A Generalizable Framework for Algorithmic Interpretation of Thin Film Morphologies in Scanning Probe Images

Wesley K. Tatum,[‡,a] Diego Torrejon,[‡,b,c] Patrick O'Neil,[‡,b,c] Jonathan W. Onorato,[a] Anton B. Resing,[a] Sarah Holliday,[a] Lucas Q. Flagg,[d] David S. Ginger,[d] and Christine K. Luscombe[,a,d,e].*

a ~Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, United States.

b ~BlackSky, 13241 Woodland Park Road, Suite 300, Herndon, VA 20171, United States.

c ~Department of Mathematical Sciences, George Mason University, Fairfax, VA, 22030 United States.

d ~Department of Chemistry, University of Washington, Seattle, WA 98195, United States.

e ~Department of Molecular Engineering and Sciences, University of Washington, Seattle, WA 98195, United States.

‡ ~These authors contributed equally

* Corresponding author. E-mail: luscombe@uw.edu

ABSTRACT: We describe an open-source and widely adaptable Python library that recognizes morphological features and domains in images collected *via* scanning probe microscopy. π-Conjugated polymers (CPs) are ideal for evaluating the Materials Morphology Python (m2py) library, because of their wide range of morphologies and feature sizes. Using thin films of nanostructured CPs, we demonstrate the functionality of a general m2py workflow. We apply numerical methods to enhance the signals collected by the scanning probe, followed by Principal Component Analysis (PCA) to reduce the dimensionality of the data. Then, a Gaussian Mixture Model segments every pixel in the image into phases, which have similar material-property signals. Finally, the phase labeled pixels are grouped and labeled as morphological domains using either connected components labeling or persistence watershed segmentation. These tools are adaptable to any scanning probe measurement, so the labels that m2py generates will allow researchers to individually address and analyze the identified domains in the image. Thus, allowing to describe the morphology of the system using quantitative and statistical descriptors such as the size, distribution, and shape of the domains. Such descriptors will enable researchers to quantitatively track and compare differences within and between samples.

Introduction

Thin-film and flexible electronic technologies, including devices such as photovoltaic cells, field effect transistors, and thermoelectric devices are enabled by materials such as π-conjugated polymers (CPs).[1–15] There are many promising candidate materials for active layer materials in these devices, which have highly tunable properties. All of these materials share a similar bottleneck: their device performance and stability are strongly dependent on their morphology at the micro- and nanometer- scale.[3,16–23] For any candidate material to achieve commercialization, it is necessary to predictably link parameters such as chemical structures, intermolecular interactions, processing conditions, and composition to final device performance, without the need to take an Edisonian approach to synthesis and optimization. The thin film morphology of a device active layer is a critical piece in this connection, so it is important to establish reproducible and quantitative descriptors of such morphology.

Devices made from thin-film electronic materials have greatly improved over the past decades, and CP-based devices have their electronic performance in some applications, such as organic electrochemical transistors (OECTs), conjugated polymers can outperform inorganic materials.[16,24,25] Due to the intrinsic connections between their device properties and their chemical structures, molecular weight, and intermolecular interactions, CP-based devices can be exceedingly complex. Compounding this, CP morphologies often have dimensions that range from nanometers to micrometers. Such domains are difficult to interpret in single component films and become even more complex in blends of two or more components.[26–32] Though this frustrates the analysis of polymers, it also contributes to their tunability and responsivity and allows researchers

to access specific performances, structures, and properties. This versatility also makes CPs well-suited for developing tools that quantify morphology.

Despite its importance, it has been difficult to quantitatively describe morphologies and nanometer-scale structure of thin films and interfaces. Imaging techniques, such as scanning probe microscopy (SPM), have been extensively developed and deployed to observe surfaces at this length scale for a wide variety of materials through the imaging of properties, such as conductivity or viscoelasticity, which are used in the present work.[24,33–36] The set of material properties that can be investigated by SPM techniques is rapidly expanding, and probe samples on length-scales ranging from Angstroms to millimeters.[37–41] These observations have enabled significant progress in understanding morphologies, but are traditionally constrained to qualitative or semi-quantitative sampling, as a result of the difficulty of hand-measuring features in microscopy images. Fortunately, the modularity and sensitivity of these instruments allow for a wide range of material properties to be simultaneously measured with topography. The resulting data have similarly formatted outputs, simplifying the comparison of different properties across a single scan.

There has been compelling research from various groups on utilizing computer vision and machine learning approaches to automatically identify features in micrographs and SPM measurements.[11,42-52] These approaches often use supervised models, which inherently restrict their application to materials or morphologies that are similar to the samples used in training the model. Moreover, the performance of supervised models is connected to the number of samples in the training dataset. This limits their applications further, as it is first necessary to analyze and manually label a significant amount of data to build a large training set in order to improve the accuracy of these supervised learning approaches. In contrast, the data agnostic nature of unsupervised classification is particularly useful in building base models for the characterization

of materials; such models can be easily extended to different materials, instruments, and techniques because they do not require training data.[53]

In this study, we present an open-source and adaptable Python toolkit, m2py, which is capable of interpreting and labeling morphological domains and features in images that have widely varied morphologies. Moreover, the m2py toolkit contains modular classes and functions that can be assembled into a customizable workflow to extract meaningful information from the domains and features in the SPM data. The m2py library simplifies the application of computer vision techniques; in particular, the workflow used herein enhances feature signals prior to applying a series of unsupervised methods. Once identified, m2py generates label maps of the identified features, so that they can be individually addressed and analyzed. These label maps can then be used to generate quantitative descriptions of the morphological information that has always been present in the SPM data. In this work, we have shown that domain size and major axis length are calculatable from m2py labels. Future work will also extract further morphological descriptions, such as aspect ratio, orientation, and perimeter length, so that the connection between such descriptors and device performance can be evaluated. In this work we demonstrate m2py's versatility in morphological classification by applying a generic m2py workflow to thin films of semicrystalline poly(3-hexylthiophene) (P3HT) in organic field-effect transistors (OFET) active layers, blended binary thin films of P3HT: phenyl-C_{61}-butyric acid methyl ester ($PC_{61}BM$) and poly[[4,8-bis[(2-ethylhexyl)oxy]benzo[1,2-b:4,5-b']dithiophene-2,6-diyl][3-fluoro-2-[(2-ethylhexyl)carbonyl] thieno[3,4-b]thiophenediyl]] (PTB7):$PC_{71}BM$ in organic photovoltaic (OPV) active layers, and self-assembled nanostructures of P3HT spin-coated onto inorganic SiO_2 or ITO substrates. The m2py toolkit is compatible with any SPM technique, which we demonstrate by applying the same m2py workflow to measurements from force-distance mapping AFM

(FDM), amplitude-modulated frequency-modulated bi-modal AFM (AMFM), and conductive AFM (C-AFM), three techniques that each have different degrees of spatial resolution and data dimensionality.

Results and Discussion

SPM data is structured as a stack of 2-dimensional matrices with each layer in the stack (commonly referred to as a channel), representing a different aspect of the material's response. [54,55] The methodology primarily used in this investigation, FDM, has six channels of signals: adhesion, deformation, dissipation, modulus, height, and stiffness. By comparing the values from each channel for each pixel, pixels can be sorted, filtered, or clustered by their similarity. A summary of the m2py workflow is shown in Figure 1.

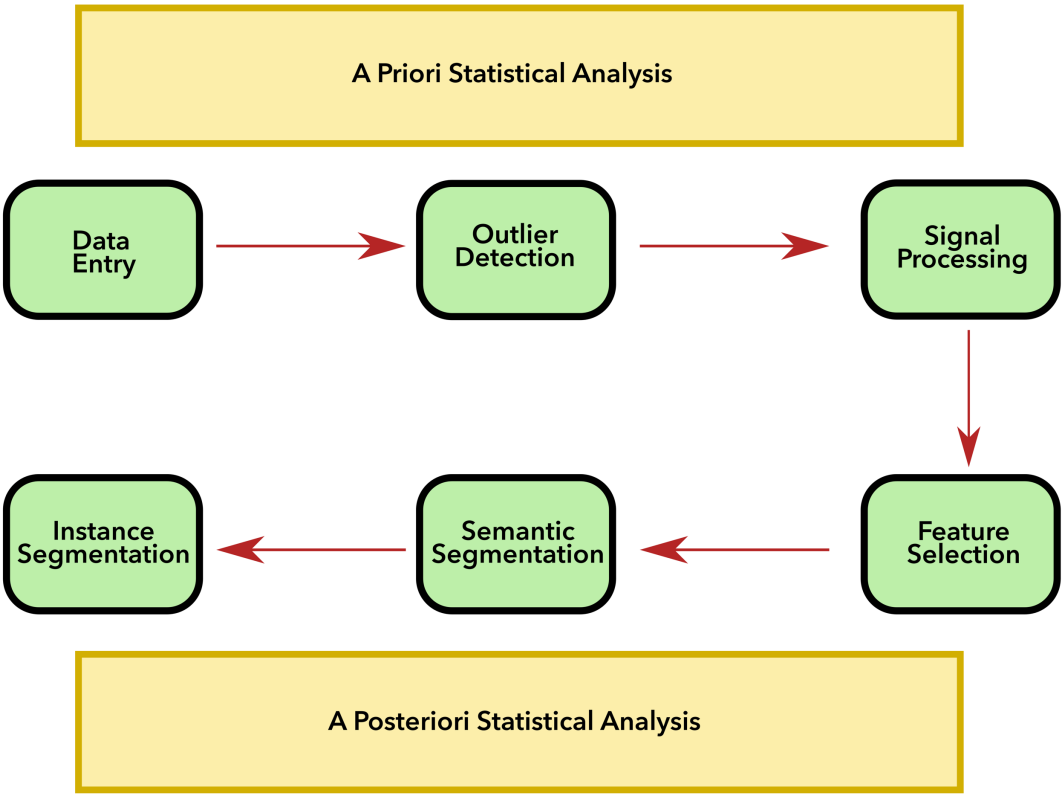


Fig. 1 A generalized m2py workflow. Raw or pre-cleaned data is fed in by the user. Next, outlier removal and fast Fourier transforms are applied to increase the sensitivity of m2py's segmentation modules. Feature selection finds the most descriptive features from the data while reducing data dimensionality to distill the material-property signals. This is followed by semantic segmentation (pixel classification) and then instance segmentation (pixel clustering). Statistical information can be extracted at any point in the workflow.

2.1 Data Intake and Pre-Processing

The SPM community has developed robust toolkits for viewing SPM data and converting between file formats. Packages such as Gwyddion can parse instrument specific and open-source file formats and perform automated routines that handle SPM data in a tensor fashion. However, these tools primarily accelerate data processing and improve interpretation of the image by the microscopist.[54–56] They do not seek to identify and label segmented domains for algorithmic analysis, nor do they enable any automation of the identification. By relying on these familiar tools for data cleaning and translation into NumPy-readable file types, m2py is able to process data from virtually any source.

Once the SPM data is converted to a NumPy format, the material-property signals need to be enhanced. Outlier detection and image denoising are essential at improving the signal to noise ratio of the raw SPM data. These outliers need to be removed early, as they can overpower material-property signals during later clustering methods. Examples of outliers includes pixels associated with surface aggregates, pin-holes, or tip-scars. The m2py method pre_processing.extract_outliers() recognizes these pixels and generates a Boolean matrix of outlier labels so that the corresponding pixels are ignored by the subsequent classification models, improving their results. This method has adjustable windows and thresholds and is typically

applied to the height channel. The output of outlier detection is illustrated in Figure S1. For further noise reduction, the method pre_processing.frequency_removal() applies image compression via fast Fourier transform to each layer using an adjustable band-pass filter to remove noise signals at high or low frequencies.[57] An example of this method being used to remove low-frequency background noise is shown in Figure S2. Additionally, the Fourier-space signal data can be accessed and used to analyze the orientation, alignment, and signals within images.[52]

2.2 Feature Selection and Semantic Segmentation

Following outlier removal and signal processing, pixels are analyzed and classified into a user-specified number of phases using a combination of Principal Component Analysis (PCA) and the Gaussian Mixture Model (GMM). PCA serves as the feature selector combining all data channels into fewer but more informative features, while retaining most of the information / variance of the system.[42,58] Even though the generated principal components are linear combinations of the input channels, they represent a complex vector-space relationship to the original data. Through the retention of only the most informative features, PCA speeds up the computationally expensive process of classification by minimizing the number of channels to examine. So, incorporating PCA into the m2py workflow is often beneficial to further enhance materials response signals prior to classification.[42,59] An example of PCA selection of material response information for FDM measurement of an annealed P3HT:$PC_{61}BM$ thin film is shown in Figure 2. Once PCA has been applied, the principal components can also be accessed and analyzed through the SegmenterGMM() class. A principal component of the OPV film from Figures 2a and b is shown in Figure 2c and an example of cross-correlation of those principal components is plotted in Figure 2d. In Figures 2b, c, and d, the red and blue colors indicate the final GMM phase classification of

each pixel. Even though the number of channels is reduced by half in this example, the cross-correlation in Figure 2d confirm that the principle components retain the signal clusters visible in the input channel cross-correlations of Figure 2b.

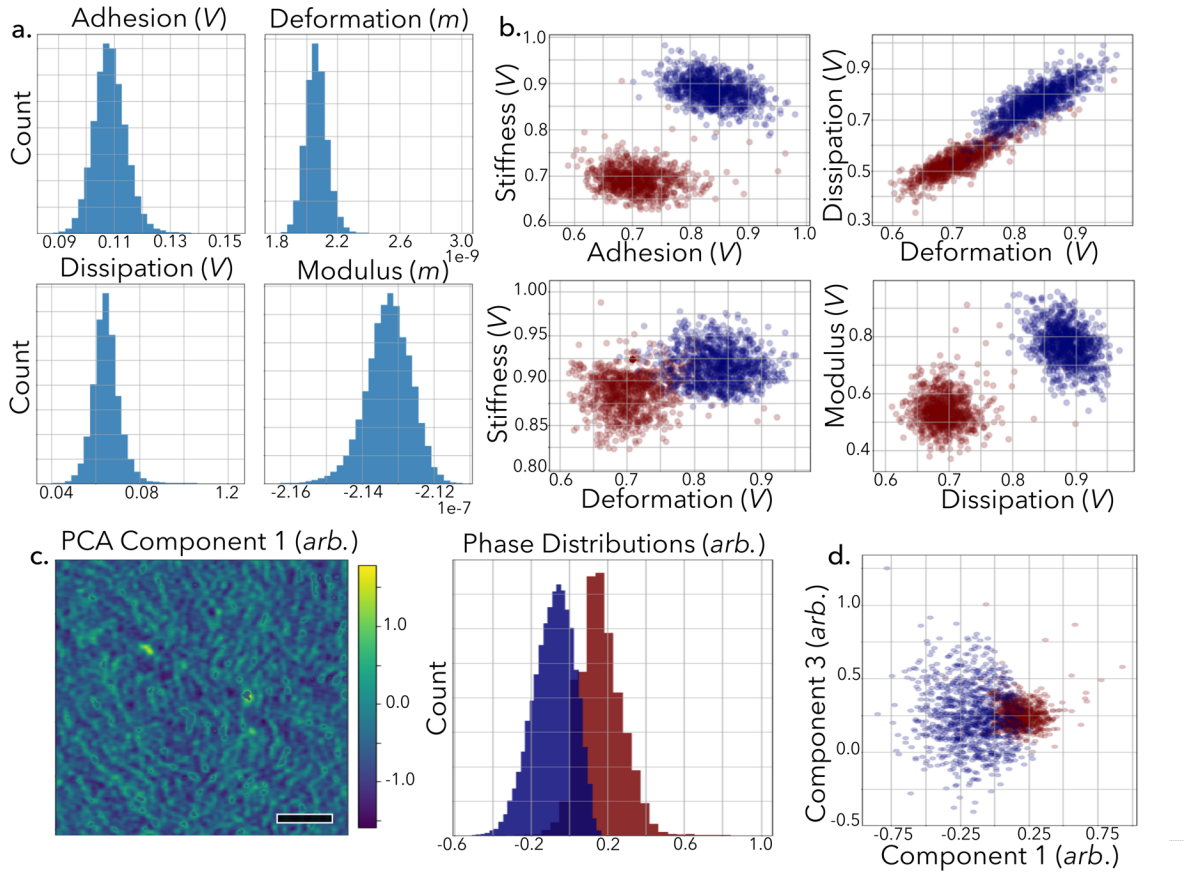


Fig. 2 Signal distributions and cross-correlations are easily extracted throughout the workflow. **a.** Four input material-property distributions from an annealed P3HT:$PC_{61}BM$ thin film. Input channel units are relative. **b.** Cross-correlations from the same sample after GMM segmentation, with red and blue representing the two GMM labeled phases. **c.** (*left*) An example principal component from PCA of the same sample. Principle component units are arbitrary. (*right*) The property distributions for each of the two GMM phase for the principal component. **d.** Cross-correlation between principal component 1 and 3, color-coded by their GMM phase labeling. Scale bar is 200 nm, images are 256 × 256 pixels.

Once the material response signals have been selected by applying PCA, GMM is used as a semantic segmenter to classify pixels by their clustered material response signals. Afterwards, m2py assigns each pixel a phase label corresponding to its GMM classification. The number of phases is easily tuned by a user; for instance, if a user is expecting a binary blend to have two phase-pure regions, and one blended region, the GMM can be made to deconvolute the SPM data into three distinct phases. A more detailed discussion of the GMM implementation is presented in the Supporting Information on page S8.

PCA and GMM are combined and implemented through the SegmenterGMM() class. Which features are used to segment and classify individual pixels has a significant impact on the overall performance. To clarify both the use of the SegmenterGMM() class and the impact of user choices on overall classification, several examples are presented below.

One common consideration is whether or not the height channel should be included during GMM phase label assignment. Depending on the magnitude of the differences in the material-response signals that the GMM segmenter uses to classify the pixels, height information can be either advantageous or disruptive when included. Omission of height data tends to be useful if the topographical features are not reflective of the underlying morphological domains, and thereby misrepresent the true morphology of the film, or when the material-response signals show only small fluctuations. This is the case with P3HT thin films, where surface features are often dictated by drying kinetics and surface-air interactions, rather than phase separation into crystalline and amorphous domains.[60–62] This is demonstrated in Figure 3, where the modulus and adhesion channels show small domains (Figure 3a.), as expected of P3HT skin-layers.[63]

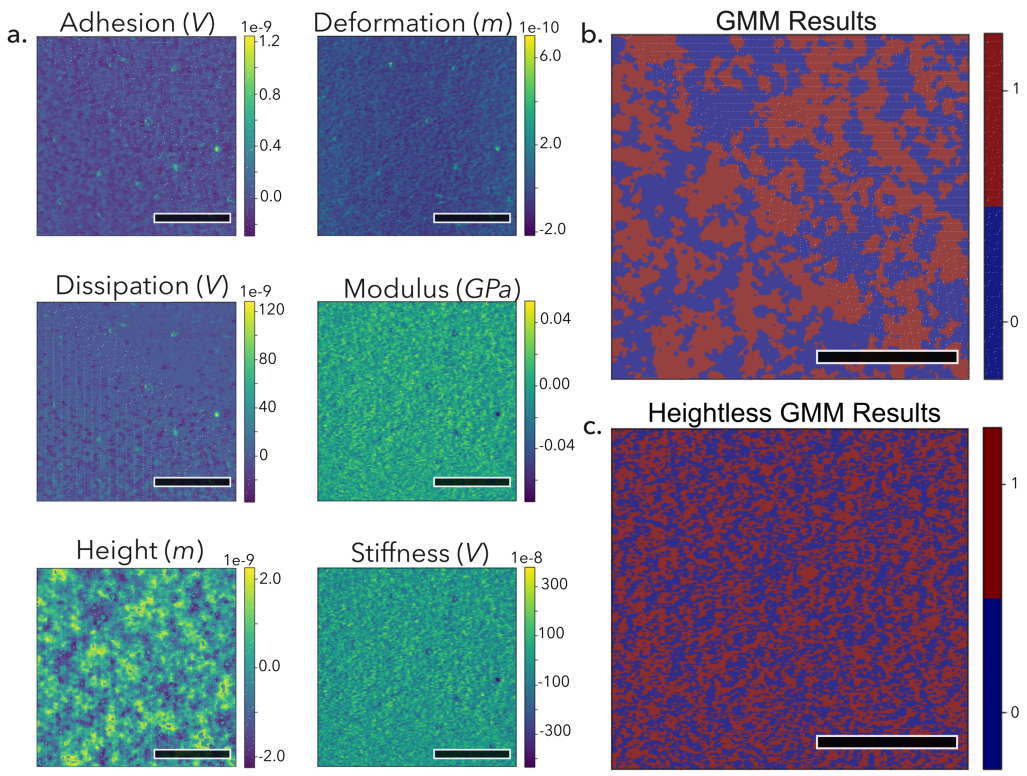


Fig. 3 Input data and the GMM label maps from FDM measurements of a P3HT OFET active layer. **a.** The input data channels after outlier and noise removal. Colorbar units are relative. **b.** GMM labeling when the height channel is included. **c.** GMM labeling when the height channel is excluded. All scale bars are 500 nm, imaged at 512 × 512 pixels.

When height information is included, GMM results show large domains (Figure 3b) that do not reflect the small, isotropic domains that are seen in the five material-property channels. Exclusion of the height channel is accomplished with the 'heightless' flag during the SegmenterGMM() class instantiation, and the corresponding results of this heightless classification, Figure 3c, are far more representative of the morphological domains visible in Figure 3a.

In cases where differences in material properties are significantly large, such as with PTB7:$PC_{71}BM$ binary blends (Figure 4), the topographical features can actually improve GMM interpretation by correcting edge-induced tip-surface interactions. Figure 4b and c show this

improvement in GMM labeling of a PTB7:$PC_{71}BM$ thin film with and without height, respectively. The spherical domains seen in the height channel of Figure 4a appear symmetric, despite being scanned left-to-right. This is not the case in the stiffness channel, where the left edges of the aggregates appear far softer than the right edges. This is caused by the probe-scanning direction–significant changes in sample thickness change the amount of the tip surface-area contacting the surface and artificially increases or reduces the measured adhesive forces as a result of the measurement's dependence on the probe tip's contact-radius.[64] This effect is graphically explained in Figure 4d.

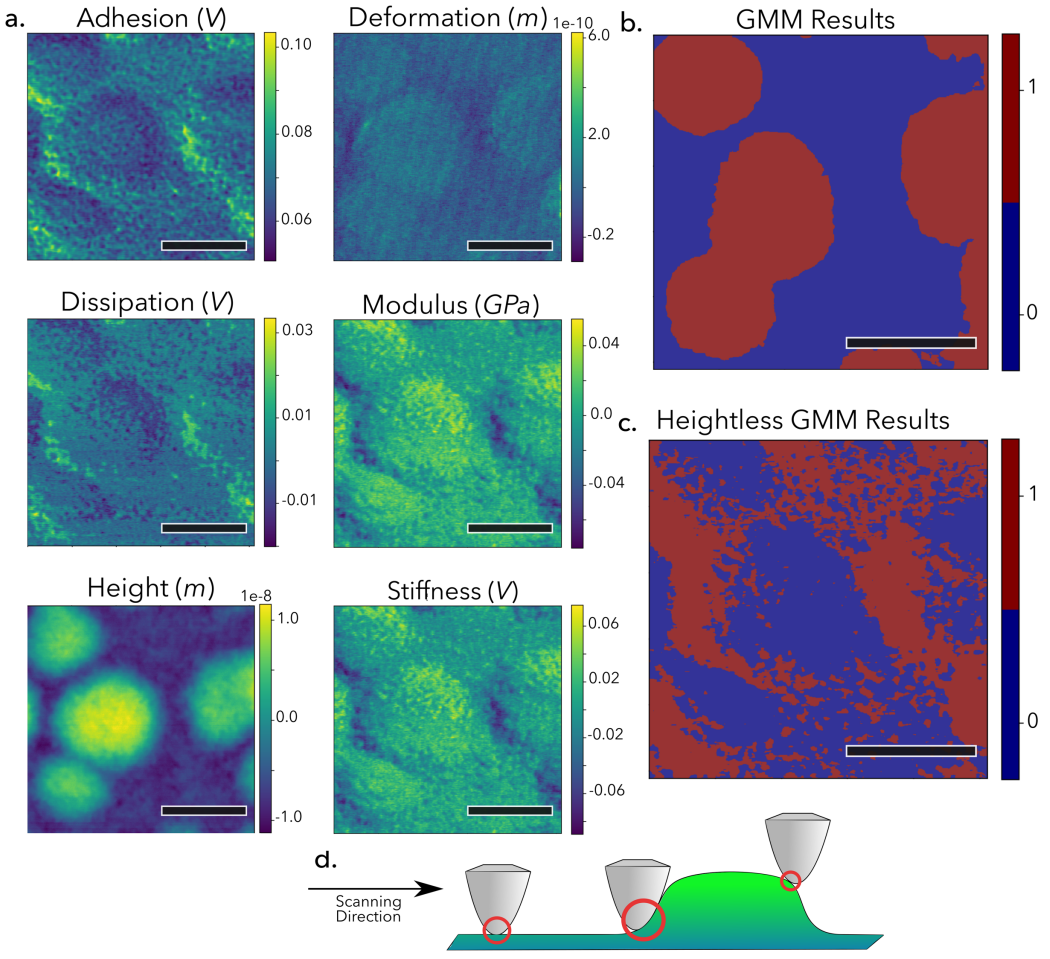


Fig. 4 Input data and GMM label maps from FDM measurements of a PTB7:$PC_{71}BM$ thin film OPV active layer. **a.** The input data channels after outlier and noise removal. Colorbar units are relative. **b.** GMM labels when the height channel is included. **c.** GMM labels when the height channel is excluded. **d.** Graphical explanation of how large features can artificially alter measured material response signals by varying the amount of surface-area contact between the probe-tip and sample surface. The scale bars are 200 nm, images are 256 × 256 pixels.

Therefore, in this example, excluding the height channel from classification eliminates edge-induced distortions from the GMM labels. Analogous effects follow for other SPM techniques, where inclusion of topography data can see similar improvements. Judicious selection of which channels to use or remove can make a significant improvement in the quality of data produced through m2py.

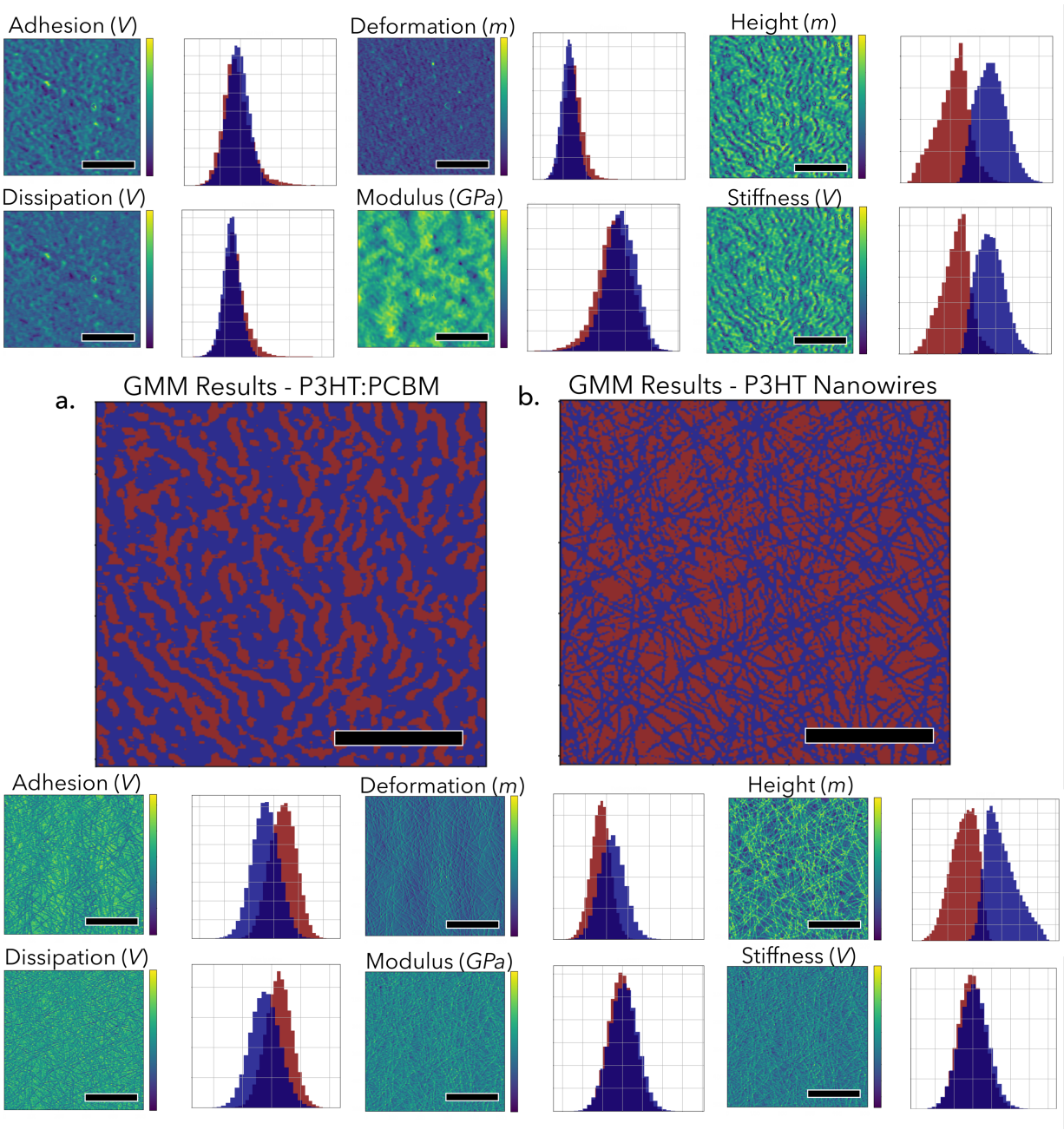


Fig. 5 Results of GMM segmentation on two different samples. **a.** 2 component classification of a P3HT:$PC_{61}BM$ thin film, where red and blue corresponds to the different components. The input channels are shown in the top two rows alongside their material-property signal distributions that are color coded by GMM component. Scale bars are 200 nm, images are 256 × 256 pixels. **b.** Results of 2 component classification of P3HT nanowires (blue) spin-coated onto a passivated SiO_2 substrate (red). The input channels are shown in the bottom two rows, alongside their

corresponding material-property signal distributions color coded by GMM phase components. Scale bars are 500 nm, images are 512 × 512 pixels.

Thus far, we have examined isotropic domains, but CP often exhibit highly anisotropic morphologies; nevertheless, the efficacy of GMM classification is not dependent on the size or shape of domains. This is because GMM classification depends on materials property signals, rather than strictly edge detection, as most supervised techniques do. An exploration of the use of m2py's GMM tool in classifying anisotropic morphologies is visible in Figure 5. A GMM label map for annealed P3HT:$PC_{61}BM$ is shown in Figure 5a, below its input channels and GMM labeled property distributions. As seen in these results, even though the bulk heterojunction (BHJ) morphology is markedly different from the morphologies in previous figures, the same workflow and GMM segmenter was able to cluster material response signals into highly anisotropic and asymmetric domains, closely matching the input data's morphology.

As an examination of a case of significant anisotropy, P3HT nanowires were also examined with the same m2py workflow. The results of the GMM classification are shown in Figure 5b, above their input channels and GMM labeled property distributions. Even though these P3HT nanowires are even more anisotropic than the BHJ domains, m2py is still able to accurately parse between the nanowires, the SiO_2 substrate, and scanning defects, such as tip-scars. This is a useful result because tip-scars occur commonly with sharply featured samples, such as nanowires, and can visually resemble the material domains. Regardless, of this effect m2py is able to distinguish nanowires from tip-scars by using differences in their material response signals. Despite the complex and varied morphologies of CP thin films, m2py's workflow has no problem interpreting both highly anisotropic CP morphologies and simpler or axisymmetric morphologies.

2.3 Instance Segmentation

The identification of each pixel's GMM phase within a scan allows a great deal of insight into SPM images, but much of the salient information lies in the size and distribution of pixel clusters, which construct the morphological domains and features. Unlike semantic segmentation, instance segmentation (pixel clustering) requires different techniques to handle different morphologies. The m2py toolkit currently implements two different instance segmentation methods: connected components labeling and persistence watershed segmentation (PWS), although additional instance segmentation methods could be easily integrated into m2py workflows.

The connected component labeling mechanism takes in the label map from GMM segmentation and applies a graphical algorithm that clusters tangential pixels that have the same GMM phase label.[65] Those connected pixels that have the same GMM label are assigned a unique domain label. The level of connectivity can be adjusted to include either only the four pixels that share an edge with the pixel in question, or to also include the four corner pixels. The total number of unique labels denotes the total number of domains present in the sample. Domains are sorted by size before receiving a label, meaning that domain 1 has the most pixels, and the domain with the fewest pixels is listed last. Adjustable thresholding is used as needed to combine domains that may be too small to be physically valid.

The other instance segmentation method, PWS, applies a watershed approach to identify domains. Watershed segmentation is commonly used in greyscale morphology description.[66] In m2py, we use a single channel – the height channel is used as default – to act as the magnitude values that form watershed-derived segments, which represent morphological domains. Other channels or labels, such as a principal component, can be used for PWS segmentation instead of height, as desired. Standard watershed approaches tend to over-segment images. To address this

issue, the PWS model employs persistent homology to merge watershed-derived segments incrementally to form larger domains.[67] Initially, too many segments arise due to noise, which indicates that a low merging threshold will result in too many identified domains. As the threshold is increased, these over-segmented regions merge and fewer domains remain. At the critical point where the most noise is eliminated and the most domains are preserved as segments, there is a sharp decrease in the number of domains, as shown in Figure 6. A rigorous discussion on selecting the appropriate instance segmenter, based on morphological structures, is given in the Supporting Information on page S10. Figure S3 shows a comparison of instance segmenters on 4 different morphologies. In short, we find that PWS is better for more isotropic morphologies and systems where tangential domains may be separate, whereas connected components labeling is ideal for highly anisotropic and irregularly shaped morphologies. Figure 6 shows the analysis of a PTB7:$PC_{71}BM$ sample and throughout the m2py workflow. The clearly visible spherical aggregates in the input height channel resembles the same information observed in the rest of the channels. GMM identifies these aggregates as a different phase than the matrix constituent, and labels them accordingly. Next, the height channel is used for instance segmentation with PWS. The individual aggregates, as well as different domain boundaries in the matrix, are identified and labeled according to their size. Finally, the PWS domains corresponding to the GMM phase 1 pixels are shown. By comparing these PWS labels against those from the initial GMM segmentation, we can see that the PWS slightly underestimates the size of these domains, as represented by the residual yellow coloration. In this case, the labelling error is less than 7% of the total image area. This error may reflect the dilation of GMM labels from tip-effects, as a learning from multiple channels, some more highly influenced by tip-radius, while PWS is based on a single channel. Erosion operations have been implemented to successfully reduce tip distortions, however

tip-deconvolution during pre-processing would more accurately reflect the surface morphology in GMM labeling. Future work will include expanding m2py's pre-processing and segmentation modules with such functions.

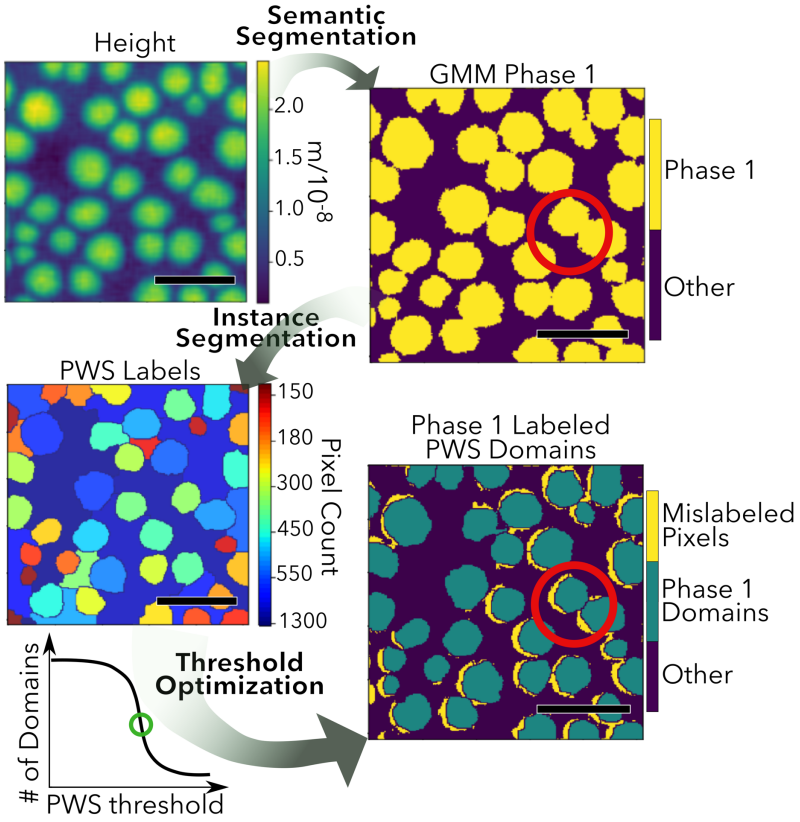


Fig. 6. A representative channel from a FDM image of PTB7:$PC_{71}BM$ is shown, followed by the results throughout m2py analysis of the image. First undergoing semantic segmentation, the resulting GMM phase labels show the spherical aggregates to be a distinct phase from the matrix. Phase 1 is shown in yellow, phases 2 and 3 are combined in purple. Next, the height channel is used in instance segmentation by the PWS method, producing individually indexed morphological domains. The color of each individual domain is selected according to its size. Finally, PWS labels that correspond to GMM phase 1 are plotted in green and the mislabeled phase 1 pixels plotted in yellow for contrast. Scale bars are 1 μm, images are 512 × 512 pixels.

Regardless of the instance segmenter that is used, a major consideration of instance segmentation is to establish a threshold for what comprises a morphological domain. Molecular and crystal structure information is useful for establishing a minimum number of pixels required to be labeled as an individual domain. In m2py, this instance segmentation threshold, label_thresh, is easily altered in the post_processing.py module to suit the parameters of the material and measurement. In our FDM experiments, the probe tips had radii of ~7 nm, which correspond to ~19 monomer units of P3HT or ~5 monomer units of PTB7. Given the community understanding of the crystallographic dimensions of these materials, the amount of polymer that interacts with the surface area of the probe tip is similar in size to a domain and could reasonably be considered an entire domain or aggregate. Therefore, a single pixel could be the optimal threshold, although higher resolution measurements would certainly provide a more detailed insight into the exact morphological structure.[68–70] The P3HT nanowires shown in Figure S1 have been measured in previous work to be 20 nm wide, or approximately 2 pixels, with lengths of up to a micron.[30,32] A threshold of 4 pixels, therefore, can provide detailed morphological insight for our imaged nanowires and P3HT thin films. Lastly, phase-separated, binary thin films may have dimensions of domains on a similar length scale, but these morphologies tend to require larger thresholds because the total area of the domain is larger. For example, the domains shown in the third row of Figure S1 contain at least 100 pixels each, with the background domain containing tens-of-thousands of pixels. These results emphasize the need for material-specific information to correctly interpret morphologies, although only minimal prior knowledge is needed for m2py to correctly identify morphologies.

2.4 Final output and transferability

As m2py classifies and clusters each pixel, it creates new label maps that can be analyzed separately or collected and appended to the original 3D image matrix for further analysis. Because these label maps correspond to the pixels in the original image, the properties of each pixel within any given phase or domain can be easily accessed. For example, histograms of the materials-property signals in each phase is easily determined from the GMM labeling. Further, by using the unique domain labels from instance segmentation each domain can be quantitatively examined. These domains can then be sorted by their corresponding GMM label, as shown in Figure 7b. Then, domain metrics are readily extracted, such as length, width, perimeter, and orientation.[65] Because the domains are uniquely labeled, these domain measurement methods can be iteratively applied to all of the domains in an image, as shown in Figures 7c-e. When sorted by GMM label, the size of the domains in each phase can also be determined, as presented in Figure 7c. In the example shown, an annealed P3HT:$PC_{61}BM$ thin film's height channel is shown (Figure 7a) and GMM segmenting labels for 2 distinct phases are created (Figure 7b). After semantic segmentation, it is clear that phase 1 domains comprise the matrix, which is colored red in the domain label map. However, it is visible that the matrix phase is not entirely co-continuous, so non-contiguous phase 1 domains are considered separate from the large matrix domain by the instance segmenter and therefore receive unique domain labels (Figure 7b, *left*). The matrix domain is the largest in the sample and is visible as the 'outlying' data point in the major-axis length and area plots, as shown in Figures 7c-7e.

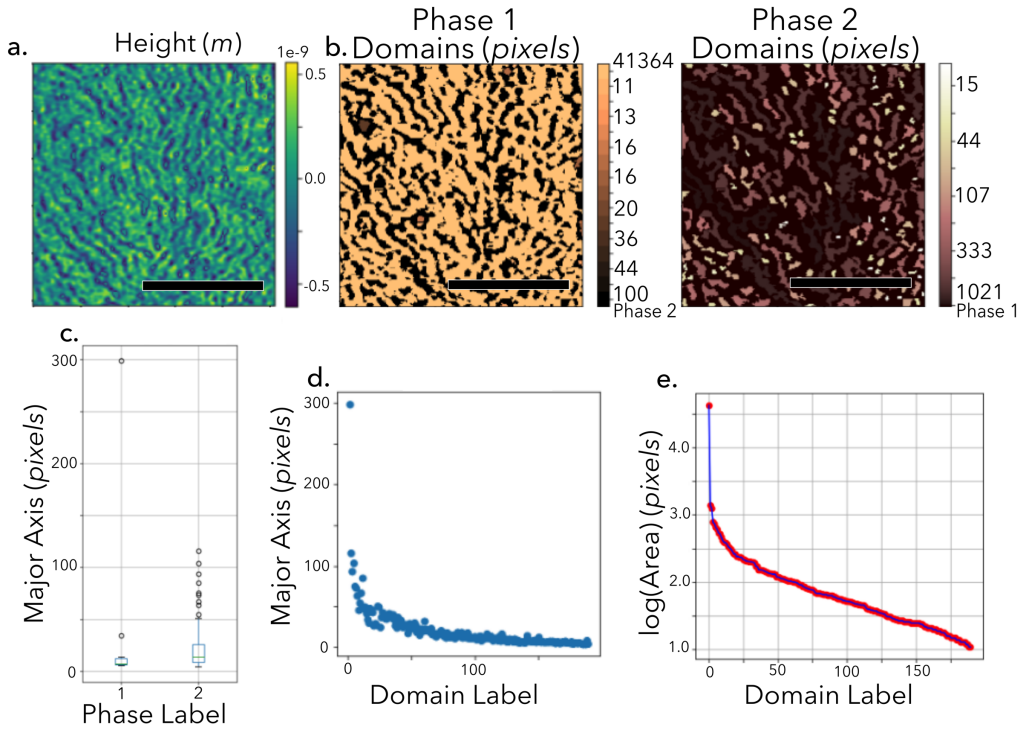


Fig. 7 Throughout the m2py workflow, labels can be used to calculate and analyze relative ratios, property distributions, and domain descriptors. Simple examples are presented for a P3HT:PC_{61}BM thin film. **a.** The height channel from FDM analysis of a P3HT:PC_{61}BM thin films is shown. Colorbar units are relative. **b.** Domain labels are determined by connected components labeling and then sorted by their GMM label. The color of each individual domain is selected according to its size. (*left*) Phase 1 domains, with black domains being phase 2. The co-continuous matrix domain is remapped to the lightest color for visibility. (*right*) Phase 2 domains, with black domains being phase 1. **c.** The major-axis length, in pixels, of connected components labeled domains, sorted by GMM label. **d.** The major-axis length, in pixels, of connected components labeled domains. **e.** The log of connected components labeled domain areas in pixels. The scale bars are 1 μm and the images are 500 × 500 nm and are 256 × 256 pixels.

These results illustrate m2py's ability to identify and quantify morphological domains in SPM FDM data. In order to ensure the compatibility of this toolkit with other SPM techniques, m2py

includes a config.py module where users can define the channels and dimensions of their unique SPM measurement. The specific parameters of each technique are defined in a python dictionary within the config.py module, where users can enter the number of channels, the type of information the channels contain (*e.g.* "Young's Modulus" or "Phase"), and the dimensions of their scans. As shown in Figure 8, m2py is already adapted to C-AFM and AMFM images. In Figure 8a, the results of AMFM imaging of P3HT nanowires on passivated ITO glass are shown next to their m2py label maps. On the far right, the domains corresponding to the nanowire phase (phase 1) are plotted in yellow. Figure 8b shows a similar sample of P3HT nanowires imaged through C-AFM next to the corresponding m2py label maps, Because the probe-tip used in C-AFM is far larger than that of AMFM, ~50 nm and ~11 nm respectively, the nanowires appear distorted in Figure 8b. However, we observe in Figures 8a and 8b that the recognized nanowire domains closely resemble the input data. By applying a morphological dilation and subsequent erosion,[65] the m2py toolkit's interpretation is not hindered by these varied probe-tip resolutions, nor by the reduced number of input data channels. To use, expand, or contribute to the m2py toolkit, readers are directed to the open-source Github repository github.com/ponl/m2py. We hope the community will aid in suggesting additions, developing modules, and optimizing the toolkit. We believe that our customizable workflow can augment and expound upon any existing SPM interpretative software and incorporate into any Python-based tools.

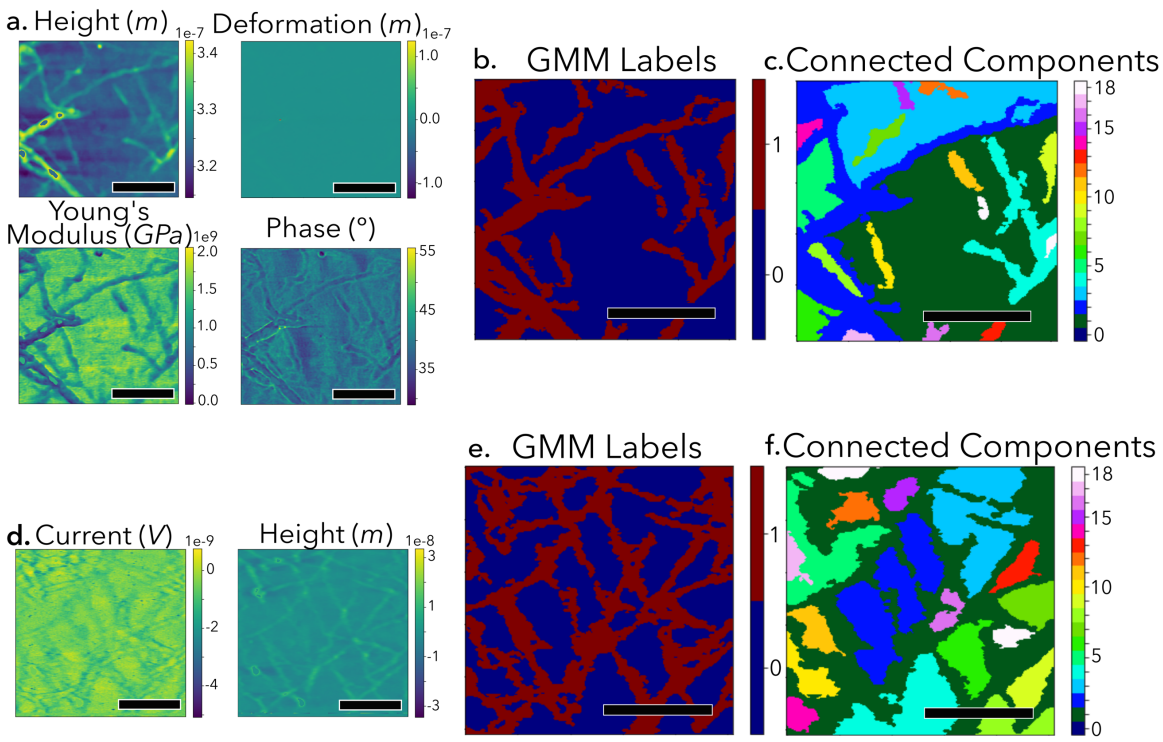


Fig. 8 Input data channels and m2py labeling results for P3HT nanowires spin-coated onto passivated ITO glass. **a.** Property channels of AMFM measurements of P3HT nanowires on ITO glass. Colorbar units are relative. **b.** GMM labels with coloring corresponding to GMM phase labels. **c.** Connected components labels. The color of each individual domain is selected according to its. **d.** Property channels for C-AFM measurements of P3HT nanowires on ITO glass. Colorbar units are relative. **e.** GMM labels with coloring corresponding to GMM phase labels. **f.** Connected components labels. Scale bars are 200 nm, images are 256 × 256 pixels.

Conclusions

The m2py library is a set of modular and adaptable tools for the quantitative description of morphologies from SPM images. By implementing a combination of numerical and unsupervised methods, m2py enhances and isolates material-property signals, which are then used to classify each pixel and cluster them into their corresponding domains. This study has highlighted m2py's

capacity to interpret and identify morphological domains across a wide range of samples of π-conjugated polymers. The resulting labels maps can be used to quantitatively describe the labeled morphologies. Our general workflow is easily modified and allows users to quantitatively evaluate and compare morphological features, domains, and phase distributions in an automated fashion. Finally, additional preprocessing, subsampling, and segmentation methods will be added to improve and expand m2py's morphological interpretations, handle RGB or electron microscopy images, and facilitate their labeling. Future work will report on these additions and seek to understand the connection of such morphological descriptions to device performance will use the m2py framework to quantitatively describe and optimize the morphologies of thin film electronic materials.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

W.K.T. and C.K.L. acknowledge NSF DMR-1708317 and NSF DESE-1633216 for funding. W.K.T., D.T., and P.O. acknowledge MATDAT18 (NSF DMR-1748198) for fostering their collaboration. D.S.G. and L.Q.F. acknowledge support of the AMFM and C-AFM polymer imaging by NSF DMR-1607242. Part of this work was conducted at the Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure site at the University of Washington, which is supported in part by the National Science Foundation (Grant ECCS-1542101), the University of Washington, the Molecular Engineering & Sciences Institute, and the Clean Energy Institute.

References

(1) Mazzio, K. A.; Luscombe, C. K. The Future of Organic Photovoltaics. *Chem. Soc. Rev.* **2015**, *44*, 78–90.

(2) Holliday, S.; Li, Y.; Luscombe, C. K. Recent Advances in High Performance Donor-Acceptor Polymers for Organic Photovoltaics. *Prog. Polym. Sci.* **2017**, *70*, 34–51.

(3) Patel, S. N.; Glaudell, A. M.; Peterson, K. A.; Thomas, E. M.; O'Hara, K. A.; Lim, E.; Chabinyc, M. L. Morphology Controls the Thermoelectric Power Factor of a Doped Semiconducting Polymer. *Sci. Adv.* **2017**, *3*, e1700434.

(4) van Franeker, J. J.; Westhoff, D.; Turbiez, M.; Wienk, M. M.; Schmidt, V.; Janssen, R. A. J. Controlling the Dominant Length Scale of Liquid-Liquid Phase Separation in Spin-Coated Organic Semiconductor Films. *Adv. Funct. Mater.* **2015**, *25*, 855–863.

(5) An, Q.; Ma, X.; Gao, J.; Zhang, F. Solvent Additive-Free Ternary Polymer Solar Cells with 16.27% Efficiency. *Sci. Bull.* **2019**, *64*, 504–506.

(6) Chortos, A.; Liu, J.; Bao, Z. Pursuing Prosthetic Electronic Skin. *Nat. Mater.* **2016**, *15*, 937–950.

(7) Tatum, W. K.; Luscombe, C. K. π-Conjugated Polymer Nanowires: Advances and Perspectives toward Effective Commercial Implementation. *Polym. J.* **2018**, *50*, 659–669.

(8) Dang, M. T.; Hirsch, L.; Wantz, G. P3HT:PCBM, Best Seller in Polymer Photovoltaic Research. *Adv. Mater.* **2011**, *23*, 3597–3602.

(9) Holliday, S.; Donaghey, J. E.; McCulloch, I. Advances in Charge Carrier Mobilities of Semiconducting Polymers Used in Organic Transistors. *Chem. Mater.* **2014**, *26*, 647–663.

(10) Ji, Y.; Xiao, C.; Heintges, G. H. L.; Wu, Y.; Janssen, R. A. J.; Zhang, D.; Hu, W.; Wang, Z.; Li, W. Conjugated Polymer with Ternary Electron-Deficient Units for Ambipolar Nanowire Field-Effect Transistors. *J. Polym. Sci. Part A Polym. Chem.* **2016**, *54*, 34–38.

(11) Flagg, L. Q.; Bischak, C. G.; Onorato, J. W.; Rashid, R. B.; Luscombe, C. K.; Ginger, D. S. Polymer Crystallinity Controls Water Uptake in Glycol Side-Chain Polymer Organic Electrochemical Transistors. *J. Am. Chem. Soc.* **2019**, *141*, 4345–4354.

(12) Persson, N. E.; Rafshoon, J.; Naghshpour, K.; Fast, T.; Chu, P.-H.; McBride, M.; Risteen, B.; Grover, M.; Reichmanis, E. High-Throughput Image Analysis of Fibrillar Materials: A Case Study on Polymer Nanofiber Packing, Alignment, and Defects in Organic Field Effect Transistors. *ACS Appl. Mater. Interfaces* **2017**, *9*, 36090–36102.

(13) Bharti, M.; Singh, A.; Samanta, S.; Aswal, D. K. Conductive Polymers for Thermoelectric Power Generation. *Prog. Mater. Sci.* **2018**, *93*, 270–310.

(14) Yao, H.; Fan, Z.; Cheng, H.; Guan, X.; Wang, C.; Sun, K.; Ouyang, J. Recent Development of Thermoelectric Polymers and Composites. *Macromol. Rapid Commun.* **2018**, *39*, 1700727.

(15) Chen, Y.; Zhao, Y.; Liang, Z. Solution Processed Organic Thermoelectrics: Towards Flexible Thermoelectric Modules. *Energy Environ. Sci.* **2015**, *8*, 401–422.

(16) Onorato, J. W.; Luscombe, C. K. Morphological Effects on Polymeric Mixed Ionic/Electronic Conductors. *Mol. Syst. Des. Eng.* **2019**, *4*, 310–324.

(17) Liu, Y.; Zhao, J.; Li, Z.; Mu, C.; Ma, W.; Hu, H.; Jiang, K.; Lin, H.; Ade, H.; Yan, H. Aggregation and Morphology Control Enables Multiple Cases of High-Efficiency Polymer Solar Cells. *Nat. Commun.* **2014**, *5*, 5293.

(18) Ye, L.; Xiong, Y.; Li, S.; Ghasemi, M.; Balar, N.; Turner, J.; Gadisa, A.; Hou, J.; O'Connor, B. T.; Ade, H. Precise Manipulation of Multilength Scale Morphology and Its Influence on Eco-Friendly Printed All-Polymer Solar Cells. *Adv. Funct. Mater.* **2017**, *27*, 1702016.

(19) Diao, Y.; Shaw, L.; Bao, Z.; Mannsfeld, S. C. B. Morphology Control Strategies for Solution-Processed Organic Semiconductor Thin Films. *Energy Environ. Sci.* **2014**, *7*, 2145–2159.

(20) Cha, H.; Wheeler, S.; Holliday, S.; Dimitrov, S. D.; Wadsworth, A.; Lee, H. H.; Baran, D.; McCulloch, I.; Durrant, J. R. Influence of Blend Morphology and Energetics on Charge Separation and Recombination Dynamics in Organic Solar Cells Incorporating a Nonfullerene Acceptor. *Adv. Funct. Mater.* **2018**, *28*, 1704389.

(21) Greco, C.; Melnyk, A.; Kremer, K.; Andrienko, D.; Daoulas, K. C. Generic Model for Lamellar Self-Assembly in Conjugated Polymers: Linking Mesoscopic Morphology and Charge Transport in P3HT. *Macromolecules* **2019**, *52*, 968–981.

(22) Collins, B. A.; Tumbleston, J. R.; Ade, H. Miscibility, Crystallinity, and Phase Development in P3HT/PCBM Solar Cells: Toward an Enlightened Understanding of Device Morphology and Stability. *J. Phys. Chem. Lett.* **2011**, *2*, 3135–3145.

(23) Onorato, J.; Pakhnyuk, V.; Luscombe, C. K. Structure and Design of Polymers for Durable, Stretchable Organic Electronics. *Polym. J.* **2017**, *49*, 41–60.

(24) Giridharagopal, R.; Flagg, L. Q.; Harrison, J. S.; Ziffer, M. E.; Onorato, J.; Luscombe, C. K.; Ginger, D. S. Electrochemical Strain Microscopy Probes Morphology-Induced Variations in Ion Uptake and Performance in Organic Electrochemical Transistors. *Nat. Mater.* **2017**, *16*, 737–742.

(25) Flagg, L. Q.; Bischak, C. G.; Onorato, J. W.; Rashid, R. B.; Luscombe, C. K.; Ginger, D. S. Polymer Crystallinity Controls Water Uptake in Glycol Side-Chain Polymer Organic Electrochemical Transistors. *J. Am. Chem. Soc.* **2019**, *141*, 4345–4354.

(26) Dong, B. X.; Nowak, C.; Onorato, J. W.; Strzalka, J.; Escobedo, F. A.; Luscombe, C. K.; Nealey, P. F.; Patel, S. N. Influence of Side-Chain Chemistry on Structure and Ionic Conduction Characteristics of Polythiophene Derivatives: A Computational and Experimental Study. *Chem. Mater.* **2019**, *31*, 1418–1429.

(27) Kohn, P.; Huettner, S.; Komber, H.; Senkovskyy, V.; Tkachov, R.; Kiriy, A.; Friend, R. H.; Steiner, U.; Huck, W. T. S.; Sommer, J.-U.; Sommer, M. On the Role of Single Regiodefects and Polydispersity in Regioregular Poly(3-Hexylthiophene): Defect Distribution, Synthesis of Defect-Free Chains, and a Simple Model for the Determination of Crystallinity. *J. Am. Chem. Soc.* **2012**, *134*, 4790–4805.

(28) Snyder, C. R.; Nieuwendaal, R. C.; DeLongchamp, D. M.; Luscombe, C. K.; Sista, P.; Boyd, S. D. Quantifying Crystallinity in High Molar Mass Poly(3-Hexylthiophene). *Macromolecules* **2014**, *47*, 3942–3950.

(29) Kang, S. D.; Snyder, G. J. Charge-Transport Model for Conducting Polymers. *Nat. Mater.* **2017**, *16*, 252–257.

(30) Mazzio, K. A.; Rice, A. H.; Durban, M. M.; Luscombe, C. K. Effect of Regioregularity on Charge Transport and Structural and Excitonic Coherence in Poly(3-Hexylthiophene) Nanowires. *J. Phys. Chem. C* **2015**, *119*, 14911–14918.

(31) Koch, F. P. V.; Rivnay, J.; Foster, S.; Müller, C.; Downing, J. M.; Buchaca-Domingo, E.; Westacott, P.; Yu, L.; Yuan, M.; Baklar, M.; Fei, Zhuping; Luscombe, Christine; McLachlan, Martyn A.; Heeney, Martin; Rumbles, Gary; Silva, Carlos; Salleo, Alberto; Nelson, Jenny; Smith, Paul; Stingelin, Natalie. The Impact of Molecular Weight on Microstructure and Charge Transport in Semicrystalline Polymer Semiconductors–Poly(3-Hexylthiophene), a Model Study. *Prog. Polym. Sci.* **2013**, *38*, 1978–1989.

(32) Tatum, W. K.; Resing, A. B.; Flagg, L. Q.; Ginger, D. S.; Luscombe, C. K. Defect Tolerance of π-Conjugated Polymer Crystal Lattices and Their Relevance to Optoelectronic Applications. *ACS Appl. Polym. Mater.* **2019**, *1*, 1466–1475.

(33) Giridharagopal, R.; Cox, P. A.; Ginger, D. S. Functional Scanning Probe Imaging of Nanostructured Solar Energy Materials. *Acc. Chem. Res.* **2016**, *49*, 1769–1776.

(34) Giridharagopal, R.; Rayermann, G. E.; Shao, G.; Moore, D. T.; Reid, O. G.; Tillack, A. F.; Masiello, D. J.; Ginger, D. S. Submicrosecond Time Resolution Atomic Force Microscopy for Probing Nanoscale Dynamics. *Nano Lett.* **2012**, *12*, 893–898.

(35) Giridharagopal, R.; Ginger, D. S. Characterizing Morphology in Bulk Heterojunction Organic Photovoltaic Systems. *J. Phys. Chem. Lett.* **2010**, *1*, 1160–1169.

(36) Pingree, L. S. C.; Reid, O. G.; Ginger, D. S. Electrical Scanning Probe Microscopy on Active Organic Electronic Devices. *Adv. Mater.* **2009**, *21*, 19–28.

(37) Baghgar, M.; Barnes, M. D. Work Function Modification in P3HT H/J Aggregate Nanostructures Revealed by Kelvin Probe Force Microscopy and Photoluminescence Imaging. *ACS Nano* **2015**, *9*, 7105–7112.

(38) Nguyen, H. K.; Ito, M.; Nakajima, K. Elastic and Viscoelastic Characterization of Inhomogeneous Polymers by Bimodal Atomic Force Microscopy. *Jpn. J. Appl. Phys.* **2016**, *5*, 08NB06.

(39) Killgore, J. P.; DelRio, F. W. Contact Resonance Force Microscopy for Viscoelastic Property Measurements: From Fundamentals to State-of-the-Art Applications. *Macromolecules* **2018**, *51*, 6977–6996.

(40) Wood, D.; Hancox, I.; Jones, T. S.; Wilson, N. R. Quantitative Nanoscale Mapping with Temperature Dependence of the Mechanical and Electrical Properties of Poly(3-Hexylthiophene) by Conductive Atomic Force Microscopy. *J. Phys. Chem. C* **2015**, *119*, 11459–11467.

(41) Passeri, D.; Rossi, M.; Tamburri, E.; Terranova, M. L. Mechanical Characterization of Polymeric Thin Films by Atomic Force Microscopy Based Techniques. *Anal. Bioanal. Chem.* **2013**, *405*, 1463–1478.

(42) Kong, J.; Giridharagopal, R.; Harrison, J. S.; Ginger, D. S. Identifying Nanoscale Structure–Function Relationships Using Multimodal Atomic Force Microscopy, Dimensionality Reduction, and Regression Techniques. *J. Phys. Chem. Lett.* **2018**, *9*, 3307–3314.

(43) Neumayer, S. M.; Collins, L.; Vasudevan, R.; Smith, C.; Somnath, S.; Shur, V. Y.; Jesse, S.; Kholkin, A. L.; Kalinin, S. V.; Rodriguez, B. J. Decoupling Mesoscale Functional

Response in PLZT across the Ferroelectric–Relaxor Phase Transition with Contact Kelvin Probe Force Microscopy and Machine Learning. *ACS Appl. Mater. Interfaces* **2018**, *10*, 42674–42680.

(44) Kalinin, S. V.; Strelcov, E.; Belianinov, A.; Somnath, S.; Vasudevan, R. K.; Lingerfelt, E. J.; Archibald, R. K.; Chen, C.; Proksch, R.; Laanait, N.; Jesse, Stephen. Big, Deep, and Smart Data in Scanning Probe Microscopy. *ACS Nano* **2016**, *10*, 9068–9086.

(45) Somnath, S.; Smith, C. R.; Kalinin, S. V.; Chi, M.; Borisevich, A.; Cross, N.; Duscher, G.; Jesse, S. Feature Extraction via Similarity Search: Application to Atom Finding and Denoising in Electron and Scanning Probe Microscopy Imaging. *Adv. Struct. Chem. Imaging* **2018**, *4*, 3.

(46) Borodinov, N.; Neumayer, S.; Kalinin, S. V.; Ovchinnikova, O. S.; Vasudevan, R. K.; Jesse, S. Deep Neural Networks for Understanding Noisy Data Applied to Physical Property Extraction in Scanning Probe Microscopy. *npj Comput. Mater.* **2019**, *5*, 25.

(47) Li, X.; Dyck, O. E.; Oxley, M. P.; Lupini, A. R.; McInnes, L.; Healy, J.; Jesse, S.; Kalinin, S. V. Manifold Learning of Four-Dimensional Scanning Transmission Electron Microscopy. *npj Comput. Mater.* **2019**, *5*, 5.

(48) Rickman, J. M.; Lookman, T.; Kalinin, S. V. Materials Informatics: From the Atomic-Level to the Continuum. *Acta Mater.* **2019**, *168*, 473–510.

(49) Mukaddem, K. T.; Beard, E. J.; Yildirim, B.; Cole, J. M. ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images. *J. Chem. Inf. Model.* **2019**, acs.jcim.9b00734.

(50) Woolley, R. A. J.; Stirling, J.; Radocea, A.; Krasnogor, N.; Moriarty, P. Automated Probe Microscopy via Evolutionary Optimization at the Atomic Scale. *Appl. Phys. Lett.* **2011**, *98*, 253104.

(51) Stirling, J.; Woolley, R. A. J.; Moriarty, P. Scanning Probe Image Wizard: A Toolbox for Automated Scanning Probe Microscopy Data Analysis. *Rev. Sci. Instrum.* **2013**, *84*, 113701.

(52) Groeber, M. A.; Jackson, M. A. DREAM.3D: A Digital Representation Environment for the Analysis of Microstructure in 3D. *Integr. Mater. Manuf. Innov.* **2014**, *3*, 56–72.

(53) Beck, D. A. C.; Carothers, J. M.; Subramanian, V. R.; Pfaendtner, J. Data Science: Accelerating Innovation and Discovery in Chemical Engineering. *AIChE J.* **2016**, *62*, 1402–1416.

(54) Somnath, S.; Smith, C. R.; Jesse, S.; Laanait, N. Pycroscopy - An Open Source Approach to Microscopy and Microanalysis in the Age of Big Data and Open Science. *Microsc. Microanal.* **2017**, *23*, 224–225.

(55) Somnath, S.; Smith, C. R.; Laanait, N.; Vasudevan, R. K.; Ievlev, A.; Belianinov, A.; Lupini, A. R.; Shankar, M.; Kalinin, S. V.; Jesse, S. USID and Pycroscopy -- Open Frameworks for Storing and Analyzing Spectroscopic and Imaging Data. *arXiv* **2019**.

(56) Nečas, D.; Klapetek, P. Gwyddion: An Open-Source Software for SPM Data Analysis. *Open Phys.* **2012**, *10*, 181–188.

(57) Virtanen, Pauli; Gommers, Ralf; Oliphant, Travis E.; Haberland, Matt; Reddy, Tyler; Cournapeau, David; Burovski, Evgeni; Peterson, Pearu; Weckesser, Warren; Bright, Jonathan; van der Walt, Stéfan J.; Brett, Matthew; Wilson, Joshua; Millman, K. Jarrod;

Mayorov, Nikolay; Nelson, Andrew R. J.; Jones, Eric; Kern, Robert; Larson, Eric; Carey, CJ; Polat, Ílhan; Feng, Yu; Moore, Eric W.; Vander-Plas, Jake; Laxalde, Denis; Perktold, Josef; Cimran, Robert; Henrikson, Ian; Quintero, E. A.; Harris, Charles R.; Archibald, Anne M.; Ribeiro, Antônio H.; Pedregosa, Fabian; van Mulbregt, Paul; and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **2020**, 17, 261-272.

(58) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(59) Takahashi, K.; Yamagishi, T.; Aoyagi, S.; Aoki, D.; Fukushima, K.; Kimura, Y. Principal Component Analysis Image Fusion of TOF-SIMS and Microscopic Images and Low Intensity Secondary Ion Enhancement by Pixel Reduction. *J. Vac. Sci. Technol. B* **2018**, *36*, 03F113.

(60) Seifrid, M. T.; Oosterhout, S. D.; Toney, M. F.; Bazan, G. C. Kinetic Versus Thermodynamic Orientational Preferences for a Series of Isomorphic Molecular Semiconductors. *ACS Omega* **2018**, *3*, 10198–10204.

(61) Rinehart, S. J.; Yuan, G.; Dadmun, M. D. Elucidating the Kinetic and Thermodynamic Driving Forces in Polymer Blend Film Self-Stratification. *Macromolecules* **2018**, *51*, 7836–7844.

(62) Na, J. Y.; Kang, B.; Lee, S. G.; Cho, K.; Park, Y. D. Surface-Mediated Solidification of a Semiconducting Polymer during Time-Controlled Spin-Coating. *ACS Appl. Mater.*

Interfaces **2017**, *9*, 9871–9879.

(63) Ruan, C. "Skin-Core-Skin" Structure of Polymer Crystallization Investigated by Multiscale Simulation. *Mater. (Basel, Switzerland)* **2018**, *11*, 610.

(64) Young, T. J.; Monclus, M. A.; Burnett, T. L.; Broughton, W. R.; Ogin, S. L.; Smith, P. A. The Use of the PeakForce ™ Quantitative Nanomechanical Mapping AFM-Based Method for High-Resolution Young's Modulus Measurement of Polymers. *Meas. Sci. Technol.* **2011**, *22*, 125703.

(65) van der Walt, Stefan; Schönberger, Johannes L.; Nunez-Iglesias, J.; Boulogne, François; Warner, Joshua D.; Yager, Neil; Gouillart, Emmanuelle; Yu, Tony; and the scikit-image contributors. scikit-image: Image Processing in Python. *PeerJ*. 2014, p e453.

(66) B.T.M., R. J.; Arnold, M. The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. *Fundam. Informaticae* **2000**, *41*, 187–228.

(67) Edelsbrunner, H.; Harer, J. *Computational Topology - an Introduction*.; American Mathematical Society, 2010.

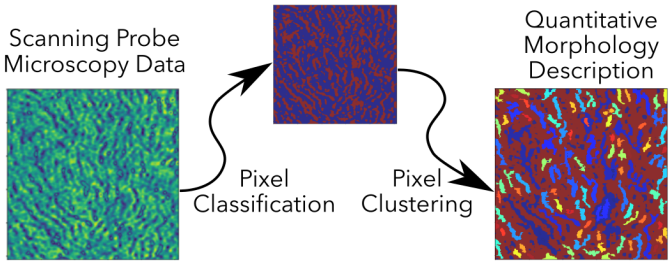
(68) Noriega, R.; Rivnay, J.; Vandewal, K.; V Koch, F. P.; Stingelin, N.; Smith, P.; Toney, M. F.; Salleo, A. A General Relationship between Disorder, Aggregation and Charge Transport in Conjugated Polymers. *Nat. Mater.* **2013**, *12*, 1038–1044.

(69) Noriega, R. Efficient Charge Transport in Disordered Conjugated Polymer Microstructures. *Macromol. Rapid Commun.* **2018**, *39*, 1800096.

(70) Mollinger, S. A.; Krajina, B. A.; Noriega, R.; Salleo, A.; Spakowitz, A. J. Percolation, Tie-

Molecules, and the Microstructural Determinants of Charge Transport in Semicrystalline Conjugated Polymers. *ACS Macro Lett*. **2015**, *4*, 708–712.

For Table of Contents Use Only



A Generalizable Framework for Algorithmic Interpretation of Thin Film Morphologies in Scanning Probe Images

Wesley K. Tatum,[‡,a] Diego Torrejon,[‡,b,c] Patrick O'Neil,[‡,b,c] Jonathan W. Onorato,[a] Anton B. Resing,[a] Sarah Holliday,[a] Lucas Q. Flagg,[d] David S. Ginger,[d] and Christine K. Luscombe[,a,d,e].*