a background. Depending on the material, these bands may look more like the sharp, well-behaved bands of simple metals, or more like the variable, self-energy modified bands of correlated materials like the cuprates. In either case, there are predictable commonalities. ARPES spectra are smooth, at least when enough electrons are collected, and they are composed of a discrete set of bands and possibly backgrounds. Under this breakdown of the spectrum as "signal plus background", it can sometimes be meaningful to ask for sparse bands, which pushes diffuse intensity into a background term. Finally, if we have information about the spatial domain pattern, which might come from another experimental technique like atomic force microscopy, scanning electron microscopy, or transmission electron microscopy, then this information can be incorporated too.

Incorporating smoothness is the most straightforward to accomplish. We can choose any number of regularization terms λJ which incentivize smoothness of the component spectra. Among available options, we can pick the total variation norm, which amounts to putting a penalty on the discrete pixel-wise derivative in each spectrum. We will consider $S_m(\theta_i)_{j,k}$ as one spectrum, so that m, i, j, and k index the measurement, the domain label/component variant, the energy, and the momentum respectively. For convenience, we will drop the measurement index (sample index) and we will use $\boldsymbol{t}$ and $\boldsymbol{s}$ as subscripts to refer to the spaces spanned by the component label indices and spectral indices, respectively. Then the total variation norm can be written as

$$J_{\mathrm{TV}}(S, \beta) = \sum_i \|\nabla_s \mathrm{S}(\theta_i) + \beta\|_* \tag{5.13}$$

for some norm $\|\cdot\|_*$. The hyperparameter β is included for numerical stability reasons and to ensure differentiability for some selections of norm. We will use the L_2 norm so that

$$J_{\mathrm{TV}}(S, \beta) = \sum_{ijk} \left(\left| \mathrm{S}(\theta_i)_{j,k} - \mathrm{S}(\theta_i)_{j,k+1} \right|^2 + \left| \mathrm{S}(\theta_i)_{j,k} - \mathrm{S}(\theta_i)_{j+1,k} \right|^2 + \beta \right)^{\frac{1}{2}}. \tag{5.14}$$

Other regularization approaches which attempt to enforce smoothness can also be considered.

Incorporating local structure is more complicated than directly regularizing the spectral features. In part, this is because ARPES effectively records images and sophisticated machinery has been devised to handle denoising in images over many decades. However, it is also hard to regularize A because a reasonable distribution to the values of A requires some knowledge of the underlying domain structure. In the case where this is known, then the distribution of values A_{mi} can be inferred. For instance, in the discussion of Figure 5.7, we observed that the distribution of values A_{mi} should be a normal whose major axis can be inferred from beam resolution and the

domain size. The Kullback-Leibler (KL) divergence provides a nonsymmetric notion of distance between probability distributions P and Q defined over the same range $\{x\}$

$$D_{KL}(P\|Q) \equiv \sum_x P(x)\log\big(P(x)\big) - P(x)\log\big(Q(x)\big). \tag{5.15}$$

If NMF produces an empirical distribution A of the abundances, we can use as a regularization term

$$J_{KL}(A,\mu,\sigma) = D_{KL}\big(A\|N(\mu,\sigma)\big). \tag{5.16}$$

In practice if we assume that the empirical distribution of A produced by the decomposition is roughly Gaussian, we enforce a term like this by taking the empirical mean and standard deviation, μ_e and σ_e at which point

$$J_{KL}(A,0,1) = \sum_i {\mu_{e,i}}^2 + \sigma_{e,i}^2 - 1 - \log\big(\sigma_{e,i}^2\big). \tag{5.17}$$

In the simple case where there are two domain structures, then $\sum_i \alpha_{m,i} \approx 1$ and this regularization term need only be enforced for a single component.

Consequently, we will examine the effect of these nanoARPES specific regularization terms in simulation by using projected gradient descent to optimize the objective

$$\min_{S\geq 0, A\geq 0} \|M - SA\|_F + \lambda_1 J_{TV}(S) + \lambda_2 J_{KL}(A,\mu,\sigma). \tag{5.18}$$

Our evaluation criteria should be more sophisticated than merely measuring the reconstruction error $\|M - SA\|_F$, because not all pixels in an ARPES spectrum are created equal. We should instead take care to recover the spectrum well in places where there is degeneracy in the solution which might belie the existence of a gap or anticrossing. We may also care to recover summary features—features which can be extracted from the spectrum which characterize the spectrum or some collection of bands in the spectrum—more accurately. As an example, we usually would prefer to recover the band velocity and dispersion peak locations, which via the bare band and self-energy have a direct connection to the physics, over reproducing the intensity along the band.

Figure 5.9 presents some results following this program for intersecting quadratic bands. In Figure 5.9a, we start with the definition of two quadratically dispersing bands S_1 and S_2, as well as a background term B. From these independent components which play the role of the energy split spectra in Figure 5.5, we generate infinitely sampled mixed spectra. These smooth spectra are then sampled with a finite number of electron arrivals and smoothed by a point spread function representing the experimental resolution degradation. These sampled spectra can be seen representatively in Figure 5.9b. Off the shelf NMF solvers generally do not support the incorporation of specific regularization terms in the objective function. Primarily, this is because NMF is usually optimized with special purpose iterative solvers for specific objectives to increase stability and performance.

To solve the objective function including J_{KL}, we implemented a projected gradient descent optimizer providing GPU acceleration on top of JAX (*163*). Our projected gradient descent optimizer implements hyperparameter scheduling and several different regularized objectives including Kullback-Leibler regularization on the mixing coefficients and more standard options like Tikhonov regularization and total variation regularization. To improve performance and to improve the stability of the gradient descent, we implement warm starting and objective annealing so that regularization terms can be slowly introduced as a perturbation from the unregularized NMF objective. In practice, this very reliably can be used to optimize otherwise challenging objective functions quickly.

Figure 5.9c and Figure 5.9d show the recovered component spectrum in closest agreement with S_2 when solving without the J_{KL} regularization term (purple) and when including the term and enforcing a standard deviation of α comparable to the simulated dataset's true distribution (green). While the component spectrum in Figure 5.9c shows significant contamination of the spectrum by S_1 because of degenerate solutions to the factorization problem, the spectrum in Figure 5.9d is in excellent agreement with S_2, differing only in the intensity along the band near the crossing location.

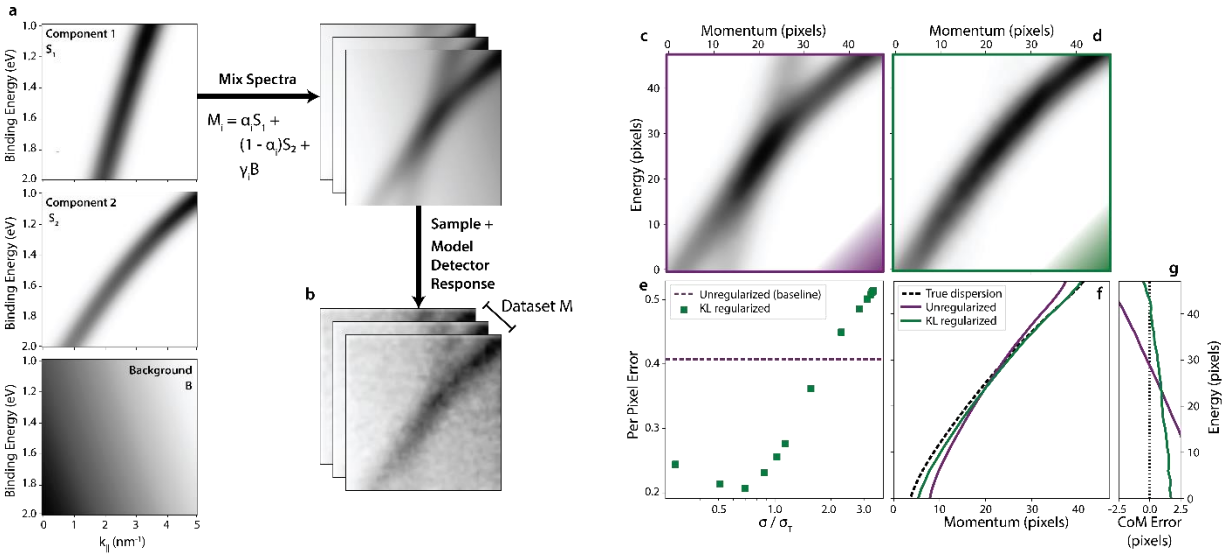


Figure 5.9 **2D ARPES Simulations Exploring Kullback-Leibler Regularization for ARPES Spectrum Unmixing. [a]** Generative structure for datasets used in the 2D simulations. Underlying component spectra are generated which define ground truth for recovery of the ARPES spectrum. Brightness variations are applied to simulate the effect of matrix elements in the photoemission process. Background components which vary in both momentum and energy are included but have no dispersive components or scattering from simulated bands. Spectra are then mixed according to abundances α and background intensities γ. For each simulated spectrum, a detector model is applied which models both finite electron yield in each exposure and the point spread of the analyzer and photoelectron detection scheme. These noisy simulated spectra constitute the dataset used in the remainder of experiments. **[c,d]** Approximate S_2 spectra obtained using [c] iterative optimization of the standard NMF objective and [d] a projected gradient descent optimized solution to the objective including the J_{KL} term. **[e]** The per-pixel normalized error for the $J_{KL}(,\mu_T,\sigma_T)$ regularized solution as a function of the target σ_T. **[f]** A comparison of the dispersion recovered from the spectra [c,d] and from the ground truth spectrum S_2. The residual center of mass location relative to the true spectrum is shown on the right-hand side.

Using the center of mass (CoM) as a function of the energy coordinate as a proxy for the MDC peak location, we can assess how well each objective reproduces the true band dispersion. Figure 5.9f shows the CoM dispersion for each of the spectra in Figure 5.9c (purple line) and Figure 5.9d (green line) as compared to the CoM extracted from the ground truth spectrum S_2 (dashed black line). Although the Kullback-Leibler regularized spectrum does not perfectly reproduce the underlying dispersion, it is in much better agreement with the true dispersion over the entire energy window and agrees nearly perfectly in terms of the band velocity, as can be seen by looking at the error residuals in the panel on the right-hand side. To verify that the Kullback-Leibler term really is controlling the output variance in the mixing coefficients and that good results are obtained when the output coefficients are distributionally matched to the real mixing coefficients, we repeated experiments as a function of the target distribution σ_T. The per pixel reconstruction error is plotted as a function of σ_T/σ, so that $\sigma_T/\sigma = 1$ indicates that the output distribution has the same standard deviation as was used to generate the dataset. For reference, the reconstruction error for the spectrum Figure 5.9c is shown as a dashed purple line. Excellent agreement is achieved for a window near, although not exactly at, $\sigma_T/\sigma = 1$. On either side of this minimum, the agreement with the ground truth spectrum decreases suggesting that this suggested regularization approach is effective in the recovery of highly mixed spectra when the domain structure is known from other experimental techniques or can be estimated ex situ.

5.4 Impact and Outlook for Spectral Unmixing in ARPES

In applications to date, spectral unmixing has occupied a narrow niche in condensed matter physics experiments. Primarily, unmixing and related approaches have found application in dimensionality reduction and have not been used to satisfy the primary aim of the experiments: to contact ground truth concerning physical properties of materials and systems. As we have seen, this represents a missed opportunity where the resolution of practical experiments can be dramatically improved without modifications to the apparatus and hardware used. A reluctance to use machine learning based tools to extend the reach of spectroscopy is nonetheless prudent. Hardware improvements continue to deliver appreciable gains—albeit at increased complexity—and concerns already abound about whether existing analysis approaches which rely on postprocessing the acquired photoelectron spectrum, like second derivative analysis and gap fitting, authentically represent the original data.

As we saw, not only can unmixing approaches provide physical insight into material properties below the resolution limits, as in the case of micropatterned gold arrays, information about the structure of ARPES spectra—dispersive peaks, smooth spectra, nondispersive backgrounds—can be encoded into the methods. When comparing to techniques we already rely upon in ARPES, this flexibility paints unmixing less as a blunt instrument than as a powerful situational tool. Furthermore, resolution improvements which come from the application of our unmixing approach in photoemission are cooperative and compatible with existing approaches, like deconvolution. In deconvolution's case, this is because spectral unmixing directly leverages the variations in the different spectra, whereas deconvolution essentially leverages spatially variation and knowledge of the beam shape to improve the resolution. As a brief example of how these techniques can be applied cooperatively: unmixing can be applied to reduce the dimensionality

of the deconvolution problem. In practice, this leads to much more stable and robust results than arise by applying deconvolution directly to the entire photoemission dataset. Alternatively, as we discussed, the spatial structure can be baked into the regularization terms we use in the unmixing process to obtain better results.

6 Conclusion and Thoughts on the Future Coevolution of ARPES Data and Its Interpretation

This dissertation has attempted to maintain a linear narrative about tools developed to improve the state-of-the-art in photoemission measurements as its spine. Although this is structurally useful, it is a shallow deception. It's clear by glancing at the publication order for the papers in the curriculum vitae that although the analysis tools in PyARPES predate the experimental studies that use it, AutodiDAQt and the work on ML came later. As presented, the ordering represents an ideal conception of how I would have used my time at Berkeley. With perfect information, I could have built the tools I needed first—and quickly. Alessandra would have been happy to see more experimental projects downstream from PyARPES.

Instead, PyARPES as presented here is much different than it was first conceived. It is simpler, more flexible, and contains vastly more utility covering the needs of others in the group rather than mine solely. It got this way because of a dialogue between the tools and the experiments it was employed on. The ML tools and early spectral unmixing project code grew out of a struggle to interpret the heterostructure WS_2/WSe_2 data. Dramatically improved performance characteristics, especially in momentum conversion and curve fitting, evolved out of challenges encountered by others on slower hardware or with larger datasets than mine. Data visualization tools improved with each experiment I worked on. Figure annotation, drafting, and data provenance were driven by trying to struggle through learning how to use Adobe Illustrator fluently. Although the details are not relevant, a similar story transpired with AutodiDAQt—I drafted early versions of the software in 2016 but large design changes were driven by my experiences with LabView and assisting in the development of the group's AR-PEEM.

In retrospect, I've come to realize that this is just how lab-scale experimental science works: it is beneficial for the quality of the tools and science when the two are in dialogue. Those of us working in condensed matter physics are lucky not to have to design the experiment, compute, and data infrastructure correctly in advance, as is the norm in high energy physics and astrophysics. Although this iteration is better for the scientists who can use the rapidly improving hardware apparatus and software tools, it makes my job more challenging here because predicting how this dialog might unfold is difficult. Thankfully, many changes which will likely take a decade or more to materialize are already ongoing and these give a place to start.

6.1 Machine Learning to Establish Stronger Connections with Theory

Using machine learning directly to make material properties predictions directly from measured spectral functions is has too many technical challenges and is underspecified in the general case. For one, most regression-oriented machine learning techniques require large volumes of data to calibrate, and these will not soon be arriving from current generation ARPES experiments because ARPES data is currently in the "wide data" rather than the "big data" limit.

The alternative approach is to use data from theoretical models as a crutch. There are two ways this can work, the first is to use experiments as a high signal constraint on theoretical models and to extrapolate properties away from the smaller core of materials where we have experimental

data. This is, to be overly reductive, the approach of the Materials Project (*164*), pymatgen (*74*), and many individual experimental efforts (*165*, *166*) and it is bearing fruit. So far, ARPES has mostly not participated in this growing trend. My conjecture is that this will change as experiments get less expensive and the reduction of ARPES data becomes simpler and more routine. PyARPES (*15*) represents our attempt to handle the second half of that problem, but there are other efforts as well (*167*, *168*). A second approach to integrating machine learning is to directly try to pattern match experimental band structures against theoretical models. This approach is also bearing fruit in the direct correspondence of mapped electronic bands and predicted band structures (*169*), but other property correspondences which try to match low energy spectral features and self energy modifications—supposedly ARPES' bailiwick—will be very interesting to watch for as well.

Although I've said that direct spectra-to-property measurements will be difficult, current approaches to ARPES analysis essentially try to do this with a human in the loop and with the simplest possible machine learning model: Lorentzian curve fitting. I expect that with mature software environments for analysis, the field will see more progress to extracting material parameters robustly and "autonomously". Over time, we may make progress on the general property prediction problems by slow emergence of more powerful analysis techniques.

6.2 Autonomous Experimentation

Viewing current analysis approaches for photoemission data also suggests some ways in which machine learning will become more tightly integrated into data acquisition itself. We have discussed approaches to AITL and the advantages it can bring, but it is also conceivable that beamlines specializing in certain aspects of the experiment, like very high spectral resolution and spatial resolution, may opt to integrate property prediction using ML once these techniques have been proved out. As an example, if robust gap analysis can be performed on very undersampled spectra using phenomenological models and global fitting, experiments may provide rapid physical property mapping to speed up sample surveying for otherwise impossibly complex experiments on superconductors and 2D correlated materials, as is shown schematically in Figure 6.1. If models and analysis methods used can measure or predict confidence intervals on their predictions (e.g. heteroscedastic methods), then measurements could be configured to acquire just enough data at even point in configuration space to satisfy experimental objectives set before measurement begins, making much better use of acquisition time. Even if beamlines and user facilities don't do this directly, users may import this functionality themselves via AITL as it is adopted into experiments via more data acquisition systems.

In a more limited context, this is not merely academic. There is a push towards autonomous surveys of sample electronics to make parsimonious use of limited experiment time and in service of the broader objective of autonomous end-to-end materials discovery. At my host institution, experiments have already integrated Gaussian Process regression to maximize the marginal utility of beam for single-sample surveying (*82*), but these approaches need to be adapted to better distinguish between physically meaningful and incidental sources of difference in the measured spectra, or else experiment time will be wasted autonomously mapping binding energy variations and mis-rotations on the sample surface instead of the material desiderata.

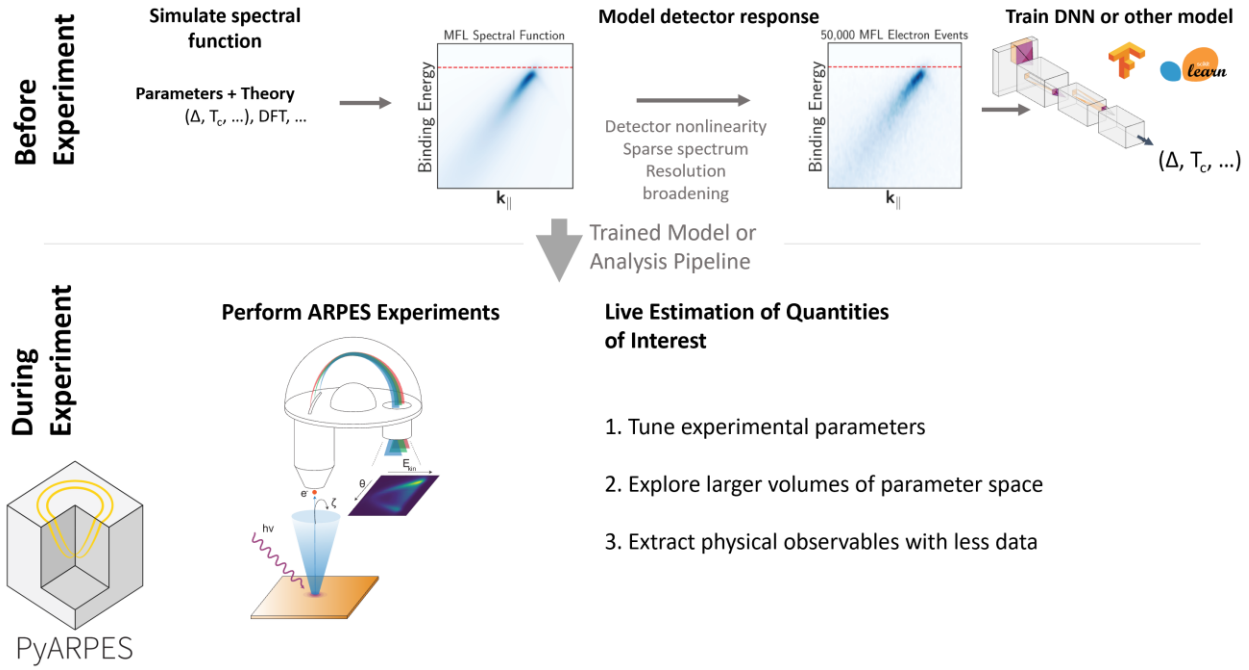


Figure 6.1 **How analyses might end up integrated into experiments in the future. [top]** More tightly coupled theoretical models which are fed through phenomenological models of the detector response and experimental geometry can be used to generate datasets and tune performance of ARPES analyses and novel ML-driven approaches. These can be used in analysis or prepared before an experiment to be used to assess the features of a sample during measurement. **[bottom]** By applying training ML models or analysis pipelines in experiments, we can adjust control parameters of the experiment as in autonomous experimentation. This application can improve the volume of the configuration space which is sampled from. Additionally, using heteroscedastic models, a minimum of data can be collected in a given configuration—just enough to perform some intended analysis—which would permit even further efficiency gains.

6.3 Machine Learning to Improve Performance of Experiments

A very straightforward application of machine learning to ARPES will be in extending the limits of performance of the technique. Spectral denoising (*170*), the removal of aberrations from the spectrometer, the removal of transients like decaying detector gain, and the automatic treatment of early data preparation steps will all benefit immensely from the integration of machine learning and automated analysis tooling. Challenges here are primarily political rather than technical, because few places aside from user facilities place much interest in improving these details, but careful treatment of most of these topics will require a much larger expense of effort and time to bring to fruition than the smoothing techniques ARPES scientists are currently used to.

6.4 Domain Specific Improvements in Analysis Tools

NanoARPES, pump-probe ARPES, ARPEEM, and ARPES conducted with in situ control of electrostatic doping or sample strain are each properly new branches under a broader experimental tree. There's a common core of analysis techniques invoked in each—data preparation, visualization, curve fitting, and momentum conversion—but already scientists are developing approaches to analysis to pick loftier fruit from each branch. Our spectral unmixing experiments can be seen as a way of leveraging intersectional properties of a technique

(nanoARPES) and a material class (domain structured materials). NanoARPES will probably see the most ML application in the short term because it encourages the highest data volume and encourages sparse acquisition. Electrostatic control of doping places more interpretational challenges on measurements because it imposes new kinds of measurement defects: degraded resolution and photoelectron lensing. Although not necessarily ML driven, I am confident we will see analysis methods (or possibly hemispherical lenses) evolve to better handle these challenges.

7 References

1. H. Hertz, Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung. *Ann. Phys. Chem.* **267**, 983–1000 (1887).

2. A. Einstein, Über einem die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*. **322**, 132–148 (1905).

3. R. A. Millikan, A Direct Photoelectric Determination of Planck's h. *Phys. Rev.* **7**, 355–388 (1916).

4. C. Nordling, E. Sokolowski, K. Siegbahn, Precision Method for Obtaining Absolute Values of Atomic Binding Energies. *Phys. Rev.* **105**, 1676–1677 (1957).

5. C. N. Berglund, W. E. Spicer, Photoemission Studies of Copper and Silver: Experiment. *Phys. Rev.* **136**, A1044–A1064 (1964).

6. D. Lu, I. M. Vishik, M. Yi, Y. Chen, R. G. Moore, Z.-X. Shen, Angle-Resolved Photoemission Studies of Quantum Materials. *Annual Review of Condensed Matter Physics*. **3**, 129–167 (2012).

7. J. A. Sobota, Y. He, Z.-X. Shen, Angle-resolved photoemission studies of quantum materials. *Rev. Mod. Phys.* **93**, 025006 (2021).

8. A. Damascelli, Probing the Electronic Structure of Complex Systems by ARPES. *Phys. Scr.* **2004**, 61 (2004).

9. S. Hüfner, R. Claessen, F. Reinert, T. Straub, V. N. Strocov, P. Steiner, Photoemission spectroscopy in metals:: band structure-Fermi surface–spectral function. *Journal of Electron Spectroscopy and Related Phenomena*. **100**, 191–213 (1999).

10. F. J. Himpsel, Angle-resolved measurements of the photoemission of electrons in the study of solids. *Advances in Physics*. **32**, 1–51 (1983).

11. C. Jozwiak, Y. L. Chen, A. V. Fedorov, J. G. Analytis, C. R. Rotundu, A. K. Schmid, J. D. Denlinger, Y.-D. Chuang, D.-H. Lee, I. R. Fisher, R. J. Birgeneau, Z.-X. Shen, Z. Hussain, A. Lanzara, Widespread spin polarization effects in photoemission from topological insulators. *Phys. Rev. B.* **84**, 165113 (2011).

12. Y. Cao, J. A. Waugh, X.-W. Zhang, J.-W. Luo, Q. Wang, T. J. Reber, S. K. Mo, Z. Xu, A. Yang, J. Schneeloch, G. D. Gu, M. Brahlek, N. Bansal, S. Oh, A. Zunger, D. S. Dessau, Mapping the orbital wavefunction of the surface states in three-dimensional topological insulators. *Nature Phys*. **9**, 499–504 (2013).

13. T. J. Nummy, J. A. Waugh, S. P. Parham, Q. Liu, H.-Y. Yang, H. Li, X. Zhou, N. C. Plumb, F. F. Tafti, D. S. Dessau, Measurement of the atomic orbital composition of the near-fermi-

level electronic states in the lanthanum monopnictides LaBi, LaSb, and LaAs. *npj Quant Mater*. **3**, 1–6 (2018).

14. A. Bostwick, E. Rotenberg, J. Avila, M. C. Asensio, Zooming in on Electronic Structure: NanoARPES at SOLEIL and ALS. *Synchrotron Radiation News*. **25**, 19–25 (2012).

15. C. Stansbury, A. Lanzara, PyARPES: An analysis framework for multimodal angle-resolved photoemission spectroscopies. *SoftwareX*. **11**, 100472 (2020).

16. J. L. Wiza, Microchannel Plate Detectors. *Nucl. Instrum. Methods*. **162**, 60–1 (1979).

17. M. Lampton, O. Siegmund, R. Raffanti, Delay line anodes for microchannel-plate spectrometers. *Review of Scientific Instruments*. **58**, 2298–2305 (1987).

18. T. J. Reber, N. C. Plumb, J. A. Waugh, D. S. Dessau, Effects, determination, and correction of count rate nonlinearity in multi-channel analog electron detectors. *Review of Scientific Instruments*. **85**, 043907 (2014).

19. N. Mannella, S. Marchesini, A. W. Kay, A. Nambu, T. Gresch, S.-H. Yang, B. S. Mun, J. M. Bussat, A. Rosenhahn, C. S. Fadley, Correction of non-linearity effects in detectors for electron spectroscopy. *Journal of Electron Spectroscopy and Related Phenomena*. **141**, 45–59 (2004).

20. X. Urbain, D. Bech, J.-P. Van Roy, M. Géléoc, S. J. Weber, A. Huetz, Y. J. Picard, A zero dead-time multi-particle time and position sensitive detector based on correlation between brightness and amplitude. *Review of Scientific Instruments*. **86**, 023305 (2015).

21. B. Winter, S. J. King, M. Brouard, C. Vallance, A fast microchannel plate-scintillator detector for velocity map imaging and imaging mass spectrometry. *Review of Scientific Instruments*. **85**, 023306 (2014).

22. J. D. Koralek, J. F. Douglas, N. C. Plumb, Z. Sun, A. V. Fedorov, M. M. Murnane, H. C. Kapteyn, S. T. Cundiff, Y. Aiura, K. Oka, H. Eisaki, D. S. Dessau, Laser Based Angle-Resolved Photoemission, the Sudden Approximation, and Quasiparticle-Like Spectral Peaks in Bi2Sr2CaCu2O(8+δ). *Phys. Rev. Lett.* **96**, 017005 (2006).

23. X. Zhou, S. He, G. Liu, L. Zhao, L. Yu, W. Zhang, New developments in laser-based photoemission spectroscopy and its scientific applications: a key issues review. *Rep. Prog. Phys.* **81**, 062101 (2018).

24. Z. Yong, W. Gui-Ling, L. Cheng-Ming, P. Qin-Jun, C. Da-Fu, X. Zu-Yan, W. Xiao-Yang, Z. Yong, C. Chuang-Tian, L. Guo-Dong, D. Xiao-Li, Z. Xing-Jiang, Sixth Harmonic of A Nd:YVO 4 Laser Generation In KBBF for ARPES. *Chinese Phys. Lett.* **25**, 963–965 (2008).

25. S. V. Borisenko, "One-cubed" ARPES User Facility at BESSY II. *Synchrotron Radiation News*. **25**, 6–11 (2012).

26. D. Roy, D. Tremblay, Design of electron spectrometers. *Rep. Prog. Phys.* **53**, 1621–1674 (1990).

27. P. V. Nguyen, N. C. Teutsch, N. P. Wilson, J. Kahn, X. Xia, A. J. Graham, V. Kandyba, A. Giampietri, A. Barinov, G. C. Constantinescu, N. Yeung, N. D. M. Hine, X. Xu, D. H. Cobden, N. R. Wilson, Visualizing electrostatic gating effects in two-dimensional heterostructures. *Nature*. **572**, 220–223 (2019).

28. C.-Y. Lin, L. Moreschini, A. Lanzara, Present and future trends in spin ARPES. *EPL*. **134**, 57001 (2021).

29. J. Pendry, Theory of photoemission. *Surface Science*. **57**, 679–705 (1976).

30. S. Moser, An experimentalist's guide to the matrix element in angle resolved photoemission. *Journal of Electron Spectroscopy and Related Phenomena*. **214**, 29–52 (2017).

31. M. P. Seah, W. A. Dench, Quantitative electron spectroscopy of surfaces: A standard data base for electron inelastic mean free paths in solids. *Surface and Interface Analysis*. **1**, 2–11 (1979).

32. S. Borisenko, Q. Gibson, D. Evtushinsky, V. Zabolotnyy, B. Büchner, R. J. Cava, Experimental Realization of a Three-Dimensional Dirac Semimetal. *Phys. Rev. Lett.* **113**, 027603 (2014).

33. Z. K. Liu, B. Zhou, Y. Zhang, Z. J. Wang, H. M. Weng, D. Prabhakaran, S.-K. Mo, Z. X. Shen, Z. Fang, X. Dai, Z. Hussain, Y. L. Chen, Discovery of a Three-Dimensional Topological Dirac Semimetal, Na3Bi. *Science*. **343**, 864–867 (2014).

34. B. Q. Lv, H. M. Weng, B. B. Fu, X. P. Wang, H. Miao, J. Ma, P. Richard, X. C. Huang, L. X. Zhao, G. F. Chen, Z. Fang, X. Dai, T. Qian, H. Ding, Experimental Discovery of Weyl Semimetal TaAs. *Phys. Rev. X*. **5**, 031013 (2015).

35. D. S. Sanchez, I. Belopolski, T. A. Cochran, X. Xu, J.-X. Yin, G. Chang, W. Xie, K. Manna, V. Süß, C.-Y. Huang, N. Alidoust, D. Multer, S. S. Zhang, N. Shumiya, X. Wang, G.-Q. Wang, T.-R. Chang, C. Felser, S.-Y. Xu, S. Jia, H. Lin, M. Z. Hasan, Topological chiral crystals with helicoid-arc quantum states. *Nature*. **567**, 500–505 (2019).

36. A. Bostwick, T. Ohta, T. Seyller, K. Horn, E. Rotenberg, Quasiparticle dynamics in graphene. *Nature Phys*. **3**, 36–40 (2007).

37. S. Y. Zhou, G.-H. Gweon, A. V. Fedorov, P. N. First, W. A. de Heer, D.-H. Lee, F. Guinea, A. H. Castro Neto, A. Lanzara, Substrate-induced bandgap opening in epitaxial graphene. *Nature Mater*. **6**, 770–775 (2007).

38. I. Tanabe, M. Gomez, W. C. Coley, D. Le, E. M. Echeverria, G. Stecklein, V. Kandyba, S. K. Balijepalli, V. Klee, A. E. Nguyen, E. Preciado, I.-H. Lu, S. Bobek, D. Barroso, D. Martinez-Ta, A. Barinov, T. S. Rahman, P. A. Dowben, P. A. Crowell, L. Bartels, Band

structure characterization of WS2 grown by chemical vapor deposition. *Appl. Phys. Lett.* **108**, 252103 (2016).

39. P. Dudin, P. Lacovig, C. Fava, E. Nicolini, A. Bianco, G. Cautero, A. Barinov, Angle-resolved photoemission spectroscopy and imaging with a submicrometre probe at the SPECTROMICROSCOPY-3.2L beamline of Elettra. *Journal of Synchrotron Radiation*. **17**, 445–450 (2010).

40. B. Rösner, P. Dudin, J. Bosgra, M. Hoesch, C. David, Zone plates for angle-resolved photoelectron spectroscopy providing sub-micrometre resolution in the extreme ultraviolet regime. *J Synchrotron Rad*. **26**, 467–472 (2019).

41. S. Ciocys, T. Morimoto, J. E. Moore, A. Lanzara, Tracking surface photovoltage dipole geometry in Bi2Se3 with time-resolved photoemission. *J. Stat. Mech.* **2019**, 104008 (2019).

42. E. Frantzeskakis, S. V. Ramankutty, N. de Jong, Y. K. Huang, Y. Pan, A. Tytarenko, M. Radovic, N. C. Plumb, M. Shi, A. Varykhalov, A. de Visser, E. van Heumen, M. S. Golden, Trigger of the Ubiquitous Surface Band Bending in 3D Topological Insulators. *Phys. Rev. X*. **7**, 041041 (2017).

43. A. A. Kordyuk, T. K. Kim, V. B. Zabolotnyy, D. V. Evtushinsky, M. Bauch, C. Hess, B. Büchner, H. Berger, S. V. Borisenko, Photoemission-induced gating of topological insulators. *Phys. Rev. B*. **83**, 081303 (2011).

44. E. Rotenberg, A. Bostwick, microARPES and nanoARPES at diffraction-limited light sources: opportunities and performance gains. *Journal of Synchrotron Radiation*. **21**, 1048–1056 (2014).

45. M. Aeschlimann, C. A. Schmuttenmaer, H. E. Elsayed-Ali, R. J. D. Miller, J. Cao, Y. Gao, D. A. Mantell, Observation of surface enhanced multiphoton photoemission from metal surfaces in the short pulse limit. *J. Chem. Phys.* **102**, 8606–8613 (1995).

46. L. Miaja-Avila, C. Lei, M. Aeschlimann, J. L. Gland, M. M. Murnane, H. C. Kapteyn, G. Saathoff, Laser-Assisted Photoelectric Effect from Surfaces. *Phys. Rev. Lett.* **97**, 113604 (2006).

47. P. D. C. King, J. A. Rosen, W. Meevasana, A. Tamai, E. Rozbicki, R. Comin, G. Levy, D. Fournier, Y. Yoshida, H. Eisaki, K. M. Shen, N. J. C. Ingle, A. Damascelli, F. Baumberger, Structural Origin of Apparent Fermi Surface Pockets in Angle-Resolved Photoemission of Bi2Sr(2-x)La(x)CuO(6+δ). *Phys. Rev. Lett.* **106**, 127005 (2011).

48. S. V. Borisenko, M. S. Golden, S. Legner, T. Pichler, C. Dürr, M. Knupfer, J. Fink, G. Yang, S. Abell, H. Berger, Joys and Pitfalls of Fermi Surface Mapping in ${\mathrm{Bi}}_{2}{\mathrm{Sr}}_{2}{\mathrm{CaCu}}_{2}{O}_{8+\ensuremath{\delta}}$ Using Angle Resolved Photoemission. *Phys. Rev. Lett.* **84**, 4453–4456 (2000).

49. C. M. Polley, L. I. Johansson, H. Fedderwitz, T. Balasubramanian, M. Leandersson, J. Adell, R. Yakimova, C. Jacobi, Origin of the π band replicas in the electronic structure of graphene grown on 4H-SiC(0001). *Phys. Rev. B.* **99**, 115404 (2019).

50. M. Sprinkle, D. Siegel, Y. Hu, J. Hicks, A. Tejeda, A. Taleb-Ibrahimi, P. Le Fèvre, F. Bertran, S. Vizzini, H. Enriquez, S. Chiang, P. Soukiassian, C. Berger, W. A. de Heer, A. Lanzara, E. H. Conrad, First Direct Observation of a Nearly Ideal Graphene Band Structure. *Phys. Rev. Lett.* **103**, 226803 (2009).

51. C. Kastl, C. T. Chen, R. J. Koch, B. Schuler, T. R. Kuykendall, A. Bostwick, C. Jozwiak, T. Seyller, E. Rotenberg, A. Weber-Bargioni, S. Aloni, A. M. Schwartzberg, Multimodal spectromicroscopy of monolayer WS2 enabled by ultra-clean van der Waals epitaxy. *2D Materials*. **5**, 045010 (2018).

52. A. J. H. Jones, R. Muzzio, P. Majchrzak, S. Pakdel, D. Curcio, K. Volckaert, D. Biswas, J. Gobbo, S. Singh, J. T. Robinson, K. Watanabe, T. Taniguchi, T. K. Kim, C. Cacho, N. Lanata, J. A. Miwa, P. Hofmann, J. Katoch, S. Ulstrup, Observation of Electrically Tunable van Hove Singularities in Twisted Bilayer Graphene from NanoARPES. *Advanced Materials*. **32**, 2001656 (2020).

53. N. Dale, R. Mori, M. I. B. Utama, J. D. Denlinger, C. Stansbury, C. G. Fatuzzo, S. Zhao, K. Lee, T. Taniguchi, K. Watanabe, C. Jozwiak, A. Bostwick, E. Rotenberg, R. J. Koch, F. Wang, A. Lanzara, Correlation-driven electron-hole asymmetry in graphene field effect devices. *npj Quantum Mater.* **7**, 1–7 (2022).

54. F. Joucken, J. Avila, Z. Ge, E. A. Quezada-Lopez, H. Yi, R. Le Goff, E. Baudin, J. L. Davenport, K. Watanabe, T. Taniguchi, M. C. Asensio, J. Velasco, Visualizing the Effect of an Electrostatic Gate with Angle-Resolved Photoemission Spectroscopy. *Nano Lett.* **19**, 2682–2687 (2019).

55. R. Mori, P. B. Marshall, K. Ahadi, J. D. Denlinger, S. Stemmer, A. Lanzara, Controlling a Van Hove singularity and Fermi surface topology at a complex oxide heterostructure interface. *Nat Commun*. **10**, 5534 (2019).

56. A. X. Gray, J. Minár, L. Plucinski, M. Huijben, A. Bostwick, E. Rotenberg, S.-H. Yang, J. Braun, A. Winkelmann, G. Conti, D. Eiteneer, A. Rattanachata, A. A. Greer, J. Ciston, C. Ophus, G. Rijnders, D. H. A. Blank, D. Doennig, R. Pentcheva, J. B. Kortright, C. M. Schneider, H. Ebert, C. S. Fadley, Momentum-resolved electronic structure at a buried interface from soft X-ray standing-wave angle-resolved photoemission. *EPL*. **104**, 17004 (2013).

57. V. N. Strocov, L. L. Lev, M. Kobayashi, C. Cancellieri, M.-A. Husanu, A. Chikina, N. B. M. Schröter, X. Wang, J. A. Krieger, Z. Salman, k-resolved electronic structure of buried heterostructure and impurity systems by soft-X-ray ARPES. *Journal of Electron Spectroscopy and Related Phenomena*. **236**, 1–8 (2019).

58. C. H. Stansbury, M. I. B. Utama, C. G. Fatuzzo, E. C. Regan, D. Wang, Z. Xiang, M. Ding, K. Watanabe, T. Taniguchi, M. Blei, Y. Shen, S. Lorcy, A. Bostwick, C. Jozwiak, R. Koch, S. Tongay, J. Avila, E. Rotenberg, F. Wang, A. Lanzara, Visualizing electron localization of WS2/WSe2 moiré superlattices in momentum space. *Science Advances*. **7**, eabf4387.

59. M. I. B. Utama, R. J. Koch, K. Lee, N. Leconte, H. Li, S. Zhao, L. Jiang, J. Zhu, K. Watanabe, T. Taniguchi, P. D. Ashby, A. Weber-Bargioni, A. Zettl, C. Jozwiak, J. Jung, E. Rotenberg, A. Bostwick, F. Wang, Visualization of the flat electronic band in twisted bilayer graphene near the magic angle twist. *Nature Physics*, 1–5 (2020).

60. S. Lisi, X. Lu, T. Benschop, T. A. de Jong, P. Stepanov, J. R. Duran, F. Margot, I. Cucchi, E. Cappelli, A. Hunter, A. Tamai, V. Kandyba, A. Giampietri, A. Barinov, J. Jobst, V. Stalman, M. Leeuwenhoek, K. Watanabe, T. Taniguchi, L. Rademaker, S. J. van der Molen, M. P. Allan, D. K. Efetov, F. Baumberger, Observation of flat bands in twisted bilayer graphene. *Nature Physics*. **17**, 189–193 (2021).

61. A. Kaminski, H. M. Fretwell, On the extraction of the self-energy from angle-resolved photoemission spectroscopy. *New J. Phys.* **7**, 98–98 (2005).

62. A. Lanzara, P. V. Bogdanov, X. J. Zhou, S. A. Kellar, D. L. Feng, E. D. Lu, T. Yoshida, H. Eisaki, A. Fujimori, K. Kishio, J.-I. Shimoyama, T. Noda, S. Uchida, Z. Hussain, Z.-X. Shen, Evidence for ubiquitous strong electron–phonon coupling in high-temperature superconductors. *Nature*. **412**, 510–514 (2001).

63. T. P. Devereaux, T. Cuk, Z.-X. Shen, N. Nagaosa, Anisotropic Electron-Phonon Interaction in the Cuprates. *Phys. Rev. Lett.* **93**, 117004 (2004).

64. I. Horcas, R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, A. M. Baro, WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Review of Scientific Instruments*. **78**, 013705 (2007).

65. P. Zahl, M. Bierkandt, S. Schröder, A. Klust, The flexible and modern open source scanning probe microscopy software package GXSM. *Review of Scientific Instruments*. **74**, 1222–1227 (2003).

66. D. Nečas, P. Klapetek, Gwyddion: an open-source software for SPM data analysis. *Open Physics*. **10**, 181–188 (2012).

67. R. F. Laine, K. L. Tosheva, N. Gustafsson, R. D. M. Gray, P. Almada, D. Albrecht, G. T. Risa, F. Hurtig, A.-C. Lindås, B. Baum, J. Mercer, C. Leterrier, P. M. Pereira, S. Culley, R. Henriques, NanoJ: a high-performance open-source super-resolution microscopy toolbox. *J. Phys. D: Appl. Phys.* **52**, 163001 (2019).

68. M. Basham, J. Filik, M. T. Wharmby, P. C. Y. Chang, B. El Kassaby, M. Gerring, J. Aishima, K. Levik, B. C. A. Pulford, I. Sikharulidze, D. Sneddon, M. Webber, S. S. Dhesi, F. Maccherozzi, O. Svensson, S. Brockhauser, G. Náray, A. W. Ashton, Data Analysis WorkbeNch (DAWN). *J Synchrotron Rad*. **22**, 853–858 (2015).

69. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python. *PeerJ*. **2**, e453 (2014).

70. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. **17**, 261–272 (2020).

71. C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy. *Nature*. **585**, 357–362 (2020).

72. S. Hoyer, J. Hamman, xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*. **5**, 10 (2017).

73. A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Du\lak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter*. **29**, 273002 (2017).

74. S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*. **68**, 314–319 (2013).

75. S. K. Lam, A. Pitrou, S. Seibert, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* (Association for Computing Machinery, New York, NY, USA, 2015; https://doi.org/10.1145/2833157.2833162), *LLVM '15*, pp. 1–6.

76. P. Zhang, P. Richard, T. Qian, Y.-M. Xu, X. Dai, H. Ding, A precise method for visualizing dispersive features in image plots. *Review of Scientific Instruments*. **82**, 043712 (2011).

77. Y. He, Y. Wang, Z.-X. Shen, Visualizing dispersive features in 2D image via minimum gradient method. *Review of Scientific Instruments*. **88**, 073903 (2017).

78. J. Graf, S. Hellman, C. Jozwiak, C. L. Smallwood, Z. Hussain, R. A. Kaindl, L. Kipp, K. Rossnagel, A. Lanzara, Vacuum space charge effect in laser-based solid-state photoemission spectroscopy. *Journal of Applied Physics*. **107, 014912** (2010) (available at https://aip.scitation.org/doi/10.1063/1.3273487).

79. V. Lollobrigida, G. Greco, D. Simeone, F. Offi, A. Verna, G. Stefani, Electron trajectory simulations of time-of-flight spectrometers for core level high-energy photoelectron spectroscopy at pulsed X-ray sources. *Journal of Electron Spectroscopy and Related Phenomena*. **205**, 98–105 (2015).

80. A. Verna, G. Greco, V. Lollobrigida, F. Offi, G. Stefani, Space-charge effects in high-energy photoemission. *Journal of Electron Spectroscopy and Related Phenomena*. **209**, 14–25 (2016).

81. A. A. Kordyuk, S. V. Borisenko, M. Knupfer, J. Fink, Measuring the gap in angle-resolved photoemission experiments on cuprates. *Phys. Rev. B.* **67**, 064504 (2003).

82. C. N. Melton, M. M. Noack, T. Ohta, T. E. Beechem, J. Robinson, X. Zhang, A. Bostwick, C. Jozwiak, R. J. Koch, P. H. Zwart, A. Hexemer, E. Rotenberg, K-means-driven Gaussian Process data collection for angle-resolved photoemission spectroscopy. *Mach. Learn.: Sci. Technol.* **1**, 045015 (2020).

83. Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, P. Jarillo-Herrero, Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature*. **556**, 80–84 (2018).

84. Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices. *Nature*. **556**, 43–50 (2018).

85. K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, Electric Field Effect in Atomically Thin Carbon Films. *Science*. **306**, 666–669 (2004).

86. E. McCann, M. Koshino, The electronic properties of bilayer graphene. *Rep. Prog. Phys.* **76**, 056503 (2013).

87. T. Ohta, A. Bostwick, T. Seyller, K. Horn, E. Rotenberg, Controlling the Electronic Structure of Bilayer Graphene. *Science*. **313**, 951–954 (2006).

88. Y. Zhang, T.-T. Tang, C. Girit, Z. Hao, M. C. Martin, A. Zettl, M. F. Crommie, Y. R. Shen, F. Wang, Direct observation of a widely tunable bandgap in bilayer graphene. *Nature*. **459**, 820–823 (2009).

89. R. W. Havener, H. Zhuang, L. Brown, R. G. Hennig, J. Park, Angle-Resolved Raman Imaging of Interlayer Rotations and Interactions in Twisted Bilayer Graphene. *Nano Lett.* **12**, 3162–3167 (2012).

90. W. Yao, E. Wang, C. Bao, Y. Zhang, K. Zhang, K. Bao, C. K. Chan, C. Chen, J. Avila, M. C. Asensio, J. Zhu, S. Zhou, Quasicrystalline 30° twisted bilayer graphene as an incommensurate superlattice with strong interlayer coupling. *PNAS*. **115**, 6928–6933 (2018).

91. T. Ohta, J. T. Robinson, P. J. Feibelman, A. Bostwick, E. Rotenberg, T. E. Beechem, Evidence for Interlayer Coupling and Moir\'e Periodic Potentials in Twisted Bilayer Graphene. *Phys. Rev. Lett.* **109**, 186807 (2012).

92. G. Tarnopolsky, A. J. Kruchkov, A. Vishwanath, Origin of Magic Angles in Twisted Bilayer Graphene. *Phys. Rev. Lett.* **122**, 106405 (2019).

93. L. Wang, E.-M. Shih, A. Ghiotto, L. Xian, D. A. Rhodes, C. Tan, M. Claassen, D. M. Kennes, Y. Bai, B. Kim, K. Watanabe, T. Taniguchi, X. Zhu, J. Hone, A. Rubio, A. N. Pasupathy, C. R. Dean, Correlated electronic phases in twisted bilayer transition metal dichalcogenides. *Nat. Mater.* **19**, 861–866 (2020).

94. M. H. Naik, M. Jain, Ultraflatbands and Shear Solitons in Moiré Patterns of Twisted Bilayer Transition Metal Dichalcogenides. *Phys. Rev. Lett.* **121**, 266401 (2018).

95. L. An, X. Cai, D. Pei, M. Huang, Z. Wu, Z. Zhou, J. Lin, Z. Ying, Z. Ye, X. Feng, R. Gao, C. Cacho, M. Watson, Y. Chen, N. Wang, Interaction effects and superconductivity signatures in twisted double-bilayer WSe 2. *Nanoscale Horizons*. **5**, 1309–1316 (2020).

96. C. Jin, J. Kim, M. I. B. Utama, E. C. Regan, H. Kleemann, H. Cai, Y. Shen, M. J. Shinner, A. Sengupta, K. Watanabe, T. Taniguchi, S. Tongay, A. Zettl, F. Wang, Imaging of pure spin-valley diffusion current in WS2-WSe2 heterostructures. *Science*. **360**, 893–896 (2018).

97. J. N. Coleman, M. Lotya, A. O'Neill, S. D. Bergin, P. J. King, U. Khan, K. Young, A. Gaucher, S. De, R. J. Smith, I. V. Shvets, S. K. Arora, G. Stanton, H.-Y. Kim, K. Lee, G. T. Kim, G. S. Duesberg, T. Hallam, J. J. Boland, J. J. Wang, J. F. Donegan, J. C. Grunlan, G. Moriarty, A. Shmeliov, R. J. Nicholls, J. M. Perkins, E. M. Grieveson, K. Theuwissen, D. W. McComb, P. D. Nellist, V. Nicolosi, Two-Dimensional Nanosheets Produced by Liquid Exfoliation of Layered Materials. *Science* (2011), doi:10.1126/science.1194975.

98. C. Lee, H. Yan, L. E. Brus, T. F. Heinz, J. Hone, S. Ryu, Anomalous Lattice Vibrations of Single- and Few-Layer MoS2. *ACS Nano*. **4**, 2695–2700 (2010).

99. K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, A. K. Geim, Two-dimensional atomic crystals. *PNAS*. **102**, 10451–10453 (2005).

100. A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, F. Wang, Emerging Photoluminescence in Monolayer MoS2. *Nano Lett.* **10**, 1271–1275 (2010).

101. K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, Atomically Thin MoS2: A New Direct-Gap Semiconductor. *Phys. Rev. Lett.* **105**, 136805 (2010).

102. A. Kuc, N. Zibouche, T. Heine, Influence of quantum confinement on the electronic structure of the transition metal sulfide TS2. *Phys. Rev. B*. **83**, 245213 (2011).

103. D. Le, A. Barinov, E. Preciado, M. Isarraraz, I. Tanabe, T. Komesu, C. Troha, L. Bartels, T. S. Rahman, P. A. Dowben, Spin–orbit coupling in the band structure of monolayer WSe2. *J. Phys.: Condens. Matter*. **27**, 182201 (2015).

104. Y.-T. Hsu, A. Vaezi, M. H. Fischer, E.-A. Kim, Topological superconductivity in monolayer transition metal dichalcogenides. *Nat Commun*. **8**, 14985 (2017).

105. J. M. Lu, O. Zheliuk, I. Leermakers, N. F. Q. Yuan, U. Zeitler, K. T. Law, J. T. Ye, Evidence for two-dimensional Ising superconductivity in gated MoS2. *Science* (2015).

106. A. Kormányos, G. Burkard, M. Gmitra, J. Fabian, V. Zólyomi, N. D. Drummond, V. Fal'ko, k · p theory for two-dimensional transition metal dichalcogenide semiconductors. *2D Mater.* **2**, 022001 (2015).

107. N. R. Wilson, P. V. Nguyen, K. Seyler, P. Rivera, A. J. Marsden, Z. P. L. Laker, G. C. Constantinescu, V. Kandyba, A. Barinov, N. D. M. Hine, X. Xu, D. H. Cobden, Determination of band offsets, hybridization, and exciton binding in 2D semiconductor heterostructures. *Science Advances*. **3**, e1601832 (2017).

108. A. J. H. Jones, R. Muzzio, S. Pakdel, D. Biswas, D. Curcio, N. Lanatà, P. Hofmann, K. M. McCreary, B. T. Jonker, K. Watanabe, T. Taniguchi, S. Singh, R. J. Koch, C. Jozwiak, E. Rotenberg, A. Bostwick, J. A. Miwa, J. Katoch, S. Ulstrup, Visualizing band structure hybridization and superlattice effects in twisted MoS2/WS2 heterobilayers. *arXiv:2106.00403 [cond-mat]* (2021).

109. S. Ulstrup, R. J. Koch, S. Singh, K. M. McCreary, B. T. Jonker, J. T. Robinson, C. Jozwiak, E. Rotenberg, A. Bostwick, J. Katoch, J. A. Miwa, Direct observation of minibands in a twisted graphene/WS2 bilayer. *Science Advances*. **6**, eaay6104 (2020).

110. S. Xie, B. D. Faeth, Y. Tang, L. Li, C. T. Parzyck, D. Chowdhury, Y.-H. Zhang, C. Jozwiak, A. Bostwick, E. Rotenberg, J. Shan, K. F. Mak, K. M. Shen, Direct observation of distinct minibands in moiré superlattices, 18.

111. L. Waldecker, A. Raja, M. Rösner, C. Steinke, A. Bostwick, R. J. Koch, C. Jozwiak, T. Taniguchi, K. Watanabe, E. Rotenberg, T. O. Wehling, T. F. Heinz, Rigid Band Shifts in Two-Dimensional Semiconductors through External Dielectric Screening. *Phys. Rev. Lett.* **123**, 206403 (2019).

112. C. Kastl, R. J. Koch, C. T. Chen, J. Eichhorn, S. Ulstrup, A. Bostwick, C. Jozwiak, T. R. Kuykendall, N. J. Borys, F. M. Toma, S. Aloni, A. Weber-Bargioni, E. Rotenberg, A. M. Schwartzberg, Effects of Defects on Band Structure and Excitons in WS2 Revealed by Nanoscale Photoemission Spectroscopy. *ACS Nano*. **13**, 1284–1291 (2019).

113. P.-C. Yeh, W. Jin, N. Zaki, D. Zhang, J. T. Liou, J. T. Sadowski, A. Al-Mahboob, J. I. Dadap, I. P. Herman, P. Sutter, R. M. Osgood, Layer-dependent electronic structure of an atomically heavy two-dimensional dichalcogenide. *Phys. Rev. B.* **91**, 041407 (2015).

114. D. W. Latzke, W. Zhang, A. Suslu, T.-R. Chang, H. Lin, H.-T. Jeng, S. Tongay, J. Wu, A. Bansil, A. Lanzara, Electronic structure, spin-orbit coupling, and interlayer interaction in bulk MoS2 and WS2. *Phys. Rev. B.* **91**, 235202 (2015).

115. G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, D. Xiao, Three-band tight-binding model for monolayers of group-VIB transition metal dichalcogenides. *Phys. Rev. B.* **88**, 085433 (2013).

116. A. Säynätjoki, L. Karvonen, H. Rostami, A. Autere, S. Mehravar, A. Lombardo, R. A. Norwood, T. Hasan, N. Peyghambarian, H. Lipsanen, K. Kieu, A. C. Ferrari, M. Polini, Z. Sun, Ultra-strong nonlinear optical processes and trigonal warping in MoS2 layers. *Nat Commun*. **8**, 893 (2017).

117. W. Chen, X. Zhou, P. Liu, X. Xiao, G. Zhou, Effect of trigonal warping on the Berry curvature and valley/spin Hall effects in monolayer MoS2. *Physics Letters A.* **384**, 126344 (2020).

118. C. Ernandes, L. Khalil, H. Henck, M.-Q. Zhao, J. Chaste, F. Oehler, A. T. C. Johnson, M. C. Asensio, D. Pierucci, M. Pala, J. Avila, A. Ouerghi, Strain and Spin-Orbit Coupling Engineering in Twisted WS2/Graphene Heterobilayer. *Nanomaterials*. **11**, 2921 (2021).

119. G.-B. Liu, D. Xiao, Y. Yao, X. Xu, W. Yao, Electronic structures and theoretical modelling of two-dimensional group-VIB transition metal dichalcogenides. *Chem. Soc. Rev.* **44**, 2643–2663 (2015).

120. S. Fang, R. Kuate Defo, S. N. Shirodkar, S. Lieu, G. A. Tritsaris, E. Kaxiras, Ab initio tight-binding Hamiltonian for transition metal dichalcogenides. *Phys. Rev. B.* **92**, 205108 (2015).

121. J. Gusakova, X. Wang, L. L. Shiau, A. Krivosheeva, V. Shaposhnikov, V. Borisenko, V. Gusakov, B. K. Tay, Electronic Properties of Bulk and Monolayer TMDs: Theoretical Study Within DFT Framework (GVJ-2e Method). *physica status solidi (a)*. **214**, 1700218 (2017).

122. J. Á. Silva-Guillén, P. San-Jose, R. Roldán, Electronic Band Structure of Transition Metal Dichalcogenides from Ab Initio and Slater–Koster Tight-Binding Model. *Applied Sciences*. **6**, 284 (2016).

123. M. H. Naik, M. Jain, Origin of layer dependence in band structures of two-dimensional materials. *Phys. Rev. B.* **95**, 165125 (2017).

124. H. Nakamura, A. Mohammed, Ph. Rosenzweig, K. Matsuda, K. Nowakowski, K. Küster, P. Wochner, S. Ibrahimkutty, U. Wedig, H. Hussain, J. Rawle, C. Nicklin, B. Stuhlhofer, G. Cristiani, G. Logvenov, H. Takagi, U. Starke, Spin splitting and strain in epitaxial monolayer WSe 2 on graphene. *Phys. Rev. B.* **101**, 165103 (2020).

125. Y.-C. Lin, C.-Y. S. Chang, R. K. Ghosh, J. Li, H. Zhu, R. Addou, B. Diaconescu, T. Ohta, X. Peng, N. Lu, M. J. Kim, J. T. Robinson, R. M. Wallace, T. S. Mayer, S. Datta, L.-J.

Li, J. A. Robinson, Atomically Thin Heterostructures Based on Single-Layer Tungsten Diselenide and Graphene. *Nano Lett.* **14**, 6936–6941 (2014).

126. C. Zhang, C.-P. Chuu, X. Ren, M.-Y. Li, L.-J. Li, C. Jin, M.-Y. Chou, C.-K. Shih, Interlayer couplings, Moiré patterns, and 2D electronic superlattices in MoS2/WSe2 hetero-bilayers. *Science Advances*. **3**, e1601459 (2017).

127. S. Shabani, D. Halbertal, W. Wu, M. Chen, S. Liu, J. Hone, W. Yao, D. N. Basov, X. Zhu, A. N. Pasupathy, Deep moiré potentials in twisted transition metal dichalcogenide bilayers. *Nat. Phys.* **17**, 720–725 (2021).

128. S. Brem, C. Linderälv, P. Erhart, E. Malic, Tunable Phases of Moiré Excitons in van der Waals Heterostructures. *Nano Lett.* **20**, 8534–8540 (2020).

129. D. Moldovan, M. Anđelković, F. Peeters, *pybinding v0.9.4: a Python package for tight-binding calculations* (Zenodo, 2017; https://zenodo.org/record/826942#.XsBRFmhKiUk).

130. H. Nishi, Y. Matsushita, A. Oshiyama, Band-unfolding approach to moiré-induced band-gap opening and Fermi level velocity reduction in twisted bilayer graphene. *Phys. Rev. B.* **95**, 085420 (2017).

131. H. Li, S. Li, M. H. Naik, J. Xie, X. Li, J. Wang, E. Regan, D. Wang, W. Zhao, S. Zhao, S. Kahn, K. Yumigeta, M. Blei, T. Taniguchi, K. Watanabe, S. Tongay, A. Zettl, S. G. Louie, F. Wang, M. F. Crommie, Imaging moir\'e flat bands in 3D reconstructed WSe2/WS2 superlattices. *arXiv:2007.06113 [cond-mat]* (2020) (available at http://arxiv.org/abs/2007.06113).

132. A. Weston, Y. Zou, V. Enaldiev, A. Summerfield, N. Clark, V. Zólyomi, A. Graham, C. Yelgel, S. Magorrian, M. Zhou, J. Zultak, D. Hopkinson, A. Barinov, T. H. Bointon, A. Kretinin, N. R. Wilson, P. H. Beton, V. I. Fal'ko, S. J. Haigh, R. Gorbachev, Atomic reconstruction in twisted bilayers of transition metal dichalcogenides. *Nature Nanotechnology*. **15**, 592–597 (2020).

133. E. M. Alexeev, D. A. Ruiz-Tijerina, M. Danovich, M. J. Hamer, D. J. Terry, P. K. Nayak, S. Ahn, S. Pak, J. Lee, J. I. Sohn, M. R. Molas, M. Koperski, K. Watanabe, T. Taniguchi, K. S. Novoselov, R. V. Gorbachev, H. S. Shin, V. I. Fal'ko, A. I. Tartakovskii, Resonantly hybridized excitons in moiré superlattices in van der Waals heterostructures. *Nature*. **567**, 81–86 (2019).

134. C. Jin, E. C. Regan, A. Yan, M. Iqbal Bakti Utama, D. Wang, S. Zhao, Y. Qin, S. Yang, Z. Zheng, S. Shi, K. Watanabe, T. Taniguchi, S. Tongay, A. Zettl, F. Wang, Observation of moiré excitons in WSe 2 /WS 2 heterostructure superlattices. *Nature*. **567**, 76–80 (2019).

135. K. Tran, G. Moody, F. Wu, X. Lu, J. Choi, K. Kim, A. Rai, D. A. Sanchez, J. Quan, A. Singh, J. Embley, A. Zepeda, M. Campbell, T. Autry, T. Taniguchi, K. Watanabe, N. Lu, S. K. Banerjee, K. L. Silverman, S. Kim, E. Tutuc, L. Yang, A. H. MacDonald, X. Li, Evidence for moiré excitons in van der Waals heterostructures. *Nature*. **567**, 71–75 (2019).

136. Mott and generalized Wigner crystal states in WSe 2 /WS 2 moiré superlattices | Nature, (available at https://www.nature.com/articles/s41586-020-2092-4).

137. Y. Tang, L. Li, T. Li, Y. Xu, S. Liu, K. Barmak, K. Watanabe, T. Taniguchi, A. H. MacDonald, J. Shan, K. F. Mak, Simulation of Hubbard model physics in WSe 2 /WS 2 moiré superlattices. *Nature*. **579**, 353–358 (2020).

138. Identification of spin, valley and moiré quasi-angular momentum of interlayer excitons | Nature Physics, (available at https://www.nature.com/articles/s41567-019-0631-4).

139. W. Fengcheng, L. Timothy, A. H. MacDonald, Topological Exciton Bands in Moiré Heterojunctions. *Phys. Rev. Lett.* **118, 147401** (2017).

140. S. Appelfeller, K. Holtgrewe, M. Franz, L. Freter, C. Hassenstein, H.-F. Jirschik, S. Sanna, M. Dähne, Continuous crossover from two-dimensional to one-dimensional electronic properties for metallic silicide nanowires. *Phys. Rev. B.* **102**, 115433 (2020).

141. P. Puschnig, D. Lüftner, Simulation of angle-resolved photoemission spectra by approximating the final state by a plane wave: From graphene to polycyclic aromatic hydrocarbon molecules. *Journal of Electron Spectroscopy and Related Phenomena*. **200**, 193–208 (2015).

142. J. Avila, M. C. Asensio, First NanoARPES User Facility Available at SOLEIL: An Innovative and Powerful Tool for Studying Advanced Materials. *Synchrotron Radiation News*. **27**, 24–30 (2014).

143. M. Cattelan, N. A. Fox, A Perspective on the Application of Spatially Resolved ARPES for 2D Materials. *Nanomaterials*. **8**, 284 (2018).

144. R. J. Koch, C. Jozwiak, A. Bostwick, B. Stripe, M. Cordier, Z. Hussain, W. Yun, E. Rotenberg, Nano focusing of soft X-rays by a new capillary mirror optic. *Synchrotron Radiation News*. **31**, 50–52 (2018).

145. P. Hofmann, Accessing the spectral function of in operando devices by angle-resolved photoemission spectroscopy. *AVS Quantum Sci.* **3**, 021101 (2021).

146. C. Bao, H. Zhang, Q. Li, S. Zhou, H. Zhang, K. Deng, K. Zhang, L. Luo, W. Yao, C. Chen, J. Avila, M. C. Asensio, Y. Wu, S. Zhou, Spatially-resolved electronic structure of stripe domains in IrTe2 through electronic structure microscopy. *Commun Phys*. **4**, 1–6 (2021).

147. M. D. Watson, A. A. Haghighirad, L. C. Rhodes, M. Hoesch, T. K. Kim, Electronic anisotropies revealed by detwinned angle-resolved photo-emission spectroscopy measurements of FeSe. *New J. Phys.* **19**, 103021 (2017).

148. L. C. Rhodes, M. D. Watson, A. A. Haghighirad, D. V. Evtushinsky, T. K. Kim, Revealing the single electron pocket of FeSe in a single orthorhombic domain. *Phys. Rev. B.* **101**, 235128 (2020).

149. S. S. Huh, J. J. Seo, B. S. Kim, S. H. Cho, J. K. Jung, S. Kim, C. I. Kwon, J. S. Kim, Y. Y. Koh, W. S. Kyung, J. D. Denlinger, Y. H. Kim, B. N. Chae, N. D. Kim, Y. K. Kim, C. Kim, Absence of Y-pocket in 1-Fe Brillouin zone and reversed orbital occupation imbalance in FeSe. *Commun Phys*. **3**, 1–7 (2020).

150. C. E. Matt, H. Pirie, A. Soumyanarayanan, Y. He, M. M. Yee, P. Chen, Y. Liu, D. T. Larson, W. S. Paz, J. J. Palacios, M. H. Hamidian, J. E. Hoffman, Consistency between ARPES and STM measurements on ${\mathrm{SmB}}_{6}$. *Phys. Rev. B.* **101**, 085142 (2020).

151. S. Patil, A. Generalov, M. Güttler, P. Kushwaha, A. Chikina, K. Kummer, T. C. Rödel, A. F. Santander-Syro, N. Caroca-Canales, C. Geibel, S. Danzenbächer, Y. Kucherenko, C. Laubschat, J. W. Allen, D. V. Vyalikh, ARPES view on surface and bulk hybridization phenomena in the antiferromagnetic Kondo lattice CeRh2Si2. *Nat Commun*. **7**, 11029 (2016).

152. J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **5**, 354–379 (2012).

153. N. Keshava, A Survey of Spectral Unmixing Algorithms.

154. N. Keshava, J. F. Mustard, Spectral unmixing. *IEEE Signal Processing Magazine*. **19**, 44–57 (2002).

155. Y.-X. Wang, Y.-J. Zhang, Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering*. **25**, 1336–1353 (2013).

156. N. Yokoya, T. Yairi, A. Iwasaki, Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*. **50**, 528–537 (2012).

157. C. Quintano, A. Fernández-Manso, Y. E. Shimabukuro, G. Pereira, Spectral unmixing. *International Journal of Remote Sensing*. **33**, 5307–5340 (2012).

158. W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, C.-Y. Chi, A Signal Processing Perspective on Hyperspectral Unmixing: Insights from Remote Sensing. *IEEE Signal Processing Magazine*. **31**, 67–81 (2014).

159. G. A. Shaw, H. K. Burke, Spectral Imaging for Remote Sensing. **14**, 26 (2003).

160. M. Yi, H. Pfau, Y. Zhang, Y. He, H. Wu, T. Chen, Z. R. Ye, M. Hashimoto, R. Yu, Q. Si, D.-H. Lee, P. Dai, Z.-X. Shen, D. H. Lu, R. J. Birgeneau, Nematic Energy Scale and the Missing Electron Pocket in FeSe. *Phys. Rev. X*. **9**, 041049 (2019).

161. C. J. Long, D. Bunker, X. Li, V. L. Karen, I. Takeuchi, Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization. *Review of Scientific Instruments*. **80** (2009) (available at https://aip.scitation.org/doi/10.1063/1.3216809).

162. R.-F. Cai, M.-T. Chang, S.-C. Lo, C.-C. Chen, Novel spectral unmixing approach for electron energy-loss spectroscopy. *New J. Phys.* **22**, 033029 (2020).

163. J. Bradbury, R. Frostig, P. Hawkins, J. J. Matthew, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, Q. Zhang, JAX: composable transformations of Python + NumPy programs (2018), (available at https://github.com/google/jax).

164. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation: APL Materials: Vol 1, No 1, (available at https://aip.scitation.org/doi/10.1063%2F1.4812323).

165. J. E. Saal, A. O. Oliynyk, B. Meredig, Machine Learning in Materials Discovery: Confirmed Predictions and Their Underlying Approaches. *Annual Review of Materials Research*. **50**, 49–69 (2020).

166. B. Meredig, E. Antono, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hattrick-Simpers, A. Mehta, L. Ward, Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Molecular Systems Design & Engineering*. **3**, 819–825 (2018).

167. R. P. Day, B. Zwartsenberg, I. S. Elfimov, A. Damascelli, Computational framework chinook for angle-resolved photoemission spectroscopy. *npj Quantum Mater.* **4**, 1–10 (2019).

168. R. P. Xian, Y. Acremann, S. Y. Agustsson, M. Dendzik, K. Bühlmann, D. Curcio, D. Kutnyakhov, F. Pressacco, M. Heber, S. Dong, T. Pincelli, J. Demsar, W. Wurth, P. Hofmann, M. Wolf, M. Scheidgen, L. Rettig, R. Ernstorfer, An open-source, end-to-end workflow for multidimensional photoemission spectroscopy. *Sci Data*. **7**, 442 (2020).

169. R. P. Xian, V. Stimper, M. Zacharias, S. Dong, M. Dendzik, S. Beaulieu, B. Schölkopf, M. Wolf, L. Rettig, C. Carbogno, S. Bauer, R. Ernstorfer, A machine learning route between band mapping and band structure. *arXiv:2005.10210 [cond-mat, physics:physics]* (2020) (available at http://arxiv.org/abs/2005.10210).

170. Y. Kim, D. Oh, S. Huh, D. Song, S. Jeong, J. Kwon, M. Kim, D. Kim, H. Ryu, J. Jung, W. Kyung, B. Sohn, S. Lee, J. Hyun, Y. Lee, Y. Kim, C. Kim, Deep learning-based statistical noise reduction for multidimensional spectral data. *Review of Scientific Instruments*. **92**, 073901 (2021).

8 Appendix 1: Practical Reference for Deriving Angle-to-Momentum Transforms with Symbolic Algebra Packages

It took me a long time to track down an older student willing to spare time explaining how the angle-to-momentum conversion relationships were derived. I had many basic questions about the geometries and implementations of sample positioning in ARPES chambers, about the coordinate transforms themselves, about conventions on names for axes and orientations of spectrometers, and more. The answers to these questions are the entrance ticket to being able to derive $k_i(\theta_j)$ and $\theta_i(k_j)$, which are both required to carry ARPES data from angle to momentum on an analysis computer.

Through this process I learned

1. nearly every valid geometry has been instantiated on real hardware to enable some unique measurement capability – e.g. beamline 10 at the Advanced Light Source.
2. there are subtle and important differences between the coordinate transforms of seemingly identical geometries.
3. even perfect momentum transforms require careful handling of spectrometer calibrations for “serious spectroscopy applications”.
4. an experimentalist should only care about these issues in experiment design, and to debug results. The rest should be handled by a computer.

I’ve made available the Mathematica notebooks I used to derive a limited selection of the ARPES geometries which ended up in PyARPES as a supplementary file to this thesis. People seem to insist that deriving a subset of these is a rite of passage required to “properly understand” the experiment. Setting aside whether this is the case, deriving these shouldn’t be painful. You can adapt these notebooks to whatever exotic geometry you happen to choose for your ARPES experiments. When you’re done, consider adding your work to PyARPES or your local favorite analysis tool.