SpectroLab: An Open Source Matlab Based Toolbox for High Throughput Spectroscopy Analysis

Christopher Sims
School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA

(Dated: 15 March 2022)

We present an open source software package SpectroLab a Matlab based tool developed in 2018 for the analysis of spectroscopic data. In this package there are tools for derivative analysis, stacked energy contours, stacked plots for theory, 3D volumetric plots, and core level analysis, and derivative analysis. The package can currently be used for angle-resolved photoemission spectroscopy (ARPES) data, with the ability to also be used for other spectroscopic data in the future. We apply this program to the HfP, ZrSiS, and Hf2Te2P systems to demonstrate its robustness.

I. INTRODUCTION

In recent years, spectroscopic techniques such as angle resolved photoemission spectroscopy (ARPES)1–2, scanning tunneling microscopy (STM)3–5, neutron scattering6–8, etc. have improved to the point where the scientific communities use these tools extensively. These techniques reveal features in momentum and energy space that can give deeper understanding of the interactions of exotic quasi-particles that have taken interest in recent years such as topological insulators9,10, Weyl semimetals11–13, nodal-line semimetals14–16, and High order Topological insulators (HOTI)17,18. Furthermore, these techniques have been expanded to measure other dimensions such as spin, time, magnetism, etc. These techniques are always prone to noise which will decrease the ability to interpret the data correctly. Previously, researchers have employed proprietary programs in order to conduct the data analysis. These programs limit the ability for users to tune the way results are presented and create figures and graphs with a high degree of control. Here, we report a Matlab based program that is written in an easy to read way that allows for the analysis of ARPES data. Using Matlab’s built in tools, we have designed a program that gives users a high control in the way they present their data.

The data analysis of these materials is critical in order to understand complex interactions that are difficult to resolve by eye. By importing the data into Matlab, we provide a way for researchers to use the vast packages included in this program that are nontrivial to code in proprietary data analysis programs such as edge detection or neural networks. We present a suite of functions that allow for easy analysis and presentation of spectroscopic data in the Matlab suite called SpectroLab. The suite needs minimal programming experience since all of the tools included can be done in one click. As oppose to other data analysis techniques, our package is all inclusive and can be used with minimal need to modify the code in anyway. The only need for the program to work is an .ibw file that is formatted correctly. We present ARPES data in from the HfP, ZrSiS, and Hf2Te2P systems in order to show that this program is applicable to different types of ARPES data. This paper includes a detailed description of the theory behind how the main functions work in section 2. In section 3, we outline the functionalities of each of the functions in the initial release. In section 4, we go over how to install and use SpectroLab. Finally in section 5, we go over a few examples of using the software package in real systems.

II. METHODS

As the data is imported into SpectroLab it is automatically interpreted into k-space from the raw data using the following equations23:

\[ k_x = \frac{1}{\hbar} \sqrt{2mE_{kin}\sin\varphi \cos\varphi} \] (1)

\[ k_y = \frac{1}{\hbar} \sqrt{2mE_{kin}\sin\varphi \sin\varphi} \] (2)

In the program these are interpreted by the following approximation

\[ k_x = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times X \times 0.046 / 180) \] (3)

\[ k_y = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times Z / 180) \] (4)

In addition to this all supporting variables are automatically calculated within the setvariables() function.

\[ \text{KY\_Angle\_Max} = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times X_{max} / 180); \]
\[ \text{KY\_Angle\_Min} = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times X_{min} / 180); \]
\[ \text{KX\_Angle\_Max} = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times X_{max} \times 0.046 / 180) - 1.68; \]
\[ \text{KX\_Angle\_Min} = 0.5112 \times \sqrt{(E - 4.5)} \times (\pi \times X_{min} \times 0.046 / 180) - 1.68; \]

A. Sobel filter (2nd Derivative)

The common edge detection algorithms are the Sobel, Canny, Prewitt, Roberts, and fuzzy logic methods24–33.
In this paper, we use a modified version of the Sobel filter in order to conduct second derivative calculations. While it is possible to use a convolution neural network (CNN) for feature extraction\textsuperscript{34,35}, these networks require a large amount of computational resources and are only applicable to the data set they are trained on. Generative adversarial neural networks (GANs)\textsuperscript{36} are a much better alternative to create a neural network feature extraction since it can work on an infinite amount of data sets once it is properly trained. The downside is that a large amount of data sets (>1000) are required in order to properly train the neural network. Therefore, we forgo the inclusion of a neural network in our program.

The Sobel filter is a $3 \times 3$ matrix that calculates the second derivative locally in an image file (where $A$ is the image). There are two versions, the horizontal and the vertical which can be used, in the distributed program. The horizontal $F_x$ is convoluted with the image to form the second derivative of the image.

$$F_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & -1 \end{bmatrix} \ast A, \quad F_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \ast A \quad (5)$$

In order to detect edges after the Sobel filter is applied, one would find the square of the absolute value of the resultant $|F_x|^2$. By not conducting this last step we can resolve weak signals of ARPES. In the case of sharp data, the edge detection feature can be turned on with an additional line of code which we have included. All edge detection algorithms can be accesses via Matlab’s image toolbox.

```matlab
Data2D = squeeze(Mapping(:,EnergyLevel,:))';
h_Sobel = fspecial('sobel');
f_y = imfilter(double(Data2D), h_Sobel,'replicate');
imagesc(f_y)
```

\section*{B. Constant Energy Contours}

An energy contour is a slice of a 3D band mapping in ARPES ($k_x$,energy,$k_y$) where at a constant energy, the 2D $k_x$,$k_y$ surface is revealed. For constant energy contours we create a function that automatically interprets the energy levels and recreates them in an energy contour. For labeling we included a rounding function that rounds to the nearest significant figure. However, the program can interpret the exact value for publication.

```matlab
Energy_Limit = 200; %specified by user
Energy_Cut = 1:int16(Energy_Limit/6):Energy_Limit;
Energy_Contour = ypos - Energy_Cut % check limits;
hold on
subaxis(2,6,1,'sv',0.01,'sh',0.01);
f = squeeze(B(:,Energy_Contour(ii),:));
imagesc(f_y)
```

\section*{C. Volumetric plot}

The volumetric plot function takes advantage of how data is formatted for ARPES. Typically data is encoded in an (X,Y,Z,C) format. Where X,Y, and Z are $k_x$, $k_y$, and energy, while C is the interpretation of the intensity at that location. By using an alpha map, we can make weak signals transparent and reveal the 3D band structure for visualization. This has the ability to be a very powerful tool for investigating anisotropy in bands.

```matlab
%VolumetricPlot()
% edit this to get a good mapping
alphamap([0 linspace(0.0005, 0.4, 255)]);
```

\section*{D. Momentum cuts}

We have included several methods to aid the taking $k_x$ and $k_y$ momentum cuts from a 3D band mapping. We take these cuts by selecting a value of $k_x$ and $k_y$ and resolving the cut in a 2D image.

```matlab
f = squeeze(B(Kxlocation,:,:));
f = imresize(f,5);
imagesc(f',
```

\section*{E. Energy & momentum distribution curves}

Energy & momentum distribution curves (EDC & MDC) are used for the data analysis of spectroscopic data. In a set amount of intervals a line of data is taken out of the image and then plotted in order to enable researchers to resolve difficult to resolve images from ARPES or other spectroscopic data. We use the \textit{waterfall()} function in order to reconstruct the EDC and MDC plots seen in literature\textsuperscript{22}. The plot also have the benefit of being three dimensional with the ability to add a color map to further define intensity differences.
F. Stacked theory

In order to import theory, we create a 2D surface in a 3D box using Matlab’s `meshgrid()` function to import an image of the theoretical plot. To insure that the images that we import are of the same size we use the largest image dimensions as the map for all of the subsequent images. We do this because the 3D plot will not work unless we use the same size map. We then offset the images by one for each image. Now, one can scale the distance between image layers using the `daspect()` function of Matlab to scale the Z (Energy) in order to resolve the layers.

```matlab
%# plot each slice as a texture-mapped surface
% (stacked along the Z-dimension)
for k=1:size(im)
    surface('XData',X-0.5, 'YData',Y, 'ZData',
    Z.*-k, ... 'CData',im{k}, 'CDataMapping','direct', ... 'EdgeColor','none',
    'FaceColor','texturemap')
end
```

III. ANALYSIS TOOLS OF SPECTROLAB

Here we included an overview of the uses of SpectroLab. The fermi level can be found with the `fermisurface()` [Fig. 2(a)] and the ` ypos` variable

| Program        | Description                                      |
|----------------|--------------------------------------------------|
| Kx Cuts        | Creates k_x momentum slices                      |
| Ky Cuts        | Creates k_y momentum slices                      |
| Plot_KX_KY     | For quick check of data analysis                 |
| Fermi Surface  | Draws Fermi surface from 3D map                  |
| Momentum 2D    | 2D ARPES cuts                                    |
| Stacked plot   | Slices and stacks 3D momentum cuts               |
| Stacked theory | Stack theory similar to stacked plot             |
| Density Curves | EDC and MDC                                      |
| Energy Contour plot | ECP in exact steps     |
| Volumetric plot | 3D volumetric plot                              |
| Dispersion     | photon energy dependence                        |
| Inner core     | Analysis of inner core spectra                   |

A. Graphical User Interface (GUI)

We include a graphical user interface (GUI) [Fig. 1] that can be used as a general tool, more advance data selection and analysis will need to be done in code.

B. Constant Energy Contours

We include a function to extract constant energy contours (ECs) from a 3D band mapping. `StackedECP()` [Fig. 2(c,d)] is a feature that calculates ECs and stacks them on top of each other in order to visualize the bands dispersing in 3D space. The energy contour functions do the same, however the ECs are presented in a 2D format.

C. Volumetric plot

In the volumetric plot function, a 3D band mapping is interpreted with an alpha map (transparency map). This allows for all bands to be mapped in a 3D space. In addition, users can rotate and zoom in to further examine the electronic structure.

D. Momentum cuts

We have included several methods to aid the taking k_x and k_y momentum cuts from a 3D band mapping. `Plot_KX_KY`, Ky cuts (Fig. 3(a)), and Kx cuts (Fig. 3(b)) are all features that allow for one to construct these plots. `Plot_KX_KY` is a debug tool but can still be used to accomplish this feature.

E. Energy & momentum distribution curves

Energy distribution curves (EDCs) and momentum distribution curves (MDCs) are important tools for the analysis of ARPES data. Here, we construct programs that can create 3D EDC [Fig. 4(a)] and MDC [Fig. 4(b)] plots similar to those found in previous literature. The advantage of these types of plots is that they are able to give a 3D interpretation of the EDC and MDC which will allow for better understanding of the intensity at certain points (such as discerning if there is a band gap or weak feature in the band gap).

F. Photon energy dependence

Dispersion allows for one to interpret photon energy dependent cuts (not a mapping of photon energy) to include for interpretation or publication. The Dispersion map function can interpret an arbitrary amount of cuts specified by the user. [Fig. 5]

G. Stacked theory

Stacked theory is a function that allows one to import theoretical DFT energy contours and plot them in a way to compare to the `stackedECP()` function [Fig. 2(b)].
There is also a feature that allows for plotting theory in one of the energy contour functions.

IV. INSTALLATION AND USAGE

In this section, we present how to install and use SpectroLab.

A. Get SpectroLab

SpectroLab is an open source software package distributed under the MIT License. The code can be downloaded directly from the public code repository: github.com/christophersims/SpectroLab.

B. Installation and running the code

The code requires no installation, however, the folders that contain the code must be added to the path in order for them to run correctly (or the user could place all of the .m files in one directory). The user must have a Matlab version year of at least 2018 with the image toolbox add-on installed.

The code can be run by simply specifying the file path of the data one wishes to analyze. If there are no formatting errors the code should run without errors.

C. File formats

For ARPES there are three major file formats used (from the Igor suite), .txt, .ibw, and .pxt. Important data is lost in the .txt format and the .pxt format is compressed and cannot be decompressed outside of the Igor program. Therefore, this program only uses the Igor Binary Wave format (.ibw). In addition, the program only accepts files in the format $(k_x,\text{Energy},k_y)$, however, the file can be reformatted within the program.

V. EXAMPLES

A. Incorrectly formatted input

If the input of the code in not formatted correctly (in the $(k_x,\text{Energy},k_y)$ format) it is possible to still run the code by utilizing Matlab’s permute() function. This function allows for one to swap the matrices in order for the program to run correctly. In addition to using the permute the variables in variables() must also be changed in order to match the code. In figure , we present data that was formatted incorrectly but was able to run in the program and produce good results. Here we present the data of Hf$_2$Te$_2$P (Fig. 6) and ZrSiS (Fig. 7) which were imported using the permute function.

B. Plotting theory

Our implementation of plotting theory is currently an image importer, however, we have the code that allows for the input if the surface states and Fermi surface generated with the iterative surface green functions imported from Wanniertools. The calculations are not included in this package and must be calculated with tools such as Quantum Espresso, VASP, Abinit, Wannier90, etc. In figure 2(b), we show the capability to show stacked theory plots in order to coincide with the stackedECP() function. In addition, we have included a ECP function that also allows for the inclusion of theoretical calculation figures (see supplementary).

C. Volumetric plot

Here we present a volumetric plot of both a 3D band mapping correctly formatted Fig. 8(a) and an incorrectly formatted (corrected in code) in Fig. 8(b). This method is a powerful tool that will enable researchers to visualize bands in a 3D space.

VI. CONCLUSION

In conclusion, we have designed a Matlab based program that can be used to interpret spectroscopic data. Included in this tool are multiple functions which can be used for the analysis of ARPES data. To demonstrate the robustness of this package we show ARPES analysis of several data sets. This package can be used as a standalone with no other dependency on other programs. The features of this analysis suite is extensive with the addition of new analysis techniques such as the 3D volumetric plot. In the future, we aim to integrate the analysis of XRD and STM data into the SpectroLab suite.

SUPPLEMENTARY MATERIAL

See Supplementary material for further figures of SpectroLab’s output in the HfP$_2$, Hf$_2$Te$_2$P, and ZrSiS systems.

ACKNOWLEDGMENTS

The authors would like to thank the Matlab community which provided several methods and programs used in SpectroLab.
C.S. acknowledges the generous support of the Purdue ECE ASIRE fellowship and the National GEM consortium fellowship.

Correspondence and requests for additional functionality should be addressed to C.S.
(Email: Sims58@purdue.edu).

1A. Damascelli, Phys. Scr. T109, 61 (2004).
2D. Lu, I. M. Vishik, M. Yi, Y. Chen, R. G. Moore, and Z.-X. Shen, Annual Review of Condensed Matter Physics 3, 129 (2012).
3G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, Phys. Rev. Lett. 49, 57 (1982).
4J. Teresoff and D. R. Hamann, Phys. Rev. Lett. 50, 1998 (1983).
5G. Binnig and H. Rohrer, Phys. Rev. 59, 615 (1987).
6M. K. Wilkinson, E. O. Wollan, and W. C. Koehler, Annu. Rev. Nucl. Sci. 31, 303 (1981).
7E. Prince, Appl. Spectrosc. Rev. 34, 159 (2004).
8M. Strobl, R. Hartl, C. Grueneweg, R. Wocarek, and J. Plomp, Journal of Imaging 3, 64 (2017).
9M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. 82, 3045 (2010).
10Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. 5, 398 (2009).
11S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, S.-M. Huang, H. Zheng, J. Ma, D. S. Sanchez, B. Wang, A. Bansil, F. Chou, P. P. Shibayev, H. Lin, S. Jia, and M. Z. Hasan, Science 349, 613 (2015).
12B. O. Lv, H. M. Weng, B. B. Fu, R. Wang, H. Miao, J. Ma, P. Richard, X. C. Huang, L. X. Zhao, G. F. Chen, Z. Fang, X. Dai, T. Qian, and H. Ding, Phys. Rev. X 5, 031013 (2015).
13S.-M. Huang, S.-Y. Xu, I. Belopolski, C.-C. Lee, G. Chang, B. Wang, N. Alidoust, G. Bian, M. Neupane, C. Zhang, S. Jia, A. Bansil, H. Lin, and M. Z. Hasan, Nat. Commun. 6, 7373 (2015).
14A. A. Burkov, M. D. Hook, and L. Balents, Phys. Rev. B 84, 235126 (2011).
15C. Fang, Y. Chen, H.-Y. Kee, and L. Fu, Phys. Rev. B 92, 081201 (2015).
16C. Fang, H. Weng, X. Dai, and Z. Fang, Chin. Phys. B 25, 117106 (2016).
17S. Regmi, M. M. Hosen, B. Ghosh, B. Singh, G. Dhakal, C. Sims, B. Wang, F. Kabir, K. Dimitri, Y. Liu, A. Agarwal, H. Lin, D. Kaczorowski, A. Bansil, and M. Neupane, http://arxiv.org/abs/1911.03703v1.
18P. Schindler, A. M. Cook, M. G. Vergniory, Z. Wang, S. S. P. Parkin, B. A. Bernevig, and T. Neupert, Sci. Adv. 4, eaat0346 (2018).
19C. Sims, M. M. Hosen, H. Aramberri, C.-Y. Huang, G. Dhakal, K. Dimitri, F. Kabir, S. Regmi, X. Zhou, T.-R. Chang, H. Lin, D. Kaczorowski, N. Kioussis, and M. Neupane, arXiv:1906.09642 (2019).
20M. Neupane, I. Belopolski, M. M. Hosen, D. S. Sanchez, R. Sankar, M. Szlachta, S.-Y. Xu, K. Dimitri, N. Dhakal, P. Maldonado, P. M. Oppeneer, D. Kaczorowski, F. Chou, M. Z. Hasan, and T. Durakiewicz, Phys. Rev. B 93 (2016).
21M. M. Hosen, K. Dimitri, I. Belopolski, P. Maldonado, R. Sankar, N. Dhakal, G. Dhakal, T. Cole, P. M. Oppeneer, D. Kaczorowski, F. Chou, M. Z. Hasan, T. Durakiewicz, and M. Neupane, Phys. Rev. B 95 (2017).
22M. M. Hosen, K. Dimitri, A. K. Nandy, A. Aperis, R. Sankar, G. Dhakal, P. Maldonado, F. Kabir, C. Sims, F. Chou, D. Kaczorowski, T. Durakiewicz, P. M. Oppeneer, and M. Neupane, Nat. Commun. 9 (2018).
23R. Comin and A. Damascelli, ARCES: A Probe of Electronic Correlations (2015) pp. 31–71.
24N. Senthilkumar and R. Rajesh, in 2009 International Conference on Advances in Recent Technologies in Communication and Computing (IEEE, 2009).
25C. Ma, W. Gao, L. Yang, and Z. Liu, in 2010 2nd International Conference on Mechanical and Electronics Engineering (IEEE, 2010).
26W. Gao, X. Zhang, L. Yang, and H. Liu, in 2010 3rd International Conference on Computer Science and Information Technology (IEEE, 2010).
27A. Seif, M. M. Salut, and M. N. Marsono, in 2010 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (IEEE, 2010).
28D. Mintz, CVGIP: Image Understanding 59, 137 (1994).
29N. Kanopoulos, N. Vassanthevada, and R. Baker, IEEE J. Solid-State Circuits 23, 358 (1988).
30A. Huang, I. W. Tsang, and D. Xu, IEEE Trans. Neural Netw. Learn. Syst. 23, 902 (2012).
31I. Canny, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8, 679 (1986).
32A. Huertas and G. Medioni, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8, 651 (1986).
33G. T. Shirvashkan and C. Chandrasekar, in IJCSI International Journal of Computer Science Issues (2012).
34C. M. Bishop, “Neural networks: A pattern recognition perspective,” (1996).
35H. Peng, X. Gao, Y. He, Y. Li, Y. Ji, C. Liu, S. A. Ekanaha, D. Pei, Z. Liu, Z. Shen, and Y. Chen, http://arxiv.org/abs/2001.10908v1.
36J. J. Goodfellow, J. Pouget-abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” (2014).
37M. P. L. Sancho, J. M. L. Sancho, J. M. L. Sancho, and J. Rubio, J. Phys. F: Met. Phys. 15, 851 (1985).
38Q. Wu, S. Zhang, H.-F. Song, M. Troyer, and A. A. Soluyanov, Comput. Phys. Commun. 224, 405 (2018).
39P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiariotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter 21, 395502 (2009).
40G. Kresse and J. Hafner, Phys. Rev. B 48, 13115 (1993).
41X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, P. Ghosez, M. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Joblet, G. Jomard, S. Lefour, M. Mancini, S. Mazevet, M. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah, and J. Zwanziger, Comput. Phys. Commun. 180, 2582 (2009).
42A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. 178, 685 (2008).
FIG. 1. **Graphical User Interface** (a) The main page the GUI figure with the main release of SPECTROLAB.
FIG. 2. Example of functions (a) Result of the fermisurface() function. (b) Result of the stacked_theory() program. (c) Output of the stackedECP() function in the HfP$_2$ system. (d) Output of the stackedECP() function in another dataset in the HfP$_2$ system.

FIG. 3. $k_x$ and $k_y$ momentum cuts (a) Fermi surface (rightmost panel) and $k_x$ momentum cuts (cut1-cut5) taken from a 3D mapping by using the kxcuts function. (b) same as (a) but for $k_y$ momentum cuts
FIG. 4. **Energy & Momentum Distribution Curves**
(a) EDC curve  (b) MDC curve
FIG. 5. Dispersion map A dispersion map of HfP$_2$ resolved with the dispersion() function

FIG. 6. Hf$_2$Te$_2$P (a) Normal energy contour (b) 2nd derivative energy contour
FIG. 7. ZrSiS Energy Contour  (a) Normal energy contour (b) 2nd derivative energy contour

FIG. 8. Volumetric plot  (a) volumetric plot of correctly formatted code. (b) Volumetric plot of incorrectly formatted input. There is no errors in the incorrectly formatted 3D band mapping contour