

**ANALYSIS OF FIBRILLAR STRUCTURES FOR THE
ENGINEERING OF POLYMERIC TRANSISTORS**

A Dissertation
Presented to
The Academic Faculty

by

Nils Erland Persson

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
December 2017

COPYRIGHT © 2017 BY NILS PERSSON

ANALYSIS OF FIBRILLAR STRUCTURES FOR THE ENGINEERING OF POLYMERIC TRANSISTORS

Approved by:

Dr. Martha A. Grover, Advisor
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. F. Joseph Schork
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. Elsa Reichmanis
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Dr. Jye-Chyi Lu
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. J. Carson Meredith
School of Chemical & Biomolecular
Engineering
Georgia Institute of Technology

Date Approved: July 31, 2017

*To the late Erland Persson,
a man both smart and sincere;
the only grandfather I ever did know,
and a hell of an engineer.*

ACKNOWLEDGEMENTS

It could be said that my PhD began in the basement of Ormsby's Tavern on 14th and Howell Mill the evening of Saturday, March 9th, 2013, while discussing the origins of life over dinner with Professor Martha Grover during Georgia Tech's recruitment weekend. Martha is endlessly curious, relentlessly discerning, and one of the few people I trust to provide me with sincere critical feedback when I need it most. I extend my deepest gratitude to the best advisor anybody could ask for.

Professor Elsa Reichmanis signed on as my co-advisor shortly after my arrival to Tech in August, and her incredible group has provided me a home for the last four years. Elsa's steady hand has guided me through many of the more byzantine machinations of academia – when she says, “don't worry about it!”, *I do not worry about it*. I extend my sincere gratitude to her as well.

I am indebted to Professor Carson Meredith, who played a significant role in shaping the final year of my thesis work, as well as my career aspirations. I extend my thanks to Professors J.C. Lu and F. Joseph Schork, as well, who brought a diverse spectrum of ideas and insights to my studies.

My research career began, however, under the advisory of Dr. Samuel Blass and Profs. Lanny Schmidt and Aditya Bhan at the University of Minnesota (rah!). Without their mentorship during my transitional year following undergrad, and their resounding endorsement of Georgia Tech, I would not be writing this today. Thank you, Sam, Lanny, and Aditya.

Research is a team game, and consequently I have many colleagues to thank. First, to Mike McBride, a guy who I always want on my team, whether it's research or basketball: my watch has ended; you must hold the keep in my stead.

To all the members of the Reichmanis group, past and present, but especially to Dalsu and Mincheol, for showing me the ropes; to Nabil and Ping-Hsun, for always being two steps ahead of me; and to Gang and Guoyan, for including me in your incredibly innovative projects: thank you all.

Concordantly, to all the members of the Grover group, whose seemingly disparate projects actually display emergent, thought-provoking properties when (self-)assembled as a whole. More than anything, you all have shown me how to think. Special thanks to Xun, for taking me to Dim Sum and teaching me the “most important” Chinese words, and to Dan, for being open and expressive in your communications as I searched for a group before coming to Tech.

Josh, Kaylie, Rrahul, and Max: you were very brave undergrads to agree to work for me all these years. Very brave, indeed. Thank you for trusting me, even when I appeared to have no idea what I was doing.

To Dr. Tony Fast, Professor Surya Kalidindi, and the students in the FLAMEL program: thank you for stepping boldly into the nascent field of materials data science with me. The FLAMEL program has fundamentally defined the direction of my career, and I found everything we did to be – and I mean this sincerely – really, really cool. That goes doubly for my work with Robert and Perry: thank you both for entertaining my entrepreneurial ambitions for as long as you did.

What would I be without my parents? I suppose, technically, an uncollapsed waveform of consciousness still searching the universe for a bodily host, but beyond that, probably a college dropout. I extend my gratitude to my mom, Tina, for instilling in me a healthy work ethic and life-sustaining nutritional habits, and my dad, Eric, for showing me what it means to be a thoughtful, caring, and productive man of good taste and humor. Also, the food. *Thank you for the food.*

A big thank you to my little sister, Lisa, who, besides being way cooler than me, did not hesitate to fly out to Atlanta on a moment's notice to help me through a tough spot. I am always here for you, Lisa. No matter what.

Många Tack, MorMor, my favorite member of the family, for constantly reminding me that what I work on is, for many people, the stuff of science fiction and dreams.

To my friend and roommate, Kong, who has probably (along with Mara) borne the brunt of my frustration with grad school: you're there when I'm up, you're there when I'm down, you're ... actually you're in the shower right now. Thanks for being there, wherever *there* happens to be.

To Kevin, Wei, Andrew, Christine, Jeff and Tim: you're the best friends ever, tbh.

I have saved the most important for last. Midway through my first year, I met my girlfriend, Mara, an amazing human being who has strengthened my character and provided me direction, and whose work as a real, medical doctor makes my "doctoral" work look easy. I am so lucky that you and your family have welcomed me into your life, and I look forward to our adventures in Baltimore. Thank you for being unconditionally interested in, concerned with, and supportive of my research and well-being.

Finally, to Olive: one look at your smooshy face, and all my troubles melt away.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF SYMBOLS AND ABBREVIATIONS	xx
ASSOCIATED SOFTWARE	xxi
SUMMARY	xxii
Chapter 1. Introduction	1
1.1 Solution-processed Organic Electronics	1
1.1.1 Applications	1
1.1.2 Charge Carrier Mobility in the Organic Field Effect Transistor	2
1.2 Conjugated Semiconducting Polymers	5
1.2.1 Why Poly(3-hexylthiophene)?	5
1.2.2 Nanofibers and Charge Transport in P3HT	6
1.3 Processing P3HT Nanofibers	8
1.3.1 The Cambrian Explosion, P3HT-style	8
1.3.2 Studying P3HT Processing on a Standard Device Platform	10
1.3.3 Structural Characterization of P3HT	11
1.4 Extracting Fibers from Images	14
1.4.1 Classification, Segmentation, and Measurement	14
1.4.2 The Beauty of Open Source	17
1.4.3 GTFiber	19
1.5 Overview of Chapters	20
Chapter 2. Automated Analysis of Orientational Order in Polymer-nanofiber-based Organic Transistors	22
2.1 Introduction	22
2.2 Methods	25
2.2.1 Overview of Orientational Order Extraction	25
2.2.2 Calculating the Decay of Orientational Order	29
2.2.3 Obtaining the Orientation Map through Image Processing	32
2.3 Results	40
2.3.1 The Effect of Solution Processing on Morphology	40
2.3.2 The Effect of Deposition Method on Morphology	44
2.3.3 Structure-Property Relationships	47
2.3.4 Extensibility of the Relationships	51
2.4 Conclusions, Chapter 2	53

Chapter 3. Packing, Defects, and Growth Mechanisms in Conjugated Polymer Nanofibers Evaluated by High-throughput Image Analysis	55
3.1 Introduction	55
3.2 Materials and Methods	59
3.2.1 Materials	59
3.2.2 Solution Processing	59
3.2.3 OFET Device Fabrication	60
3.2.4 Electrical Characterization	61
3.2.5 Atomic Force Microscopy	61
3.3 Image Data and Analysis	62
3.3.1 Solution Processing and Thin Film Deposition	62
3.3.2 Image Processing	64
3.3.3 Structural Visualizations and Order Parameters	66
3.3.4 Interactive Exploration of Materials Image Datasets	70
3.4 Results and Discussion	72
3.4.1 Process-Structure-Property Relationships	72
3.4.2 The Role of Fiber Length	75
3.4.3 The Mechanistic Origins of Fiber Curvature	78
3.5 Meso-Scale Simulation of Polymer Packing	81
3.5.1 Simulation Details	82
3.5.2 Molecular Weight, Fiber Length Density, and Tie Chains	86
3.5.3 Transferability of Results and Future Work	89
3.6 Conclusions, Chapter 3	90
Chapter 4. Structural Similarity Metrics for Materials Image Repositories	92
4.1 Introduction	92
4.2 Methods	95
4.2.1 Database Assembly, Cleaning, and Tagging	95
4.2.2 Feature Generation: Spatial Statistics	98
4.2.3 SIFT Features	100
4.2.4 VLAD Encoding	101
4.2.5 Convolutional Neural Net Representations	102
4.2.6 Classification, Dimensionality Reduction, and Reverse Image Search	104
4.3 Results and Discussion	105
4.3.1 Classification Performance	105
4.3.2 2-D Embedding and Library Visualization	107
4.3.3 Reverse Image Search	109
4.3.4 Automated Image Analysis Dispatching	112
4.4 Conclusions, Chapter 4	114

Chapter 5. Conclusions and Future Work	115
5.1 Conclusions	115
5.1.1 Process-Structure-Property Relationships in P3HT-based OFETs	115
5.1.2 Image Analysis Strategies	117
5.1.3 Materials Informatics Infrastructure	118
5.2 Suggested Future Work	120
5.2.1 Leveraging IBM Watson for Initial Experimental Design	120
5.2.2 Generative adversarial neural nets for microstructure segmentation	122
5.2.3 In situ characterization of OFET processing and properties	124
APPENDIX A. GTFiber 1.0	126
A.1. How to Spatially Average Orientations	126
A.2. Coherence-enhancing Anisotropic Diffusion Filtering	128
A.2.1 Algorithmic Details	128
A.2.2 Effect of Orientation Smoothing	131
A.2.3 Effect of Diffusion Time	133
A.3. Top Hat Filtering, Thresholding, and Skeletonization	133
A.3.1 Top Hat Filtering	133
A.3.2 Thresholding Choices	135
A.3.3 Skeletonization	137
A.3.4 Other Imaging Systems	138
APPENDIX B. GTFiber 2.0	140
B.1. Fiber Vectorization	142
B.1.1 Segment Vectorization	143
B.1.2 Fiber Reconstruction	147
B.2. Order Parameter Calculation	152
B.3. Accuracy and Sensitivity Analysis	156
B.4. User Interface and Parameter Selection	163
APPENDIX C. The OFET Database	166
C.1. Databases Drive Reproducibility	166
C.2. Database Curation	169
C.3. Results	171
C.3.1 Mobility versus Molecular Weight Distribution	171
C.3.2 Mobility versus Channel Length	173
C.3.3 Mobility versus Deposition Method	175
C.3.4 Identifying a Standard Device Platform	177
C.4. Commentary on Materials Processing Databases	180
Chapter 6. REFERENCES	181

LIST OF TABLES

Table 1 – Top performing devices by processing regimen. ^a Polydispersity Index. ^b Regioregularity. ^c Initial concentration of polymer solution. ^d Deposition method: Spin = spin coated; Slide = slide coated; Blade = blade coated. ^e Channel length. ^f 2-methylpentane.	48
Table 2 – Classification performance on artifacts and fibers with SIFT and CNN features.	106
Table 3 – Calculating average orientation with the structure tensor.	127
Table 4 – Accuracy of GTFiber in four different cases, using the same parameters, compared with manual tracing.	158
Table 5 – Sensitivity of GTFiber parameters: percent change in each structural metric per percent change in each algorithm parameter.	160
Table 6 – Guide to parameter selection and tuning.	165
Table 7 – Processing conditions for 5 devices conforming to the standard defined above. ^a Regioregularity. ^b Initial concentration of P3HT in solution. ^c Hexamethyldisilazane.	178

LIST OF FIGURES

Figure 1 – Roll-to-roll processed organic electronic devices.	2
Figure 2 – Bottom gate, bottom contact OFET architecture, energy level diagram for hole transport, and extraction of charge carrier mobility from the OFET transfer curve.	3
Figure 3 – Above: P3HT, π - π stacking, and the structure of the P3HT nanofiber. Below: An early theory of charge transport in P3HT thin films, claiming that inter-grain tie chains were more important for charge transport than isolated crystalline domains.	7
Figure 4 – Reported mobility of devices satisfying progressively tighter constraints in the OFET database. The constraint in each column stacks on all the previous constraints. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.	9
Figure 5 – Nucleation, growth and deposition of P3HT nanofibers through various solution processing techniques, with accompanying AFM images (2 μ m) showing the resulting fibrillar morphologies.	11
Figure 6 – Structural characterization of P3HT thin films with examples of raw data, model parameters, and their physical significance.	12
Figure 7 – Characterization of P3HT nanofibers <i>via</i> AFM, with representative images provided at different length scales.	16
Figure 8 – Open source software that contributed to GTFiber. ^{71,74} CellProfiler, ADABlock, and FIRE contained strategies for the automated skeletonization of fibers, while FiberApp contained algorithms for manual vectorization and structural analysis. GTFiber unites these algorithms to provide automated fiber vectorization and analysis.	18
Figure 9 – Main results of the protocol. (A) A cropped version of the phase channel of a tapping mode AFM image. (B) False color Orientation Map extracted from the original image. Each pixel's orientation corresponds to an orientation on the attached color wheel. (C) Orientation Distribution extracted from the Orientation Map. The radial axis indicates the count of pixels of a given orientation. A diametrical black line segment indicates the average	27

orientation, and the full-frame value of $S2D$ is indicated at bottom left. (D) Decay of the orientational order parameter, $S2D$, as a function of frame size. Fitted model parameters are indicated at upper right.

Figure 10 – Calculation of $S2D$ decay function from a skeletonized image. Each point on the graph represents the average value of $S2D$ calculated from evenly spaced frames in the image. The grid step determines the spacing between each frame, which can create sparsely distributed frames or highly overlapped frames. The frame step determines the level of discretization of the x-axis (frame size). White dots in the grid step schematic indicate the centers of frames. 30

Figure 11 – Graphical representation of four extrema in the space defined by λC and $Sfull$. (A) Low $Sfull$, high λC (B) high $Sfull$, high λC (C) low $Sfull$, low λC (D) high $Sfull$, low λC . 31

Figure 12 – Outline of image processing workflow used to obtain the skeleton and subsequent Orientation Map. 34

Figure 14 – Graphical User Interface for the image analysis methods introduced here, written in MATLAB and compiled as a standalone application. 39

Figure 15 – AFM images and Orientation Maps from films spin-coated from three solutions. (A) P3HT in chloroform and 15 vol% 2-methylpentane (5 mg/mL), sonicated two minutes then aged four days. (B) P3HT in chloroform (5 mg/mL), sonicated two minutes then aged four days. (C) P3HT in chloroform (5 mg/mL), processed *via* microfluidic-cooling-UV treatment. Image processing parameters used: Gaussian smoothing, 10nm; orientation smoothing, 30nm; diffusion time, 8s; top hat size, 30nm; adaptive thresholding; noise removal, 2500 nm²; skeleton fringe removal, 40nm. Orientation Maps have been anti-aliased for printing clarity. 41

Figure 16 – Decay of orientational order and schematic of inter-fiber interaction for each of the three morphologies presented in Figure 15. Parameters used for orientational order decay calculation: frame step, 200nm; grid step, 200nm. 42

Figure 17 – The effect of spin coating on fiber alignment. (A) Original tapping mode phase image (10 x 10 μm) from one location. (B) Orientation distributions for accompanying image. (C) Map of average fiber orientation as a function of image location on final spin-coated film (30 x 30 mm). The center of the original droplet is taken to be the origin. Spin direction was counter-clockwise. 45

Figure 18 – The effect of blade coating parameters and solution aging on fiber alignment. (a–d) AFM tapping mode phase images (4 μm) of thin film morphology resulting from slide coating solutions of UV-irradiated P3HT nanofibers at various coating velocities. (e) Image order parameter S_{full} as a function of coating velocity for images a–d. (f) Image order parameter S_{full} as a function of solution aging time in images g–j. (g–j) AFM phase images (5 μm) of thin film morphology resulting from blade coating solutions of P3HT in which nanofibers were formed through UV irradiation and aging. Coating direction was horizontal in all images. 46

Figure 19 – Correlation between mobility and S_{full} , the image-based order parameter for in-plane fiber alignment. (a) Results from sonication, aging, and spin coating study. (b) Results from UV, aging, and blade coating study. (c) Schematic of charge transport in low and high alignment structures. (d) Relationship between P3HT backbone torsion angle and intra-chain mobility from DFT simulations by Lan and Huang. Adapted with permission from Ref. 34. 50

Figure 20 – Processing of P3HT nanofibers and progression of fibrillar morphologies as observed in 5 μm AFM phase images. (a) Illustrations of solution processing and thin film deposition methods considered in this study: UV irradiation and sonication induce nucleation, while aging and poor solvent addition lead to extended growth. Nucleation and growth are combined in a microfluidic-cooling-UV system. Thin film deposition methods considered include spin coating and blade coating. (b) Small fiber nuclei after sonication. (c) Longer fibers formed after aging a sonicated solution for two days. (d) Densely packed fibers formed in the microfluidic process and deposited by spin coating. (e) Highly aligned fibers formed by UV irradiation and two days of aging followed by blade coating at 3 mm/s. 63

Figure 21 – Extraction and vectorization of fibers from an AFM image using GTFiber. (a) A typical AFM phase image of P3HT nanofibers, (b) Result of anisotropic diffusion and top hat filtering, (c) Result of adaptive thresholding, (d) Result of skeletonization and fringe removal, (e) Fitting vectorized contours to the unbranched skeletal segments using Active Contours at a 30 nm step length, (f) Line plot of the vectorized segments, (g) Illustration of segment matching and fiber reconstruction, (h) Line plot of vectorized fibers in which each fiber is assigned a random color to evaluate classification accuracy. 65

Figure 22 – Processed images of fibrillar morphologies and their coordinates in the space defined by the order parameters S_{full} , ρ_{FL} , and λ_C . (a) Schematic illustrating fiber packing arrangements typical of low and high values of the three structural order parameters. (b) The complete structural image library plotted in terms of S_{full} and ρ_{FL} , with solution processes coded by color and deposition methods coded by shape. (c) Image library plotted in terms of S_{full} and λ_C . (d-j) Orientation Maps extracted from raw images at extreme values of each order parameter, with their location in each plot indicated by a gray pointer. Each image is 5 μm , and colors correspond to the in-plane orientation of each fiber's backbone, as indicated by the attached color wheel in d (i.e. red is horizontal, cyan is vertical).

68

Figure 23 – Web-based interactive data visualization for materials image data. (a) Each data point represents one image; as a user zooms in on the plot by scrolling (similar to Google Maps), the underlying image for each data point fades into view. (b) Hovering the mouse over a data marker reveals the structural measurements and processing information for that image. (c) Comparison of structural order parameters and measurements for the two images shown in b. (d) Vectorized fiber plot of left structure in b. (e) Vectorized fiber plot of right structure in b.

71

Figure 24 – Process-structure-property relationships among the image dataset, considering only 5 and 7 μm images. (a-e) Each structural metric across all solution processing conditions, sorted in order of decreasing median values of that structural metric. Each point represents one image. Slight jitter has been introduced to the x-coordinate of data points to improve readability in dense regions. Boxes span from the 25th to the 75th percentile of each point cloud, with a line at the median. (f) Correlation analysis of structural metrics with each other as well as with charge carrier mobility, where data was available. Darker green indicates strong direct correlation, while darker red indicates strong inverse correlation.

73

Figure 25 – The relationship between mean fiber length and fiber alignment across the entire image library. The gray line is included to enhance readability.

76

Figure 26 – (a) Image analysis results from a sample that was UV irradiated, then aged 24h and blade coated, measured at both the air at SiO_2 interface by peeling off the film. (b) Length distributions from each image, fit with log-normal probability density functions. (c-f) Raw AFM images (5 μm each) and colored Orientation Maps produced by GTFiber for each sample at each interface.

77

- Figure 27 – Analysis of P3HT nanofiber curvature and its mechanistic origin. (a) Cropped section of a 2 μm image taken from the same sonicated, aged and blade coated sample as Figure 26c, and a plot of vectorized fiber contours with vertices colored according to their curvature. (b) Cropped section of a 2 μm image taken from the same UV, aged and blade coated sample as Figure 26e, with the same accompanying curvature heat map. (c) Proposed nanofiber growth mechanism and accompanying simulation results. Growth is assumed to occur via a polycondensation-like route. Collisions between growing segments are assumed to occur at a narrow Gaussian distribution of locations along the growth front. Results of nanofiber growth simulations under both centered and distributed collision assumptions, plotted to scale with chain lengths sampled from a log normal distribution corresponding to an M_n of 40 kD and PDI of 2.25. 79
- Figure 28 – Simulation of polymer chain packing along a vectorized fiber backbone. Chain lengths are perpendicular to their parent segment's orientation, and are randomly sampled from a specified molecular weight distribution. Chains are spaced according to the π - π stacking distance of P3HT. 83
- Figure 29 – Visualization of image processing and simulation results. (a) Original 5x5 μm AFM image, phase channel. (b) Orientation Map, in which fibers have been extracted, vectorized, and each segment colored according to its in-plane orientation. (c) Full-scale plot of all simulated polymer chains (501,331) in a 5x5 μm window. (d) Zoomed-in view of (c) with a 500 nm window. 84
- Figure 30 – Quantifying tie chain density for a single fiber. Starting from Figure 29b, the central fiber is made invisible. That fiber's backbone is represented by a black dashed line, while red circles indicate the intersection point along its backbone of any chains from any other fiber in the entire microstructure (notice that even chains from the extreme left fiber are counted). 85
- Figure 31 – The potential for fibers to form tie chains given molecular weight distribution and fiber length density. (a) Tie chain density as a function of fiber length density, simulated for three different number-average molecular weights. Blue: $M_n = 40$ kD, green: $M_n = 25$ kD, red: $M_n = 10$ kD. PDI = 2.25. Fiber cartoons are provided on the x-axis as a visual descriptor of fiber length density. (b) Probability density functions of the three molecular weight distributions considered at a PDI of 2.25. Highlighted in light blue is the range of chain lengths corresponding to the range of inter-fiber spacings similarly highlighted in (a). 87

Figure 32 – Structural classes and artifact classes labeled in the dataset with examples.	98
Figure 33 – Illustration of 2-point autocorrelation matrix (right) for the segmented fibrillar microstructure at left. In the segmented microstructure, a white pixel indicates a fibrillar phase. In the autocorrelation heat map, the color indicates the probability that a vector with coordinates (tx, ty) will have both of its ends contained in a fibrillar phase when placed randomly in the microstructure.	99
Figure 34 – SIFT Feature Generation. Keypoints are identified by circles with radius corresponding to their scale and radial lines indicating their orientation. For each keypoint, an oriented and scaled 4×4 grid of gradients in 8 directions is calculated, yielding a 128-dimensional feature for each keypoint.	101
Figure 35 – VLAD encoding after k-means clustering of all identified image features. Colors represent distinct feature clusters.	102
Figure 36 – Convolutional neural net architecture and feature representation.	103
Figure 37 – Outline of feature extraction, classification and nearest neighbor search.	104
Figure 38 – 2-D embedding of the image database using t-SNE. (Left) Full dataset, (Right) Zoomed-in view of the neighborhood containing most of the fiber images. Blue outlines indicate a label of “fiber” while red indicate “not fiber”.	108
Figure 39 – Performance of reverse image search evaluated on a variety of microstructures, displaying first three results sorted by shortest Euclidean distance in VLAD100-encoded, t-SNE-reduced SIFT space.	111
Figure 40 – Example of dispatched image analysis search result: a query image is uploaded, classified as containing fibers that can be analyzed by GTFiber, and is dispatched for further analysis, including the extraction of orientation and fiber length distributions.	112
Figure 41 – (upper) Dynamic curation of process-property databases with IBM Watson, (lower) Automated extraction of raw data from published figures.	121
Figure 42 – Generative Adversarial Neural Net approach to universal object localization and measurement.	123

Figure 43 – <i>In situ</i> monitoring of device performance during processing.	124
Figure 44 – Diagram of the use of the structure tensor. An $m \times n$ image is converted into an $m \times n \times 4$ image, where each pixel has a corresponding J_{xx} , J_{xy} , and J_{yy} (J_{xy} is repeated and thus redundant). Gaussian filtering can be applied to each layer of the $m \times n \times 4$ image, and then each pixel's structure tensor is reformed, eigenvectors are computed, and a filtered orientation is returned.	128
Figure 45 – The effect of structure tensor smoothing. Standard deviation of orientation smoothing kernel in nm indicated at left, followed by the gray scale result of diffusion filtering, the Orientation Map, and an orientation distribution. Other parameters used for this filter: initial Gaussian smoothing, 10nm; diffusion time, 8s; top hat size, 30nm; adaptive thresholding; noise removal area, 3500 nm ² ; skeleton fringe removal, 40nm.	132
Figure 46 – The effect of the “diffusion time” parameter on diffusion filtering results, from left to right: original gray scale, 1s diffusion time, 3s, 8s. Constant parameter values: initial Gaussian smoothing, 10nm; orientation smoothing, 30nm.	133
Figure 47 – The top hat filtering process, from left to right: result of anisotropic diffusion filtering, result of image erosion, result of image dilation, and result of subtracting the dilated image from the result of anisotropic diffusion filtering.	134
Figure 48 – The effect of the “top hat size” parameter on contrast enhancement, from left to right: initial result of diffusion filtering, result of 10nm top hat filter, 30nm, 50nm. Constant parameter values: initial Gaussian smoothing, 10nm; orientation smoothing, 30nm; diffusion time, 8s.	134
Figure 49 – Thresholding of original image. Left: original gray scale. Center: Automatic global threshold. Right: Adaptive threshold surface using Yanowitz-Bruckstein method. ¹⁸⁷	135
Figure 50 – Thresholding of image after diffusion and top hat filtering. Left: gray scale top hat filter result. Center: global threshold at gray value of 0.3. Right: Noise removed if area < 3500 nm ² .	136
Figure 51 – Thresholding of image after diffusion and top hat filtering. Left: gray scale top hat filter result. Center: result of adaptive thresholding. Right: Noise removed if area < 3500 nm ² .	136

Figure 52 – Skeletonization and trimming. Left: result of adaptive thresholding and noise removal. Center: result of skeletonization. Right: skeleton after removal of branches less than 40nm long.	137
Figure 53 – Skeleton fringe trimming, before and after.	138
Figure 54 – GTFiber applied to: (A) Micro-CT imaging (top) of a fiber composite, (B) SEM imaging of cellulose fibrils (reproduced from Ref. 4 with permission from the Royal Society of Chemistry), ¹³⁴ (C) TEM imaging of fibrillar protein assemblies.	139
Figure 55 – Overview of the GTFiber algorithm stack and data structures.	141
Figure 56 – Segmentation, skeletonization, and segment isolation in GTFiber.	142
Figure 57 – Discretization of a fiber, where s is the path length coordinate and r is the tangent vector at s . F is the stretching force applied to the ends of the contour.	144
Figure 58 – Energy minimization of an active contour and visualizations of the first three stages of GTFiber. (a) Example initialization of an active contour. Red circles: three initialization points. Blue circles: distributed initialization points interpolated between the red initialization points, but with 3-pixel spacing. (b) Energy minimization of the active contour (10 iterations), with $\beta = 5$, $k1 = 20$, and $k2 = 10$. Image gradient vector field overlaid for illustrative purposes. (c) Original grayscale image. (d) Skeletonized image extracted using GTFiber’s image processing operations. (e) Plot of the vectorized version of the skeleton in (d) using Matlab’s <i>line</i> plotting function (with anti-aliasing).	146
Figure 59 – (a) A portion of a raw AFM image. (b) Line plot of the broken, vectorized segments.	148
Figure 60 – Illustration of segment matching and scoring. (a) Candidate endpoints are identified through a projected search kernel. (b) Distances are measured between the searching endpoint and its match endpoints. (c) A test fit of an active contour combining segments $s0$ and $s1$. Curvature at each point is indicated by the color scale at right in units of 1/pixel. Red boxes indicate the points considered for curvature scoring. (d) Test fit of an active contour between segments $s0$ and $s2$.	149
Figure 61 – (a) Original AFM image. (b) Vectorized fibers with curvature indicated by a color scale at each point.	152

Figure 62 – Calculation of S_{full} and λ_c .	153
Figure 63 – Calculation of fiber length density.	154
Figure 64 – Calculation of fiber width distributions.	155
Figure 65 – (a) Representative pictures of flexible white wires on a black fabric background. (b) Histogram of fiber counts from all 50 pictures. (c) Histogram of mean fiber lengths from all 50 pictures; (inset) true fiber length distribution of the wire population.	161
Figure 66 – Problematic cases for GTFiber.	162
Figure 67 – GTFiber user interface and representative visualizations.	164
Figure 68 – Diagram of the process of fabricating and characterizing a P3HT-based OFET device. Below: list of all process variables included in our database.	169
Figure 69 – (A) Hole mobility versus Mn (in kilodaltons) for 218 devices across the literature. Note the log scale on the y-axis. (B) Mobility versus polydispersity index for a subset (146) of the devices in (A). Markers are semi-transparent so that high data densities appear darker.	171
Figure 70 – Mobility versus groups of channel lengths as reported for 206 devices in the literature. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.	173
Figure 71 – Mobility versus solution deposition method for 204 devices in the literature: dip-coated, drop-cast, or spin-coated. Only devices with gold electrodes are considered in this search. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.	175
Figure 72 – Reported mobility of devices satisfying progressively tighter constraints. The constraint in each column stacks on all the previous constraints. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.	177

LIST OF SYMBOLS AND ABBREVIATIONS

AFM	Atomic Force Microscopy
CED	Coherence-Enhancing Diffusion
CNN	Convolutional Neural Network
HOMO	Highest Occupied Molecular Orbital
I_D	Drain Current
λ_C	Decay Length of Orientational Order
MGI	Materials Genome Initiative
M_n	Number-Average Molecular Weight
M_w	Weight-Average Molecular Weight
MW	Molecular Weight
μ	Charge Carrier Mobility
OFET	Organic Field Effect Transistor
P3HT	Poly(3-hexylthiophene)
PDI	Polydispersity Index
ρ	Orientation Smoothing Parameter
ρ_{FL}	Fiber Length Density
ρ_{tie}	Tie Chain Density
S_{full}	Full-image Alignment Parameter
SIFT	Scale-Invariant Feature Transform
t-SNE	t-Distributed Stochastic Neighbor Embedding
SVM	Support Vector Machine
V_{DS}	Drain-Source Voltage
V_G	Gate Voltage
VLAD	Vector of Locally Aggregated Descriptors

ASSOCIATED SOFTWARE

Name	Permalink to Code
GTfiber 2.0	https://gtfiber.github.io/
GTfiber 1.1	www.github.com/Imperssonator/GTFiber-Mac www.github.com/Imperssonator/GTFiber-Windows
OFET Database	www.github.com/Imperssonator/OFET-Database
ZoomImgs	https://zoomimgs.github.io

SUMMARY

At its core, this thesis addresses the fundamental question of how to quantify the structure of packed semi-flexible fibers in solid materials. This is accomplished through the introduction of an open source software package, GTFiber, for the automated extraction and analysis of fibers from images of poly(3-hexylthiophene) (P3HT) nanofibers.

P3HT is one of many *conjugated polymers*: polymers that possess (a) semiconducting properties, making them candidates for use in electronic devices, and (b) solubility in organic solvents, making them solution processable (*i.e.* printable from ink). P3HT crystallizes into nanofibers that enhance the charge carrier mobility of thin film transistors through remarkably complex process-structure-property relationships. To understand and control mobility, these relationships must be quantified and modeled, which necessarily proceeds through the quantification of the long-range fibrillar structure.

First and foremost, this thesis demonstrates a rigorous protocol for the analysis of P3HT nanofibers from Atomic Force Microscopy images. Using this protocol, the relationships between fiber crystallization, thin film deposition, structure, and charge transport are elucidated, and a correlation is demonstrated between fiber alignment and mobility. Additionally, processing methods are analyzed in terms of their aligning mechanisms, revealing that inter-fiber connectivity, fiber length, and depositional shear forces play crucial, interrelated roles in alignment and charge transport. Finally, the future of image analysis in materials science is explored through a large image database using modern computer vision techniques, demonstrating a new paradigm for the discovery and analysis of microstructural data.

Chapter 1. Introduction

1.1 Solution-processed Organic Electronics

Conjugated polymers are expected to give rise to a new ecosystem of large-area electronics manufacturing based on solution-processable semiconducting inks. Semiconductors are the enabling material behind transistors, which have delivered unbelievable technological advances since their development by Bell labs in 1947. Progressively smaller circuits with exponentially higher transistor densities have driven the production of mainframe computers, personal computers, laptops, smartphones, and now smart watches. Behind these advancements is the microfabrication industry, rooted in the processing of crystalline silicon. However, small, fragile wafers of crystalline silicon must be handled in a cleanroom environment and processed at high temperatures, precluding the application of crystalline silicon to devices requiring large-area deposition and high-throughput manufacturing.

1.1.1 Applications

In the late 1980s, the semiconducting properties of soluble polyalkylthiophenes were first reported,^{1,2} suggesting that semiconducting polymers could be dissolved in solutions that could be coated over large areas, enabling the production of so-called “macroelectronics.” This led to significant interest in roll-to-roll processing, a thin film deposition technique that would allow electronic devices to be printed like newspapers, as illustrated in **Figure 1**.^{3,4} In the ideal case, the substrate, electrodes, semiconductor, and

encapsulation materials could all be printable, transparent, and flexible, opening the door to a wide variety of potential applications. These include, but are not limited to: flexible solar cells;^{5,6} large-area and flexible displays, including e-papers,⁷ active matrix liquid crystal (LCD) and organic light emitting diode (OLED) displays,⁸⁻¹¹ radio-frequency identification (RFID) tags,^{12,13} sensors for biomedical applications such as electronic skins,^{14,15} and devices for control and sensing in soft robotics.^{16,17}

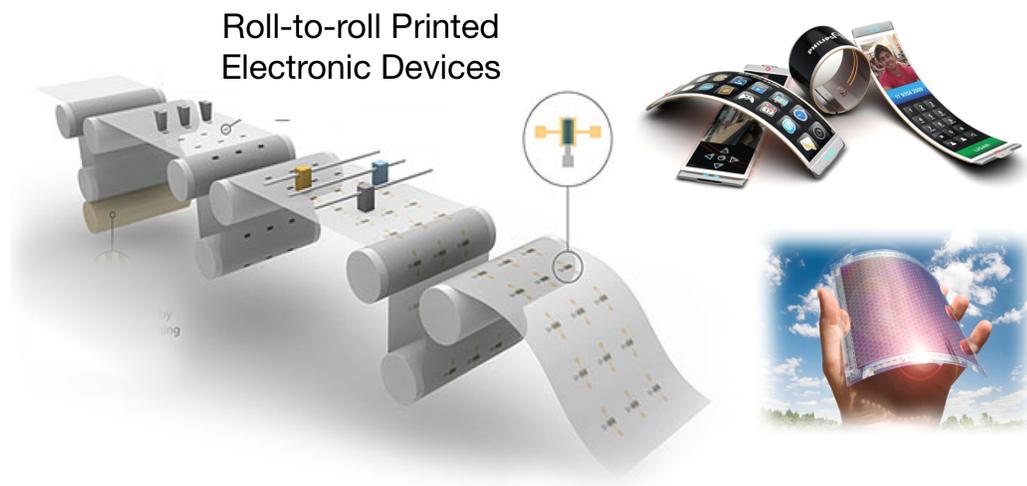


Figure 1 – Roll-to-roll processed organic electronic devices.ⁱ

1.1.2 Charge Carrier Mobility in the Organic Field Effect Transistor

The fundamental building block underlying many of these futuristic technologies is the organic field effect transistor (OFET), illustrated in **Figure 2**. They are present in control circuits for display pixels, logic circuits in RFID tags, and can function as sensors, as well.¹⁸ An OFET is essentially a switch with “on” and “off” states. Current flows from the source to the drain through the *semiconducting channel*. Here, we consider a *p*-type

ⁱ Images adapted from MURI, University of Minnesota; Android Authority

semiconductor in which the charge carriers are holes, indicated by the positive charge. In the “off” state, holes are unable to travel through the semiconductor because its Highest Occupied Molecular Orbital (HOMO) level is inaccessible from those of the electrodes. In the “on” state, an electric field is applied perpendicular to the current flow, bringing the HOMO level of the semiconductor into alignment with that of the electrodes, allowing for charge injection and transport. Thus, the gate voltage, V_G , determines whether or not a current, I_D , can flow from the source to the drain electrode.

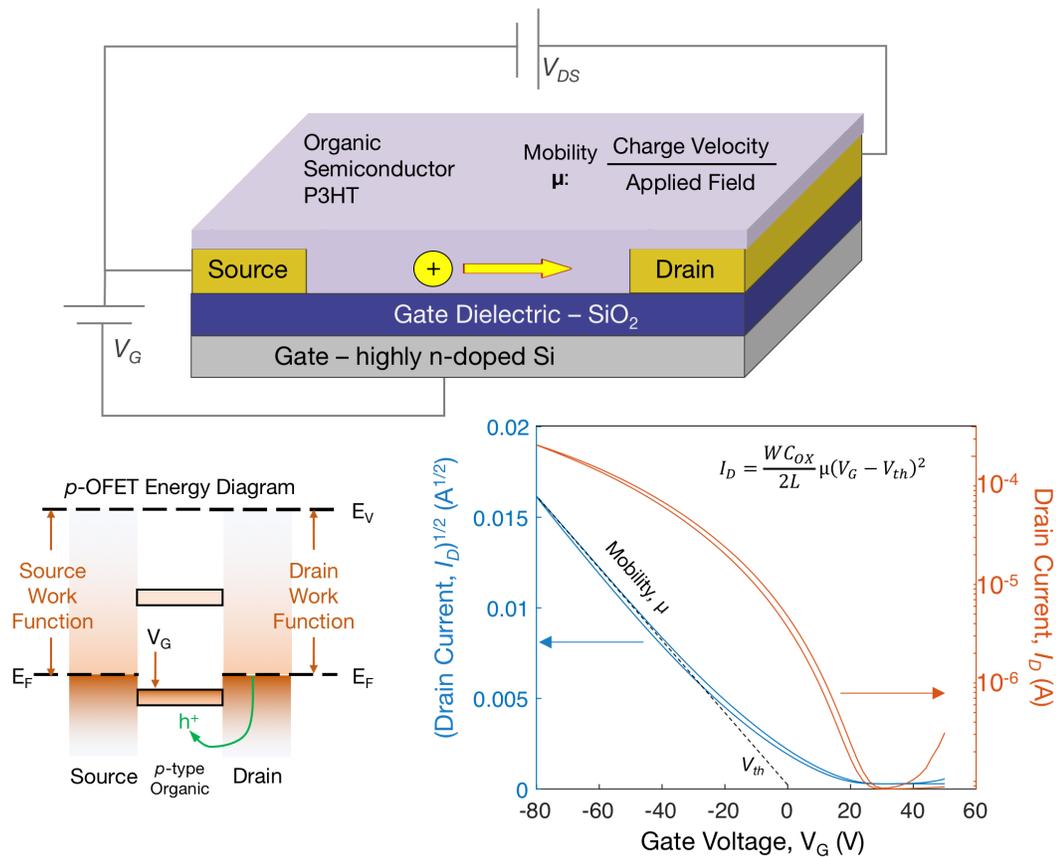


Figure 2 – Bottom gate, bottom contact OFET architecture, energy level diagram for hole transport, and extraction of charge carrier mobility from the OFET transfer curve.

An OFET is characterized by its *transfer curve*, plotted at right in **Figure 2**: with source-drain voltage held constant, V_G is swept from high to low and back again while

measuring I_D . The relationship between these quantities is modeled by the following equation, referred to as the *saturation regime* model:

$$I_{DS} = \frac{WC_{OX}}{2L} \mu (V_G - V_{th})^2 \quad (1)$$

where W and L are the width and length of the channel bounded by the source and drain electrodes, C_{OX} is the capacitance of the gate dielectric material, and μ and V_T are the **charge carrier mobility** and **threshold voltage**, which are fitted model parameters. While both of these parameters are crucial to OFET operation in real-world environments, mobility has attracted the most research interest because of its relevance to application performance and its strong but complex relationship with material chemistry and microstructure.^{19,20}

Mobility is defined as the drift velocity of a charge carrier in a material per the strength of applied electric field and has units of cm^2/Vs . In most cases, a higher mobility is better, enabling higher transistor switching rates and the control of stronger currents. In solar cells, mobility must be high enough to enable charge migration to electrodes, but not so high as to promote charge recombination.²¹ A mobility of $0.1 \text{ cm}^2/\text{Vs}$ is required for applications such as *e*-papers, while a mobility of $1 \text{ cm}^2/\text{Vs}$ would enable the use of organic transistors in the backplane control circuitry of liquid crystal displays. For the control of organic light emitting diode displays, it is estimated that a value of at least $10 \text{ cm}^2/\text{Vs}$ would be required. In comparison, the mobility of crystalline silicon for microprocessor applications is in the 100s to 1000s of cm^2/Vs .²²

1.2 Conjugated Semiconducting Polymers

1.2.1 Why Poly(3-hexylthiophene)?

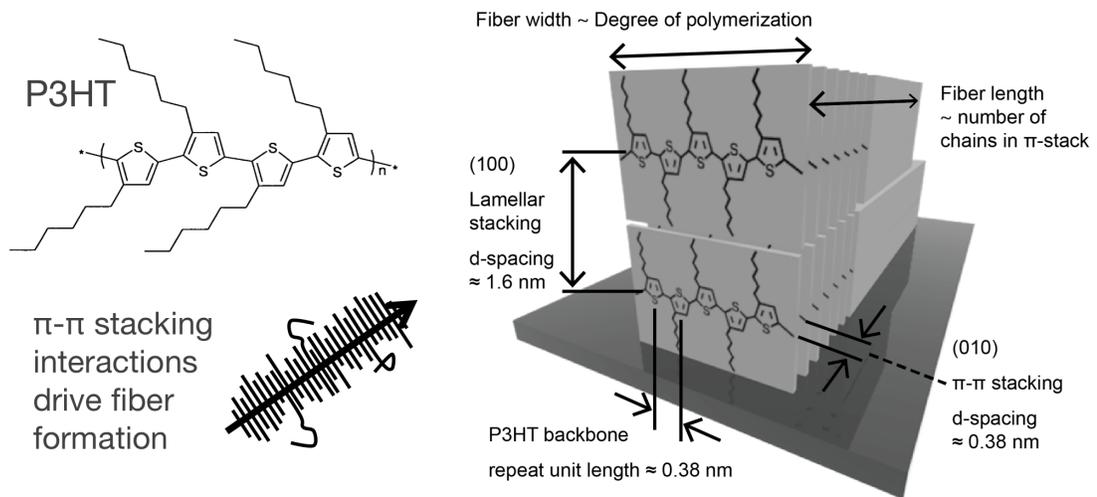
Organic semiconductors fall into two major classes: small molecules and polymers. While small molecules are capable of achieving mobilities greater than $10 \text{ cm}^2/\text{Vs}$ as bulk single crystals, polymer semiconductors are generally more processable, making them better candidates for layer-by-layer large-area printing.²³ Their rheological properties are highly tunable, facilitating greater control of flow in inks; they produce very smooth films, making deposition of further device layers easier; their solubility window is relatively narrow, enabling the selection of a wider variety of orthogonal solvents for subsequent material layers; and they have negligible vapor pressure, limiting inter-layer diffusion during heat treatment steps common to device processing.²⁴

Poly(3-hexylthiophene) (P3HT), the molecular structure of which is presented in **Figure 3**, has established itself as the canonical semiconducting polymer since its popularization in the early 2000s.^{25–28} While considerable efforts have been placed on the design and synthesis of new, often complex molecular structures with intrinsically higher mobility, P3HT has remained a source of continual research interest for the study of conjugated polymer process-structure-property relationships.^{23,29,30} This is because of its commercial availability, unique crystallization behavior, and the considerable momentum behind the theoretical treatment of its charge transport behavior.³¹

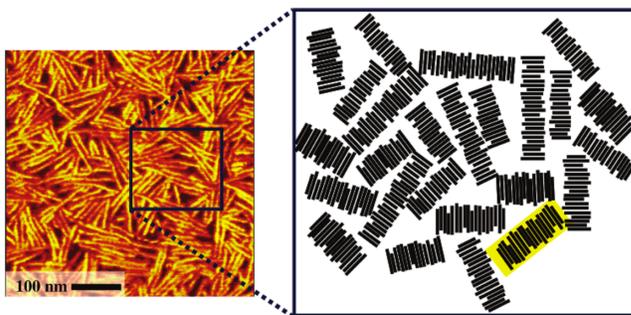
1.2.2 Nanofibers and Charge Transport in P3HT

While the hexyl side chains of P3HT lend it its solubility, the π -conjugation of the thiophene rings along P3HT's backbone yields many of its remarkable properties. The delocalized energetic states facilitate rapid charge transport, and also lead to physical π - π stacking interactions when chains are stacked co-facially, as depicted in **Figure 3**. The P3HT molecule in **Figure 3** is a *regioregular* isomer: the side chains are evenly spaced, all falling to the left of the sulfur atoms in their respective thiophene rings. Regioregular P3HT is known to form nanofibrillar structures through π - π stacking.³² ***The length of the nanofiber is perpendicular to the polymer chain backbones.*** This point cannot be stressed enough.

The most important aspects of P3HT nanofiber structure are highlighted in **Figure 3**. Stacking distance between polymer chains is 0.38 nm, and stacks – comprised of thousands of polymer chains – can grow to be microns in length. P3HT also stacks vertically, referred to as lamellar stacking. This lattice plane has a spacing of 1.6 nm, and the insulating effect of the hexyl side chains makes it practically irrelevant for charge transport.³³ By contrast, charge transport along planarized P3HT chains is theorized to have a mobility of up to 1 cm²/Vs, while transport along π - π stacks is governed by an activated charge hopping mechanism and thus has slightly lower mobility of 0.01 – 0.1 cm²/Vs.³⁴ For this reason, P3HT nanofibers are frequently drawn as viewed from above, with line segments representing the chains of the π - π stack, ignoring the lamellar stacking dimension.



$M_n = 3 \text{ kD}$
 $\mu = 10^{-4} \text{ cm}^2/\text{Vs}$



$M_n = 30 \text{ kD}$
 $\mu = 10^{-2} \text{ cm}^2/\text{Vs}$

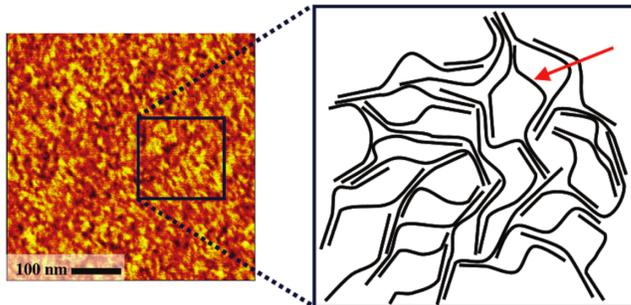


Figure 3 – Above: P3HT, π - π stacking, and the structure of the P3HT nanofiber. Below: An early theory of charge transport in P3HT thin films, claiming that inter-grain tie chains were more important for charge transport than isolated crystalline domains. Adapted with permission from Ref. 35. Copyright 2005 American Chemical Society.

The molecular weight (MW) of P3HT has a well-characterized but complex effect on its mobility. In an early study on P3HT morphology, illustrated in **Figure 3**, Kline *et al.* found that low-MW P3HT formed distinct nanofibers with a mobility of $10^{-4} \text{ cm}^2/\text{Vs}$,

but higher molecular weight P3HT had higher mobility of $0.01 \text{ cm}^2/\text{Vs}$ with a disordered morphology.³⁵ It was theorized that while charge transport was faster within the low-MW fibers, long-range transport was ultimately limited by the lack of inter-grain connectivity. However, nanofibers with higher molecular weight were soon observed, leading to OFET devices whose structure and mobility could be tuned through the control of the crystallization of P3HT. Today, it is known that both high intra-fiber crystallinity as well as inter-grain connectivity are important factors in determining mobility.³⁶

1.3 Processing P3HT Nanofibers

1.3.1 The Cambrian Explosion, P3HT-style

Over the next decade, studies on the P3HT nanofiber proliferated. This was due in part to its increasing commercial availability, enabling researchers to study its complex processing behavior without first having to synthesize it.³⁷ But a large part of its appeal was that its crystallization could be controlled in many interesting ways, making it an endlessly revelatory case study on polymer crystallization. P3HT has been subjected to, in no particular order: thermal cycling,³⁸ solvent-vapor annealing,³⁹ aging,⁴⁰ mixed solvents,⁴¹ magnetic fields,⁴² inkjet printing,⁴³ and mechanical rubbing,⁴⁴ among many other treatments. Its thin films have been deposited by spin coating,⁴⁵ drop casting,⁴⁶ dip coating,⁴⁷ blade coating,⁴⁸ and spray coating (again, among many other methods).^{49,50} These processing studies were conducted in the context of both OFETs and bulk heterojunction solar cells, where P3HT finds frequent use as an electron donor material.⁵¹

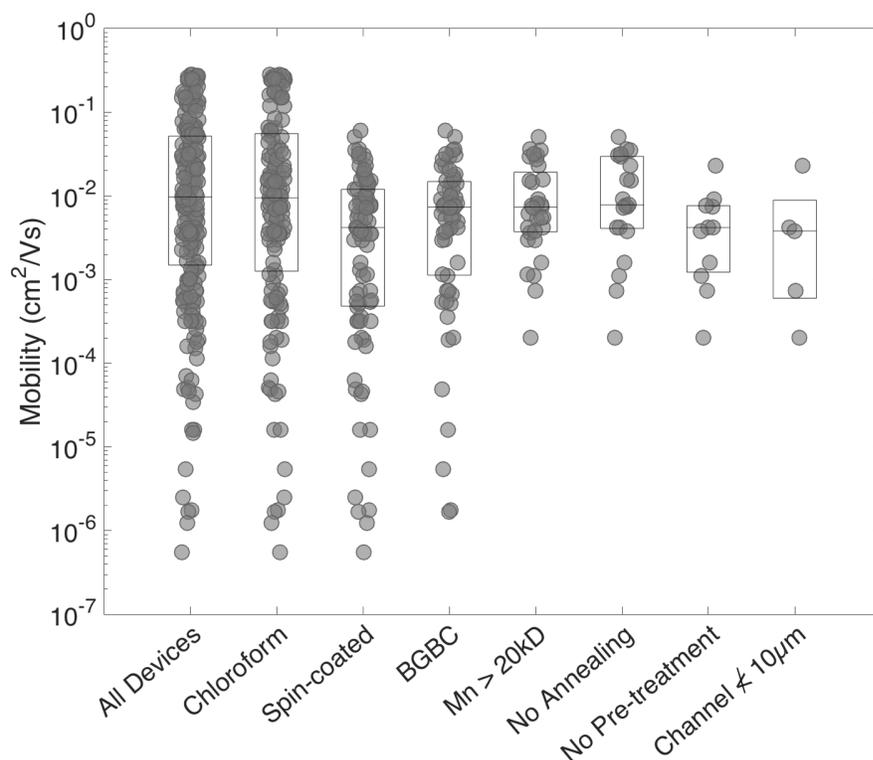


Figure 4 – Reported mobility of devices satisfying progressively tighter constraints in the OFET database. The constraint in each column stacks on all the previous constraints. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.

A major problem with these one-parameter-at-a-time studies is that they were performed on non-standardized OFET device platforms. The literature on P3HT-based OFETs was analyzed through the construction of a searchable database of OFET device processing and electrical properties, detailed in **Appendix C**.⁵² The major result is illustrated in **Figure 4**: in considering over 200 individual devices from over two dozen studies, reported mobility spans *seven orders of magnitude*. By filtering this database through successively more stringent constraints on processing and device architecture, a relatively standardized process was identified: P3HT with $M_n > 20$ kD, dissolved in chloroform with no pre-treatment, spin-coated on a bottom gate, bottom contact device

substrate with channel length $> 10 \mu\text{m}$. Even among the five devices identified in this search, the variance in mobility was still two orders of magnitude. Clearly, quantitative, meaningful process-structure-property relationships could not be reliably extracted from these isolated studies.

1.3.2 Studying P3HT Processing on a Standard Device Platform

In the years leading up to the inception of this project, the Reichmanis group began to demonstrate facile methods to control the nucleation and growth of high molecular weight P3HT nanofibers directly in solution, resulting in the fabrication of OFETs with mobilities up to $0.1 \text{ cm}^2/\text{Vs}$, as shown in **Figure 5**. Solution-based preparation of nanofibers offered a way to obtain devices with high mobility, without requiring complex post-deposition treatments such as solvent-vapor annealing that would significantly complicate roll-to-roll processing. In the lab experiments, thin films were deposited *via* spin coating on a standard OFET platform: bottom gate, bottom contact architecture with gold electrodes, an untreated SiO_2 dielectric layer, and $50 \mu\text{m}$ channel length. Preeminent among the new solution processing methods were sonication, UV irradiation, and poor solvent treatment of solutions of P3HT in chloroform. These treatments resulted in surface morphologies with observable nanofiber structures, similar to the fibers observed by Kline *et al.* and many others. In addition, aggregation of P3HT in the solution and solid states was confirmed by its well-characterized effect on its UV-visible absorbance spectrum.

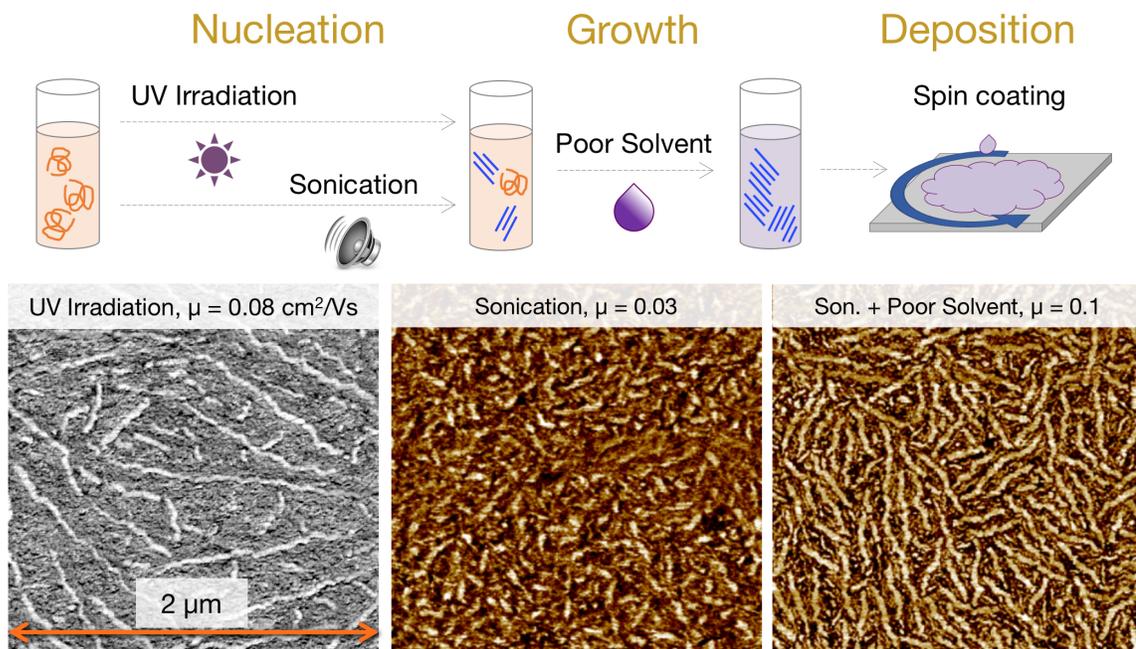


Figure 5 – Nucleation, growth and deposition of P3HT nanofibers through various solution processing techniques, with accompanying AFM images (2 μm) showing the resulting fibrillar morphologies.

1.3.3 Structural Characterization of P3HT

Structural characterization of P3HT thin films proceeds through three major techniques, illustrated in **Figure 6**. The aggregation of P3HT is reflected by its UV-vis absorbance spectrum in both the solution and solid state. The percentage of P3HT involved in aggregates is directly correlated with the percentage of the spectrum accounted for by a fitted Frank-Condon progression (purple area), with amorphous chains making up the balance.⁵³ The ratio of the two largest peak areas is used to calculate a sample's exciton bandwidth, which is correlated with the conjugation length and increased planarization of the P3HT backbone, as well as a strengthening of J-aggregate behavior relative to H-aggregate.⁵⁴ This can also be interpreted as intramolecular electronic interactions becoming stronger relative to intermolecular interactions. The aggregate fraction and conjugation

lengths tend to rise as more P3HT chains become incorporated into nanofibers.⁵⁵ A dichroic ratio can also be calculated from polarized UV-Vis measurements, indicating the degree of bulk anisotropy in P3HT chains' in-plane orientations.

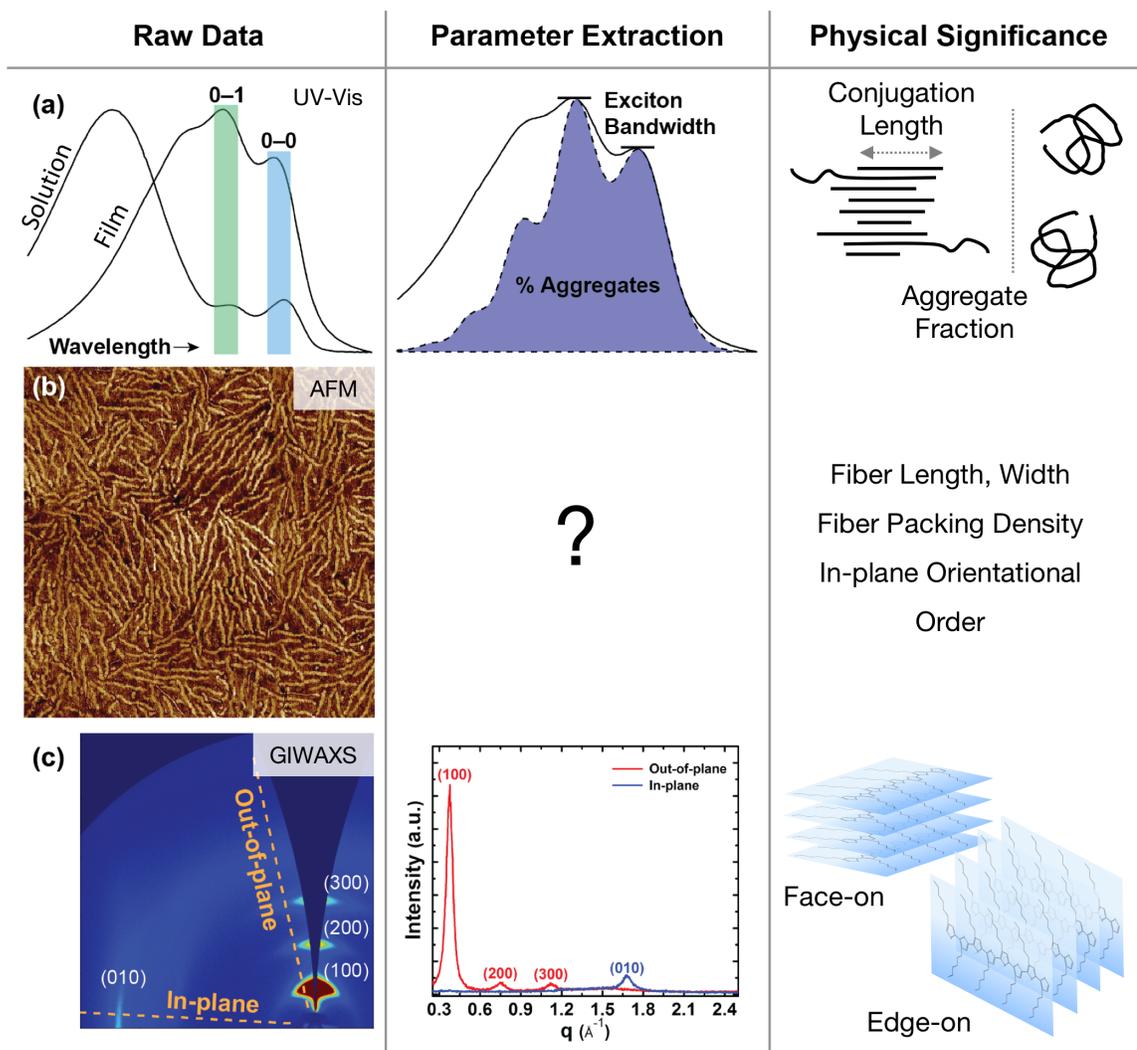


Figure 6 – Structural characterization of P3HT thin films with examples of raw data, model parameters, and their physical significance.

Grazing incidence wide angle X-ray scattering (GIWAXS) can be used to analyze the spacing, size and bulk orientation of crystalline grains in the thin film. Scherrer's equation allows estimation of the grain size along each lattice vector using the full-width-

half-max of each peak, and Herman's orientation factor quantifies the out-of-plane orientation of crystalline grains.⁵⁶ P3HT crystals can adopt many orientations relative to their substrate, the two most common being face-on and edge-on. Since charge transport occurs primarily through polymer chains and π - π stacks, the edge-on orientation is desirable for OFETs.

While UV-Vis and GIWAXS have provided significant insight into the crystallization of P3HT, they cannot directly characterize what is arguably the most important aspect of P3HT thin films: *the fibers themselves*. Since the early days of research on P3HT, AFM images have been presented as evidence of its intriguing fibrillar morphology. The phase channel of AFM imaging provides excellent contrast between P3HT's nanofibrillar phase and the surrounding disordered polymer matrix, yielding images like those shown in **Figure 5**, **Figure 6**, and **Figure 7**.

Properties of individual fibers such as their width have been measured from images by eye, but a comprehensive analysis of the long-range packing and alignment behavior of P3HT nanofibers had never been performed.⁵⁷ This was especially troubling given the clear importance of grain boundaries and orientational order to charge transport.⁵⁸ Researchers in the simulation community agreed that a quantitative understanding of meso-scale microstructure was a major missing piece in bridging the gap between molecular simulations and device-scale models of charge transport.⁵⁹ Furthermore, it was anticipated that as research on nanofiber processing in the Reichmanis group progressed, the observed fibrillar structures would grow in complexity, requiring higher resolution imaging as well as a quantitative model to relate their structure to solution processing and electrical properties. This ultimately required advancements in AFM imaging of P3HT,

microstructural image analysis, and the management of raw process-structure-property data. The greatest among these challenges was the extraction of fibers from images.

1.4 Extracting Fibers from Images

1.4.1 Classification, Segmentation, and Measurement

A grayscale image is a two-dimensional unsigned 8-bit integer array in which each element corresponds to an intensity. Images do not intrinsically contain any high-level information about what objects they depict. As humans equipped with 100-billion-parameter neural networks, we can glance at an image in **Figure 5** and immediately recognize that it contains fibers. If we sit down for a time, we can probably trace every single fiber in each image. This is because, evolutionarily, our brains have developed and refined complex visual pathways, some of which are dedicated to the recognition of fibers. In contrast, in the field of computer vision, trained convolutional neural nets are just beginning to match human accuracy in the former, simpler task of recognizing that an image contains fibers at all.⁶⁰ This is referred to as *classification*, and is demonstrated in **Chapter 5** for fiber recognition, as well as for the classification of image artifacts.

Object localization, or *segmentation*, as it is called in materials science, refers to the task of identifying which pixels belong to each object in an image. This seems easy at first glance; the fibers are the brighter pixels in the images. However, Atomic Force Microscopy of soft materials is inherently noisy and low contrast, partially because it is not truly an imaging technique in the strictest sense of the word. Instead, AFM generates an image array from the response of a tip to a drive signal as it traverses the surface of a sample, and thereby comes with a wide variety of artifacts, which are analyzed in detail in

Chapter 4. In many cases, these artifacts look very fiber-like to a computer vision algorithm, which complicates the segmentation process.

Furthermore, P3HT nanofibers exhibit several unique structural features that, while interesting physical phenomena in and of themselves, also cause unexpected difficulty in the extraction process. A blown-up 5 μm image representative of these difficulties is provided in **Figure 7**. Nanofibers exhibit a wide range of lengths and packing densities. The case of high packing density presents extreme difficulty, even for a human performing manual tracing. Oftentimes in larger images, it is virtually impossible for a human to successfully deconvolute densely packed fibers, and a computer vision approach is the only way to obtain a reliable estimate of their structure. Finally, P3HT nanofibers exhibit abrupt nano-scale kinks (high curvature bends) that differentiate them from the smooth fibers considered in previous approaches to the fiber extraction problem.^{61–63} This characteristic is examined from a crystallization perspective in **Chapter 4**.

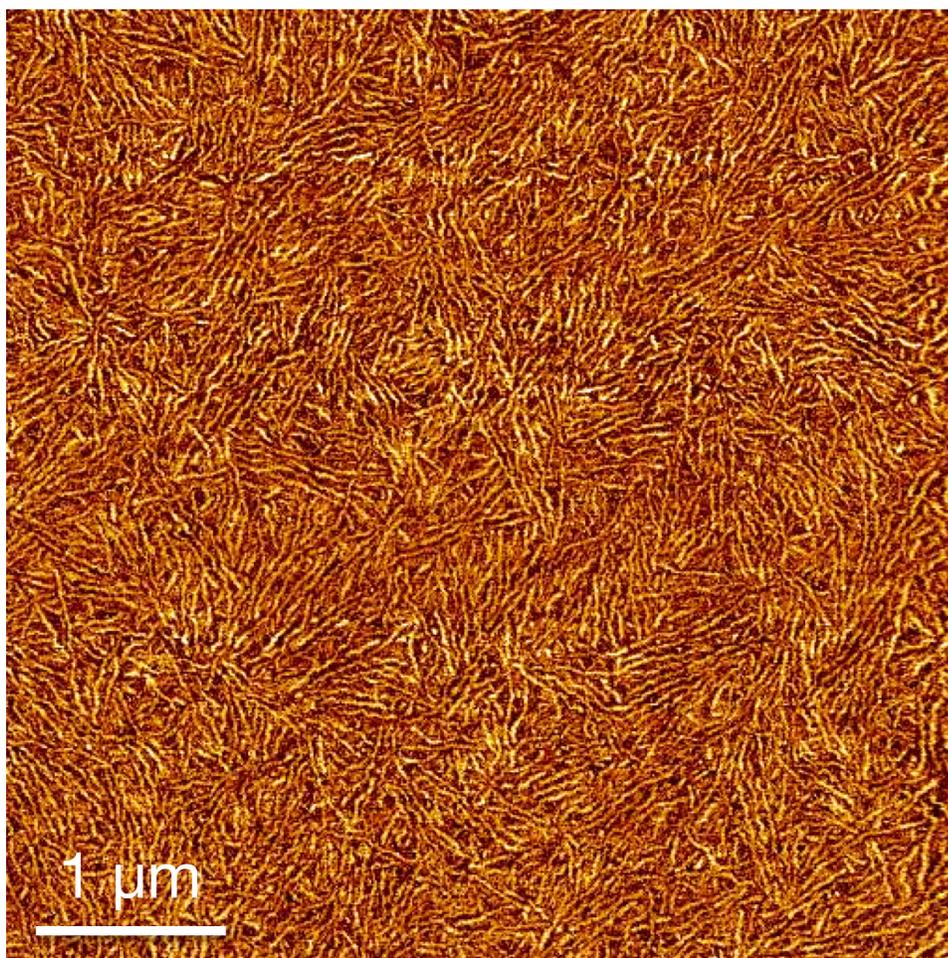
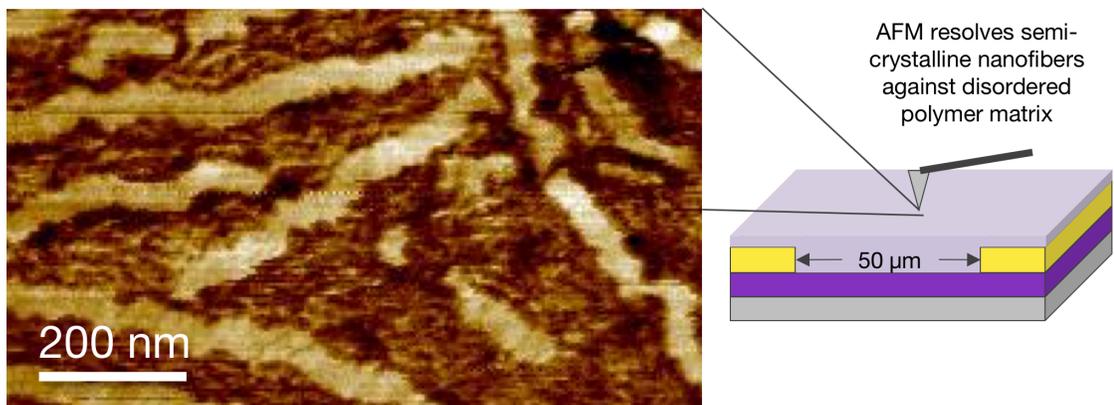


Figure 7 – Characterization of P3HT nanofibers *via* AFM, with representative images provided at different length scales.

1.4.2 *The Beauty of Open Source*

Even once we extract the fibers from an image, it is not immediately obvious which structural order parameters will correlate meaningfully with processing, a common problem in the analysis and control of complex systems and materials.^{64,65} Fortunately, this study is not the first time a researcher has tried to extract fibers from their images, and it certainly will not be the last.^{62,63,66} Advances in imaging technology⁶⁷ and a steady increase in research on fibrillar materials^{68–70} and other soft materials with complex, oriented structures indicate that this problem will only grow in magnitude as time goes on. To this end, several open source software tools have been developed with the aim of mitigating the current difficulties with fiber analysis. Four programs stand out in particular, each developed for a specific application and targeting one or two aspects of the problem with relative success: the CellProfiler Worm Toolbox,⁷¹ FIRE,⁷² ADAblock,⁷³ and FiberApp.^{74,75}

The relationship between these packages is illustrated in **Figure 8**. For the purpose of analyzing P3HT nanofiber images, each package has advantages and drawbacks but none are able to accomplish all of the desired goals. Without going into algorithmic detail, the package with the best analytical capabilities was FiberApp: designed for generalizability, it contains models for nearly every aspect of fibrillar structure one could reasonably want to quantify. Its main drawback is that fiber extraction is only possible through enhanced manual tracing. Besides a lack of intellectual contribution, accurate manual tracing is time consuming (regularly taking up to an hour per image), and it was not possible to estimate how many images would eventually be contained in the Reichmanis image library (hundreds). Manual tracing also lacks any semblance of

statistical rigor or reproducibility. That being said, FiberApp was used extensively throughout this work to generate ground truth training images for accuracy testing and optimization, and contributed to a fuller understanding of the theory of fibrillar structure quantification.

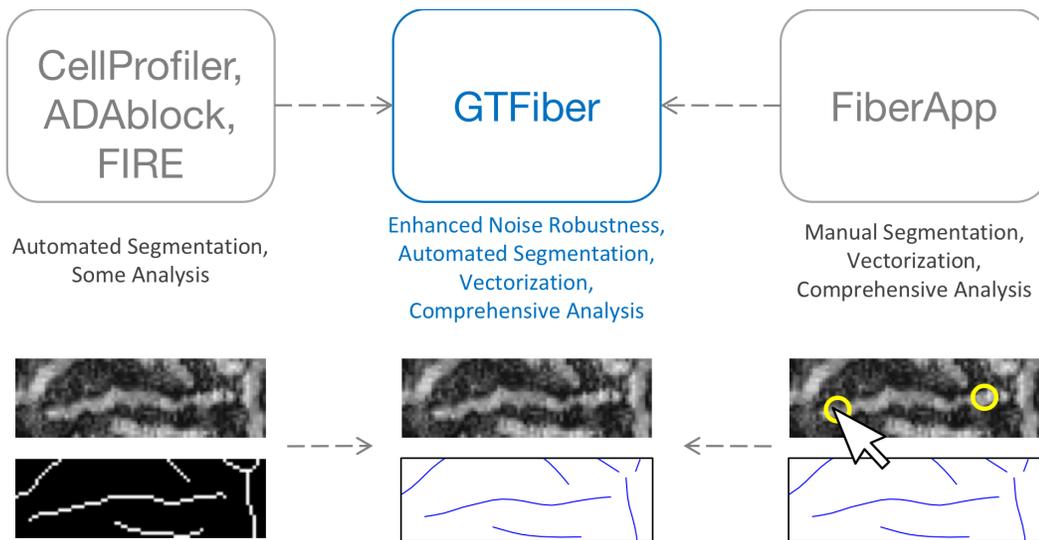


Figure 8 – Open source software that contributed to GTFiber.^{71,74} CellProfiler, ADAblock, and FIRE contained strategies for the automated skeletonization of fibers, while FiberApp contained algorithms for manual vectorization and structural analysis. GTFiber unites these algorithms to provide automated fiber vectorization and analysis.

The other three packages have strategies for automated extraction but are mismatched in their analysis capabilities or are not robust to the idiosyncrasies of AFM images of P3HT nanofibers. CellProfiler, FIRE, and ADAblock generally followed the same heuristic model of extraction, with CellProfiler incorporating a machine learning element. However, CellProfiler was designed specifically to analyze fat content in worms, ADAblock for block copolymer defect analysis, and FIRE for collagen network analysis, and are thus inherently limited in their applicability to P3HT nanofibers. In addition, some of these packages were tested on relatively clean images in which the fibers were already

strongly contrasted with their background, meaning that our images created problems for their extraction procedures.

1.4.3 GTFiber

Since any workable solution would involve combining pieces from all of these packages, FiberApp was chosen as the base off of which to build the image analysis pipeline because it was closest to the desired result. Through extensive trial and error incorporating the best elements from each of these open source software packages, a moderately robust image analysis pipeline was developed. Each new image produced from my own OFET processing experiments as well as those of my collaborators in the Reichmanis group introduced either a new structural feature to analyze or a new imaging artifact that broke the image processing algorithms. Detailed analysis of the new images also suggested new experiments to perform that likely would not have come up otherwise. Through this iterative process, both the image analysis program and our understanding of the process-structure-property relationships for nanofibrillar transistors improved.

A breakthrough occurred with the implementation of a graphical user interface (GUI). In attempting to create an interface that others could navigate, the problem, approach, and goals became clearer. Thus, GTFiber was born, and is consequently one of the primary contributions of this thesis. GTFiber is a comprehensive image analysis software package for fibrillar microstructures, compiled as a standalone application for both Windows and Mac OS and available at [gtfiber.github.io]. It contains segmentation algorithms from CellProfiler, FIRE, and ADABlock, the vectorization and analysis suite from FiberApp, and is bolstered by additional pre-processing filters of my own curation

that make it robust to the particular artifacts common to AFM and P3HT. Further optimization and accuracy quantification for the present image library was implemented through a machine learning approach. GTFiber is thus an improved iteration of fiber analysis in an open source software ecosystem that can provide immediate benefit to some researchers, and algorithmic strategies and implementations for others.

1.5 Overview of Chapters

Chapters 2 and 3 outline the development of GTFiber, and include examples of its application to the process-structure-property relationships of P3HT OFETs along the way. Early in the program's development, it became evident that orientational order is much simpler to analyze than fiber length distributions; it can be measured without extracting individual fibers. Thus, **Chapter 2** describes the analysis of orientational order in polymeric transistors, while **Chapter 3** builds on this by adding fiber length, width, and packing density to the mix. The volume of image data considered also increases with each additional chapter, as experimental capabilities grew in both complexity and throughput over time.

The development of GTFiber and the accompanying analysis of process-structure-property relationships in P3HT-based transistors was performed as part of a traineeship with the NSF FLAMEL program: From Learning, Analytics, and Materials to Entrepreneurship and Leadership – a program under the Materials Genome Initiative (MGI).⁷⁶ The MGI is an effort to halve the time and cost of developing new materials by leveraging data science, machine learning and supercomputing to handle the rapidly expanding volume of data produced by materials research.⁷⁷

With the aims of the MGI in mind, **Chapter 4** grapples with the broader problem of handling large volumes of materials image data. It explores large image databases from a high-level perspective through the analysis of 3,000 images from a single AFM instrument. Cutting-edge tools from modern computer vision such as convolutional neural nets are utilized to demonstrate automated classification of microstructures and image artifacts, as well as reverse image search functionality to connect researchers whose experiments produce similar structures. It is expected that the methods and results presented herein will hold broad appeal both in the organic electronics and materials informatics communities.

Chapter 2. Automated Analysis of Orientational Order in Polymer-nanofiber-based Organic Transistors

In this chapter, we introduce GTFiber 1.0, the first standardized protocol for the quantification of structural AFM images in our field. In this version of GTFiber, fibers were segmented as an “image skeleton,” which permitted the analysis of orientational order, but not yet fiber length. However, even at this stage significant insight was gained from its application to our group’s structures.

2.1 Introduction

Automated curve and line detection for structural analysis in biological and materials applications has received a steady level of attention for nearly two decades. Applications can be found in fields ranging from *C. elegans* behavioral analysis⁷¹ and tomography of fibrous materials^{61,78} to fingerprint enhancement and analysis.⁷⁹ Each application makes use of different image features for analysis: fingerprints and block copolymer structures are analyzed for defects (discontinuities and branching points in curves),⁷³ fibrous materials may be characterized by persistence lengths and fiber orientations, and liquid crystalline fibers are analyzed for orientational order across length scales.⁷⁵ While the mathematical tools for performing these analyses have been known for some time, it is only within the last year or two that they have been implemented as user-friendly software packages. As imaging technology continues to advance, it is imperative

that research groups have access to these tools to easily extract as much quantitative knowledge from their images as possible.

The software introduced in this chapter is designed to fill a need for image analysis in the field of organic electronics. Their electrical properties are extremely sensitive to their structure at multiple length scales, as demonstrated by the broad range of charge carrier mobility values reported in **Appendix C**.^{28,35,80-82} Shear-induced alignment of π - π -stacked polymer nanofibers has shown potential as a scalable method to obtain high mobility OFETs.^{49,83} In order to rationally design such processes to obtain devices with controllable mobility, the process-structure-property relationships must be understood, requiring rigorous quantification of the thin film morphology present at the dielectric interface of the semiconducting channel. This data would also provide a bridge between the experimental and simulation communities.^{36,59}

AFM grants access to the real-space polymeric structure inside the channel of a device, which would be expected to correlate directly with that device's measured performance, as well as provide data on the long-range orientational order of fibrillar aggregates. In contrast, UV-visible (UV-vis) absorption and X-ray techniques are not easily applied to the structure present in device channels. This is not to discount the value of such techniques. Thin film structural features found to correlate well with measured hole mobility include exciton bandwidth,⁵⁴ paracrystalline disorder,³⁶ and edge-on orientation of crystalline domains.⁸⁴ Exciton bandwidth, a measure of electronic delocalization, can be extracted from UV-visible absorption spectra (UV-vis). Paracrystalline disorder can be extracted from X-ray diffraction peak shape analysis, and edge-on orientation is quantified

through Herman's orientation factor, which can be calculated from grazing incidence wide angle X-ray scattering data.⁵⁶

AFM images are often presented as evidence of a fibrillar or nodular structure, but have also been used to calculate quantities such as fiber widths and lengths.^{35,85} Verilhac *et al.* found that P3HT fiber width is directly correlated with chain length up to about 30 nm, after which chain folding perpendicular to the π - π stacking direction limits fiber width.⁸² Surin *et al.* used AFM to characterize fiber lengths and widths, finding a similar correlation, and demonstrating that films with the fibrillar morphology displayed the highest mobilities.²⁸ Surin *et al.*, Park *et al.*, and Cho *et al.* each observed an increase in fiber width after thermal annealing, as extrapolated from AFM images.^{28,86,87} In each case, physical parameters were measured by eye, using a few especially well-resolved regions of images. Singh *et al.* obtained high-resolution AFM images spanning OFET device channels, revealing the influence of source and drain contacts on fiber deposition and contact resistance. These studies represent some of the earlier uses of AFM for structural knowledge extraction.

In recent years, our group has developed solution-based methods for the formation of P3HT nanofibers that allow the fabrication of high-mobility OFETs with rapid thin film deposition methods such as spin-coating and blade coating. This includes the use of sonication, poor co-solvent addition, ultraviolet radiation, and microfluidic processing to attain solutions with a high degree of aggregation, and in some cases, lowered π - π -stacking distances.^{80,81,83,88-90} In films deposited from solutions containing nanofibers, various degrees of fiber alignment can be observed at a variety of length scales *via* AFM. A question that frequently arises is how much of this alignment can be attributed to processes

in solution, possibly due to the liquid crystalline behavior of P3HT, and how much comes as a result of flow-induced alignment during thin film deposition.^{91,92} Furthermore, can P3HT nanofiber alignment be linked to the measured charge transport properties of each film, namely the hole mobility? Herein lies the motivation for the development of the holistic image analysis procedure introduced in this article.

Image analysis has been used previously in a variety of fields to quantify fiber alignment. For tissue engineering, differences in alignment of gelatin scaffolds due to the electrospinning process were demonstrated using large-scale image analysis.^{93,94} Strain-induced alignment of collagen gels and its effect on their mechanical properties has been quantified through microscopy analysis.^{63,69} In polymer-fiber composites and composites of carbon nanotubes, micro-CT imaging has been used to obtain both fiber length and orientation distributions.^{62,70} In this work, we present a robust image analysis workflow that combines the best practices of each of these fields and demonstrate its application to the analysis of the thin film morphology of P3HT-based OFETs. It has been compiled from MATLAB as a standalone application so that readers with no computational experience can reproduce the protocol and easily apply it to their own images, and the source code has also been made available for those who wish to study and modify the procedure.

2.2 Methods

2.2.1 Overview of Orientational Order Extraction

P3HT nanofibers form a densely packed fibrillar structure that exhibits wide variations between different thin film deposition methods, as observed *via* large-scale AFM images. The phase channel of tapping mode imaging, in particular, can resolve which

regions of a film are composed of the more crystalline nanofibers, and which regions contain more amorphous material. When performing automated analysis on densely packed structures, fiber orientation and alignment are the most reliably quantifiable features. The contour length and persistence length of fibers is difficult to reliably estimate when significant overlap is present, and fiber width estimation becomes increasingly inaccurate as the length scale of the image increases (unless resolution is increased accordingly). However, as long as fiber backbones are resolved, images of increasing length scales reveal a wealth of information on fiber packing, orientation, and orientational order, which offer insight into the mechanisms of fiber formation and interaction, as well as the effects of solution processing on long range order and its contribution to device-scale charge transport.

In **Figure 9**, the main results that fall out of the analysis procedure are introduced. While many structural features can be extracted from digitized fibrillar microstructures with varying levels of confidence, orientational order is the most robust to noise and imperfect segmentation, and its extraction requires the fewest assumptions and specialized treatments. **Figure 9A** is a cropped $5 \times 5 \mu\text{m}$ section of a $10 \times 10 \mu\text{m}$ AFM phase channel micrograph of a P3HT thin film deposited from a solution of P3HT in chloroform (5 mg/mL) *via* spin-coating. It was also purposefully chosen as an example with low contrast and high levels of image noise to demonstrate the minutiae and robustness of the protocol; other images presented later are much cleaner. This image is used as an example as each image processing step is explored, including the effects of various parameters as presented in **Appendix A**. The Orientation Map, **Figure 9B**, is the final result of the processing workflow: an image in which each fiber is represented by a backbone of single pixel width,

and each of those pixels is labeled with an orientation from 0 to 180°, as indicated by the attached color wheel. The software introduced in this article generates the Orientation Map as well as analytical results. Analysis of the Orientation Map yields two principle results: an orientation distribution (**Figure 9C**) and the decay of orientational order as a function of frame size (**Figure 9D**).

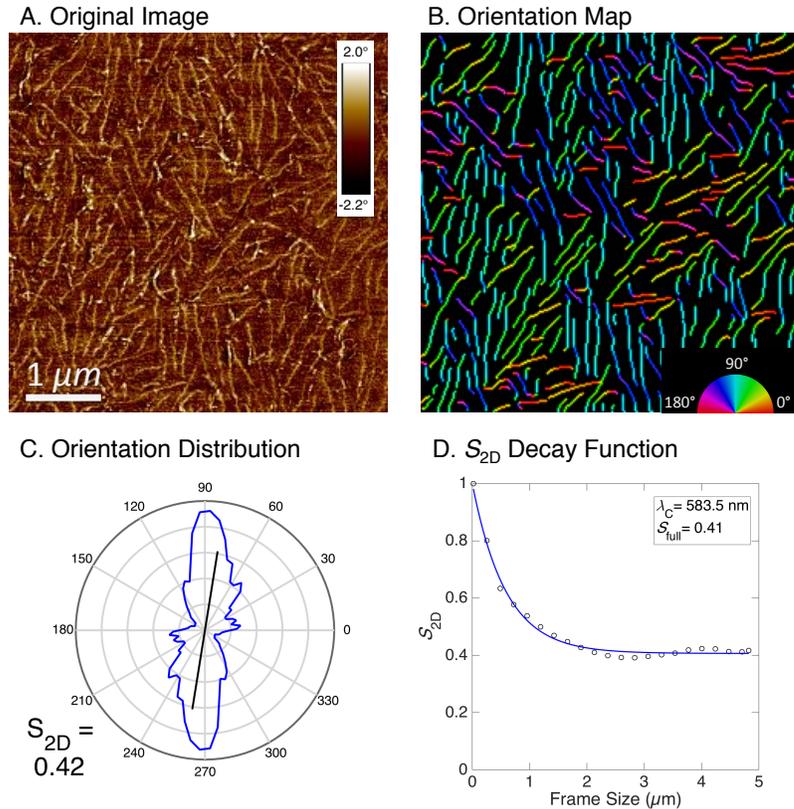


Figure 9 – Main results of the protocol. (A) A cropped version of the phase channel of a tapping mode AFM image. (B) False color Orientation Map extracted from the original image. Each pixel’s orientation corresponds to an orientation on the attached color wheel. (C) Orientation Distribution extracted from the Orientation Map. The radial axis indicates the count of pixels of a given orientation. A diametrical black line segment indicates the average orientation, and the full-frame value of S_{2D} is indicated at bottom left. (D) Decay of the orientational order parameter, S_{2D} , as a function of frame size. Fitted model parameters are indicated at upper right.

The orientation distribution in **Figure 9C** indicates that this structure has an average orientation of $\approx 80^\circ$ off of the horizontal, and that the distribution is anisotropic. The radial

axis of the orientation distribution counts pixels whose orientations fall into each of 36 bins (width 5°). The distribution is forced to be symmetric to 180° rotation because the fibers do not have any inherent directionality that is accessible through the image. In other words, an orientation vector of $[1,1]$ should count toward the same bin as a vector of $[-1,-1]$. Because the radial axis counts pixels, a long fiber will contribute more to the distribution than a short fiber of the same orientation. Skeletonization of the fibers to single-pixel thickness ensures that fiber width does not bias this distribution.

The degree of anisotropy of the orientation distribution is captured by the orientational order parameter S_{2D} , similar to Herman's Orientation Factor in GIWAXS analysis.^{73,75} Mathematically, S_{2D} is defined as:

$$S_{2D} = 2 \langle \cos^2 \theta_n \rangle - 1 \quad (1)$$

where θ_n is the angle between an individual fiber pixel and the image's overall director, \vec{n} , which is chosen as the average orientation of the population. In the context of this analysis, S_{2D} varies between 0 and 1; the expected value of $\cos^2 \theta$ for uniformly random angles is 0.5, yielding an S_{2D} of 0 invariant to the selection of the director, and a uniform oriented population generates an $\langle \cos^2 \theta_n \rangle$ of 1, yielding an S_{2D} of 1. The director is plotted in each Orientation Distribution as a centered black line segment, as in **Figure 9C**.

2.2.2 Calculating the Decay of Orientational Order

Figure 9D plots the value of the orientational order parameter S_{2D} as a function of frame size. This analytical method was introduced by Usov et al. in the FiberApp software package – it has been adapted for use here, where fibers are discretized as backbone pixels instead of segment vectors.⁷⁴ Since S_{2D} can be calculated for any population of oriented objects, it is defined for a frame as small as a single pixel or as large as an entire image. As illustrated in **Figure 10**, this calculation can be averaged over many frames tiled over an image and performed at increasing frame sizes to obtain the decay function of S_{2D} . Sampled frames that contain no fiber pixels are excluded from the average.

The two relevant computational parameters for this procedure are the grid step, which determines how many frames are sampled from the image to obtain the average S_{2D} for that frame size, and the frame step, which determines the level of discretization of the frame size axis. The decay function plotted here uses a grid step of 200nm and a frame step of 400 nm, although this is not to scale in the schematics in **Figure 10**. The schematics demonstrate how the grid step and frame size affect the sampling of S_{2D} over the image. Since the grid step remains constant, the smallest sampling frames may have space between them, while larger frames will overlap one another. When the frame size becomes too large to accommodate multiple samples at the specified grid step, a single centered frame is used. While this numerical implementation leads to an under-sampling of the fibers at the edges of the image, the values of the decay function are remarkably invariant to the grid step size, provided that it is sufficiently small. Undersampling begins to influence the results for grid steps greater than one-fifth the image width.

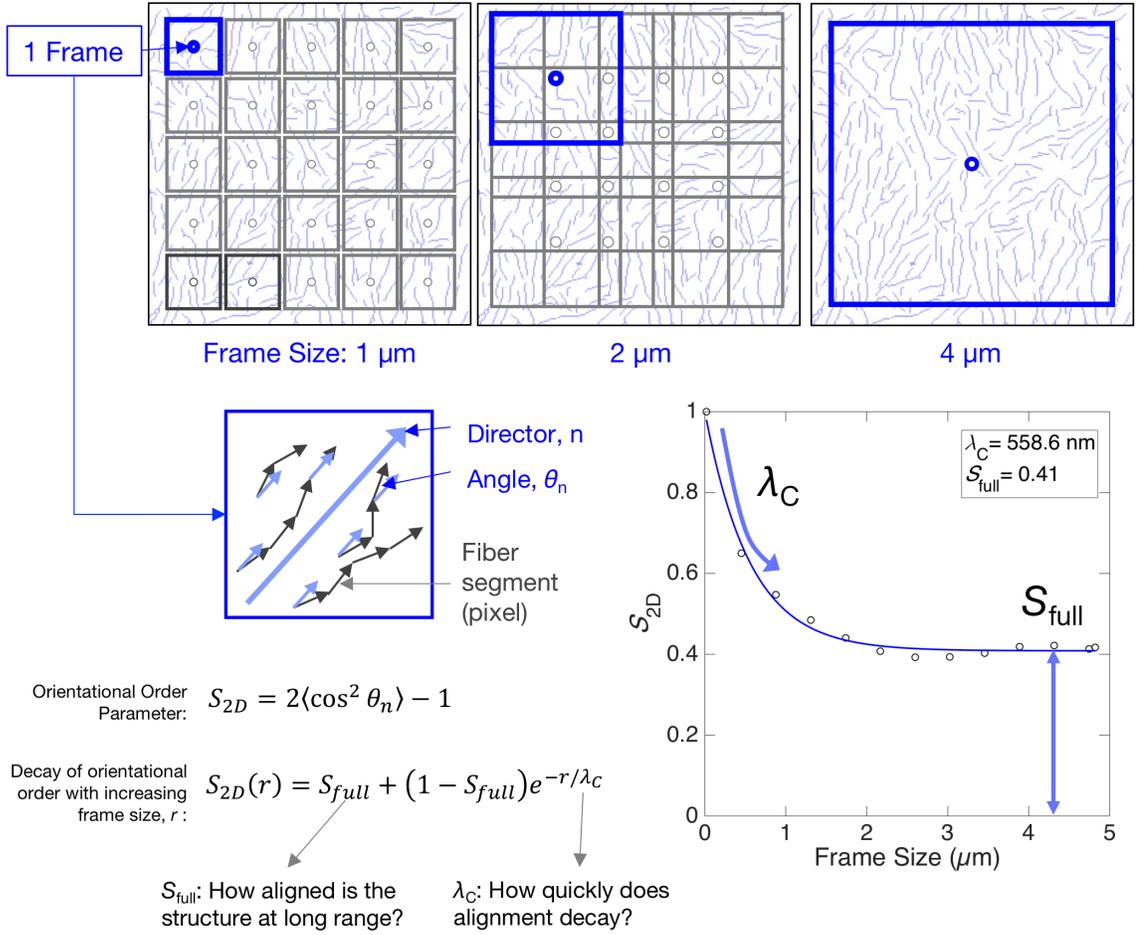


Figure 10 – Calculation of S_{2D} decay function from a skeletonized image. Each point on the graph represents the average value of S_{2D} calculated from evenly spaced frames in the image. The grid step determines the spacing between each frame, which can create sparsely distributed frames or highly overlapped frames. The frame step determines the level of discretization of the x-axis (frame size). White dots in the grid step schematic indicate the centers of frames.

Finally, an exponential decay function is fit to S_{2D} with a constant term to account for the asymptotic full-frame value of S_{2D} , which does not necessarily decay to zero:

$$S_{2D}(r) = S_{full} + (1 - S_{full})e^{-r/\lambda_C} \quad (2)$$

where S_{full} is the asymptotic value of S_{2D} , r is the frame size of calculation, and λ_C is the length scale of orientational order decay in the image, referred to hereafter as the *decay length*. Similar exponential decay functions have been implemented in other fields to

analyze orientational order.^{73,75} As shown in **Figure 10**, S_{2D} asymptotically approaches a full-frame S_{full} of 0.41 with a decay length of 559nm. It is important to distinguish between S_{2D} and S_{full} : S_{2D} is a value that is calculated from a single frame of fiber segments, while S_{full} is a model parameter fit to the set of all values of S_{2D} calculated at all frame sizes.

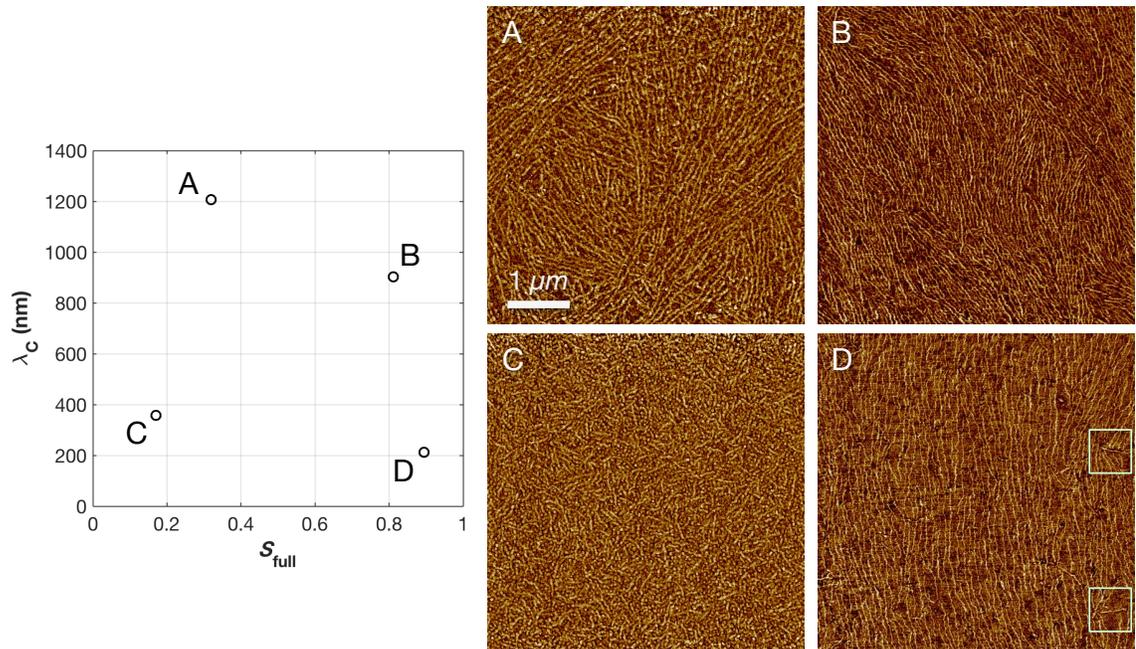


Figure 11 – Graphical representation of four extrema in the space defined by λ_C and S_{full} . (A) Low S_{full} , high λ_C (B) high S_{full} , high λ_C (C) low S_{full} , low λ_C (D) high S_{full} , low λ_C .

These parameters can be used to quantitatively analyze and compare images of fibrillar structures. The four exemplary structures shown in **Figure 11** represent extrema in both S_{full} and λ_C . S_{full} is visually intuitive: structures **B** and **D** clearly possess a greater degree of alignment (more fibers pointing in the same direction) than **A** and **C**. Decay length is more nuanced: it represents the length scale (frame size) at which S_{2D} has decayed 63% of the way to the final value of S_{full} . Images with a short decay length (**C** and **D**) possess the same degree of alignment when measured in both small and large frames, on

average. In **Figure 11C**, a small frame likely contains a few randomly oriented fibers, while the entire image simply contains more randomly oriented fibers. This is in contrast to **Figure 11A**, in which small frames contain fibers of similar orientations, while the full image contains fibers of many different orientations. In this way, λ_C represents the length scale of the persistence of orientational order.

The difference between **Figure 11B** and **D** is more subtle. **Figure 11D** has a short decay length because of the defects highlighted by the overlaid boxes. These short horizontal fibers lower the average calculated S_{2D} at short and long length scales, thereby lowering the value of λ_C . On the other hand, the orientation of the fibers in **Figure 11B** varies more gradually from the top of the image to the bottom, thus larger frames must be sampled to see the lowest alignment, yielding a higher value of λ_C . In this way, λ_C quantifies the influence of strong local defects on overall alignment. A third axis could be included to quantify the direction of alignment, which becomes more meaningful at higher values of S_{full} .

2.2.3 Obtaining the Orientation Map through Image Processing

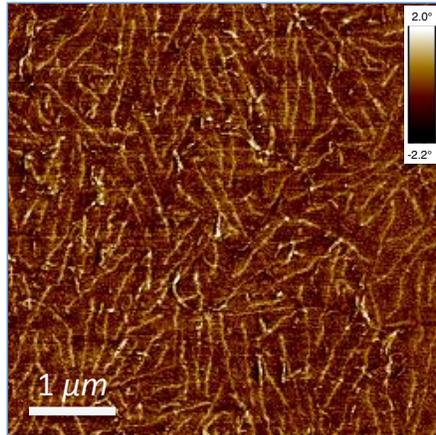
The primary image processing challenge in this protocol is to obtain the binary skeleton, a valuable data structure for the analysis of fibrillar microstructures.^{61,71,73} Once obtained, a binarized skeleton can be used for spatial correlation analysis, path finding, percolation analysis, and virtually any other relevant structural analysis desired.^{71,95,96} The skeleton is a binary image in which fibers are thinned to single pixel width so that pixels take on a value of “1” if they lie on a fiber’s backbone, and “0” if they do not. For orientational analysis, each pixel should also carry a value for its orientation; this can be

represented on a domain from 0 to 180°. The combination of the skeleton and orientations yields the Orientation Map, as introduced in **Figure 9B**. The process of obtaining the skeleton from the original is made especially difficult given the low contrast of the AFM phase channel and the presence of substantial noise. These complications are quite common, so a robust procedure for extracting the binary skeleton would be widely beneficial. The approach is outlined below, and has been implemented with a graphical user interface in MATLAB. A schematic of the image filters and morphological operations that are used to accomplish this is provided in **Figure 12**, using a 5×5 μm region of an originally 10×10 μm image as an example.

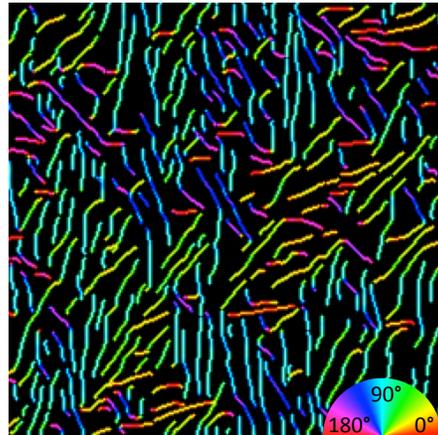
Stages 1 and 2: Coherence-Enhancing Anisotropic Diffusion Filtering

To generate a skeleton, a binarized image must first be obtained in which white pixels represent original pixels that are within fibers, and black pixels represent non-fibrillar original pixels. An example of a simple binarized image is shown in Stage 4 of **Figure 12**. This could usually be obtained by applying a global black and white threshold value, but nonidealities such as long-range image gradients, brighter noise from amorphous regions, and darker noise from within fibers make this a near impossibility. Contrast enhancements through mean and median filtering are also ineffective due to the high levels of noise and low contrast. Examples of failed thresholding are provided in **Appendix A**.

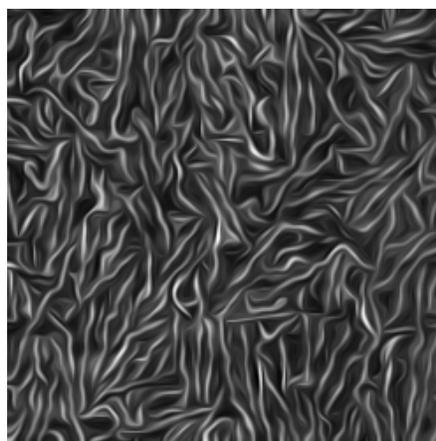
1. Original Image



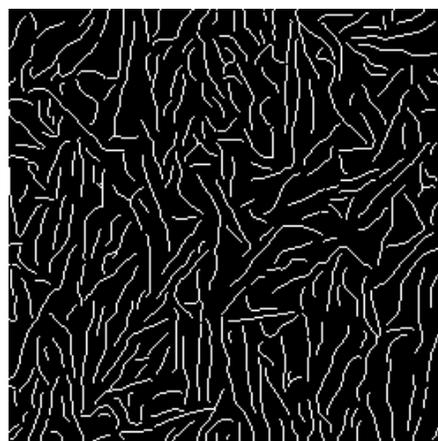
6. Orientation Map



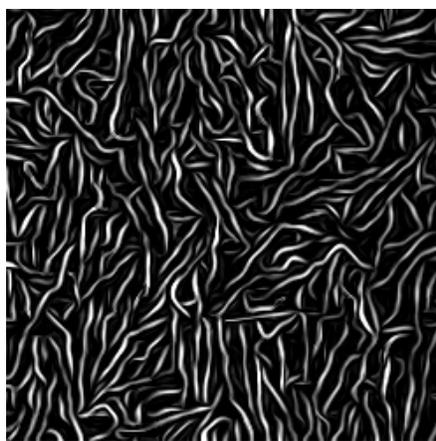
2. Diffusion Filter



5. Skeletonization



3. Top Hat Filter



4. Threshold & Clean

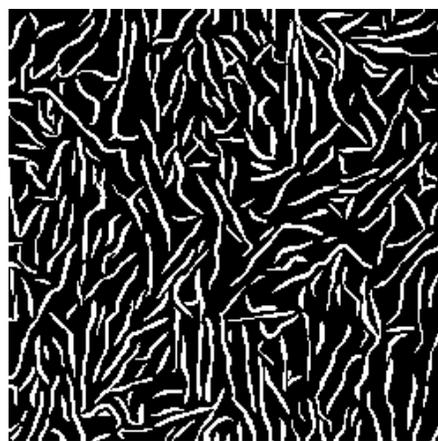


Figure 12 – Outline of image processing workflow used to obtain the skeleton and subsequent Orientation Map.

Effective contrast enhancement and smoothing can be attained through coherence-enhancing anisotropic diffusion filtering, an advanced filtering method first introduced by Weickert, implemented here with the algorithm by Perona and Malik.^{79,97} The result of the anisotropic diffusion filtering step is shown in Stage 2, “Diffusion Filter,” for the example image in **Figure 12**. Fibrillar regions appear substantially smoother, with more uniform brightness. This method diffuses pixel intensity (gray values) using a 2-D diffusion tensor proportional to the local anisotropy of the image. At each time step, pixels’ gray values are diffused in a direction that corresponds to the nearest fiber, effectively smoothing the image along the direction of fiber travel. The greatest advantage of this approach is that the orientation calculated at each pixel can be recovered from the diffusion tensor after filtering and applied directly to the final Orientation Map. A discussion of the diffusion tensor and its unique properties, as well as the numerical implementation of anisotropic diffusion filtering, can be found in **Appendix A**.

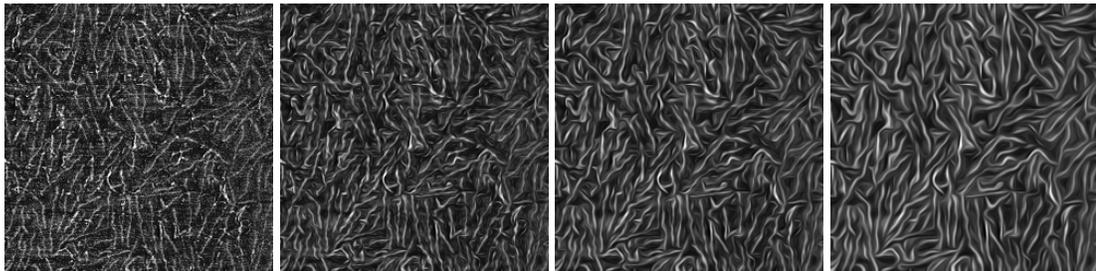


Figure 13 – The effect of the “diffusion time” parameter on diffusion filtering results, from left to right: original gray scale, 1s diffusion time, 3s, 8s. Constant parameter values: initial Gaussian smoothing, 10nm; orientation smoothing, 30nm.

While there are seven parameters that can be adjusted with diffusion filtering, there are only two that have a significant influence on the results: orientation smoothing, and diffusion time. Increasing diffusion time is illustrated in **Figure 13**, while orientation smoothing is illustrated in **Appendix A**, and can be explored *via* the graphical user

interface as well. Orientation smoothing refers to an isotropic Gaussian filter applied to each element of the diffusion tensor, so that each pixel has an orientation that represents the weighted average orientation of its surrounding pixels. In this way, sharply resolved fiber edges help inform the orientation of a fiber's backbone. The standard deviation of this Gaussian filter, ρ , determines the weight of each pixel's contribution to its neighbors, and how large a sampling neighborhood is used. If ρ is too small, noisy fiber edges will limit filter performance, but if ρ is too large, fibers with high contrast will influence the orientation of their neighbors. A good heuristic is to set orientation smoothing to the typical width of a single fiber.

Diffusion time determines how many time steps of the diffusion filter are performed, although a "time step" has no physical meaning here and simply represents how finely the integration method is discretized. A default time step of 0.15 "seconds" is typically used. Diffusion times between 1 and 10 seconds are typically appropriate. Longer diffusion times tend to blend smaller fibers into the background. Even after diffusion filtering, however, contrast is still relatively low, and thresholding yields an unsatisfactory result.

Stages 3 and 4: Top Hat Filtering and Thresholding

Once fibrillar regions have attained more uniform local gray values, traditional contrast-enhancing filters can be applied. Top hat filtering is a feature-enhancing filter that has been used to process angiograms and mineralogical images, among other applications.⁹⁸ Top hat filtering performs a morphological image opening (erosion followed by dilation) with a disk-shaped structuring element. The result of the opening is subtracted from the original gray scale image, resulting in an enhancement of sharp peaks

in the original. This can be observed in Stage 3, “Top Hat Filtering” in **Figure 12**, and the individual steps involved are illustrated in **Figure 47** of **Appendix A**. Only one parameter affects this step: the top hat filter size, which is the radius of the disk-shaped structuring element used for erosion and dilation. The size is specified in nm by the user, which is converted to pixels and rounded to an integer value. A size less than the typical fiber width will enhance only the narrowest fibers, while values greater than the maximum fiber width attain more broadly effective contrast enhancement, so the only guideline for this parameter is to set it moderately above the estimated fiber width.

After Stage 3, the image is now in an acceptable state for binarization – that is, discretely classifying which pixels belong to fibers and which do not. While a global threshold can produce acceptable results at this point, variations in intensity between fibers may still cause complications. A more robust method is to apply an adaptive threshold, especially when fibers are very dense. The result of this thresholding procedure is shown in Stage 4 of **Figure 12**, “Threshold and Clean.” The cleaning step removes contiguous white regions (of 8-connectivity) smaller than an area specified in nm^2 . The typical setting used in this protocol is 2500 nm^2 , which corresponds to an area of 26 pixels in a $5 \times 5 \mu\text{m}$, 512×512 pixel image. The performance of various thresholding procedures is illustrated in **Appendix A, Figure 51**.

Stages 5 and 6: Skeletonization and Orientation Mapping

The previous steps were all in service to generating a realistic, meaningful skeleton, which requires a smooth binarized image of the fibrillar structure. Skeletonization is a crucial step to obtain meaningful physical analyses of many images. It has been used to analyze human motion, medical images, and of course, fibers and other fiber-like

objects.^{61,71,99–101} Skeletonization precludes fiber width from biasing measurement of the orientation distribution: since each pixel contributes one count to an orientational bin, a thicker fiber's orientational contribution would scale linearly with its thickness. The process of skeletonization identifies the endpoints of a connected white object and erodes the boundaries until that object is represented as (possibly branched) line segments of single-pixel thickness. It is implemented through a built-in MATLAB function. Bumps along the edges of the thresholded image (Stage 4) lead to the identification of false fringe branches in the skeleton. A trimming algorithm was written to remove extraneous fringe branches from the main fiber backbones – the user can specify the maximum length of fringe branches to remove. Details of this step are contained in **Appendix A**.

Finally, the orientation map is generated from the skeleton by recovering orientations from the diffusion tensors that were calculated at the final time step of diffusion filtering at each pixel in the skeleton. These values are used to generate a false-color image of the Skeleton in which orientations off of the horizontal are represented by their hue on a color wheel. Every step of this procedure can be reproduced and implemented by researchers using a standalone application, shown in **Figure 14**. An image is loaded, with dimensions provided by the user in nm. Settings for each filtering step can be chosen, or the default values can be used. A single button runs the entire filtering regime and produces the skeleton. The user can choose to display the result of each step of filtering in order to visualize the effect of their chosen parameters. After filtering, the Orientation Map can be produced with a single button push, and the decay function can be plotted as well. Once parameters have been found that produce an acceptable result, a batch of images can be run from a folder with the chosen settings, with the results saved in a .csv file readable

by Excel. Publishable figures can be saved as well. The software is available for download at [gtfiber.github.io].

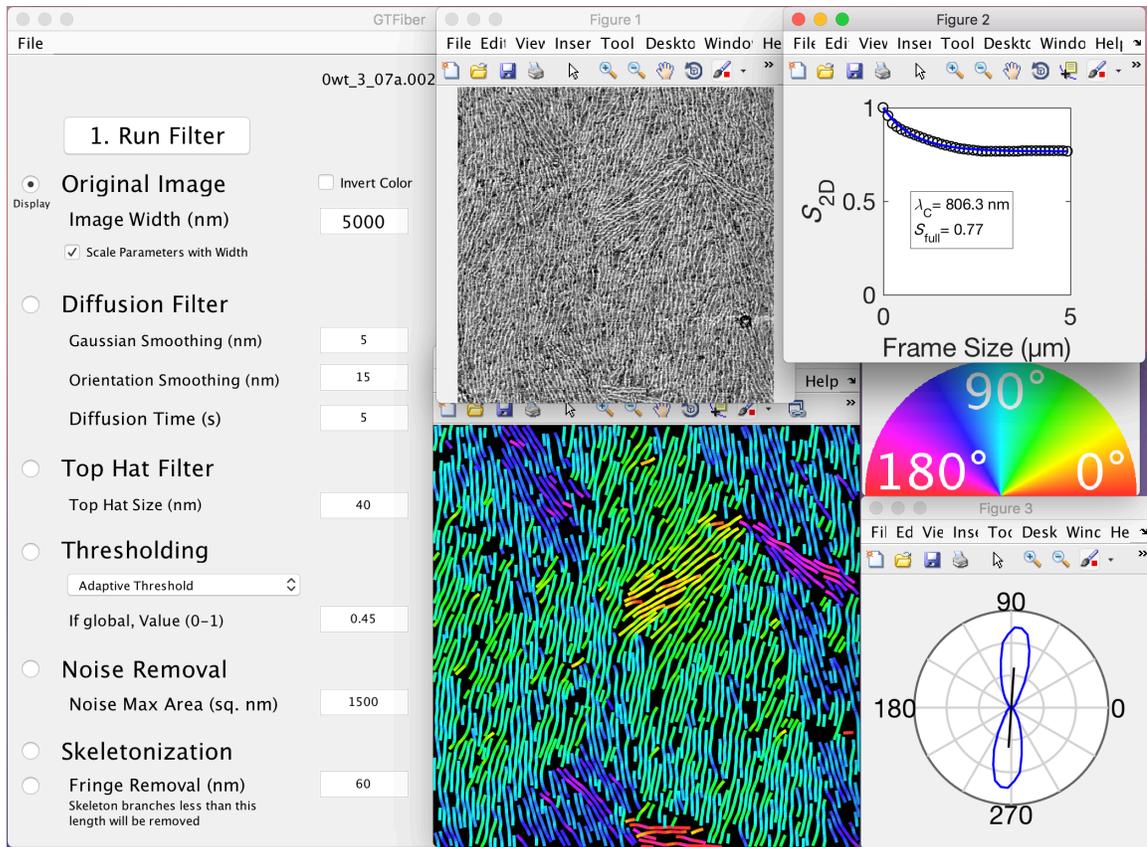


Figure 14 – Graphical User Interface for the image analysis methods introduced here, written in MATLAB and compiled as a standalone application.

The contour extraction method presented here is extremely robust. From any image with one-dimensional features whose brightness contrasts moderately with the background, this software can reliably extract the backbones and orientations (the Orientation Map). Examples of the application of the software to Micro-CT, SEM and TEM images are provided in **Appendix A**. Application to three-dimensional structures and data is also possible and is the subject of an ongoing collaboration. Other ongoing research efforts

include the analysis of structural parameters for spiral-like and elliptical morphologies, as well as the use of machine learning for fiber classification and clustering.

2.3 Results

2.3.1 *The Effect of Solution Processing on Morphology*

The image analysis protocol outlined above has been instrumental in analyzing the long range ordering and orientation of thin films of P3HT nanofibers. The many processing steps involved in solution processing of organic transistors makes it difficult to deconvolute the origin of structural order: does it arise in solution, or during deposition? What is the contribution of each step? In **Figure 15** we present the results of three high-resolution 5×5 μm AFM images that have been analyzed with our software. **Figure 15A** shows a film deposited from a solution of P3HT in chloroform and 15 vol% 2-methylpentane, a poor co-solvent. The solution was sonicated for two minutes and allowed to age for four days. Similarly, **Figure 15B** shows a film deposited from a P3HT/chloroform solution that was sonicated for two minutes aged for four days, without a poor co-solvent. **Figure 15C** shows a film deposited from a P3HT/chloroform that was processed under UV irradiation in a microfluidic flow system.⁸³ These processing methods are detailed in **Figure 20** of **Chapter 3**; for now, the images simply serve as an example of the variance captured by the order parameters introduced here. All films were deposited *via* spin-coating for 60 s at 1500 rpm.

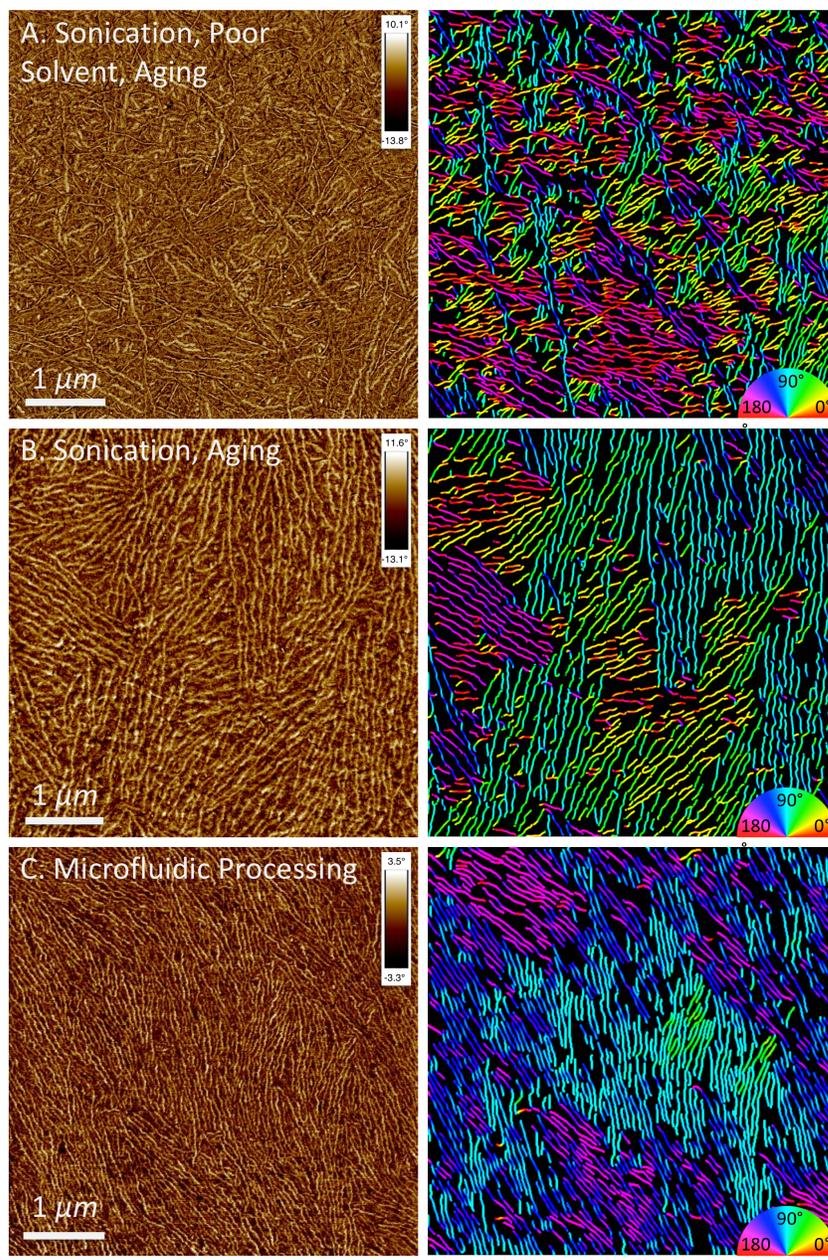


Figure 15 – AFM images and Orientation Maps from films spin-coated from three solutions. (A) P3HT in chloroform and 15 vol% 2-methylpentane (5 mg/mL), sonicated two minutes then aged four days. (B) P3HT in chloroform (5 mg/mL), sonicated two minutes then aged four days. (C) P3HT in chloroform (5 mg/mL), processed *via* microfluidic-cooling-UV treatment. Image processing parameters used: Gaussian smoothing, 10nm; orientation smoothing, 30nm; diffusion time, 8s; top hat size, 30nm; adaptive thresholding; noise removal, 2500 nm²; skeleton fringe removal, 40nm. Orientation Maps have been anti-aliased for printing clarity.

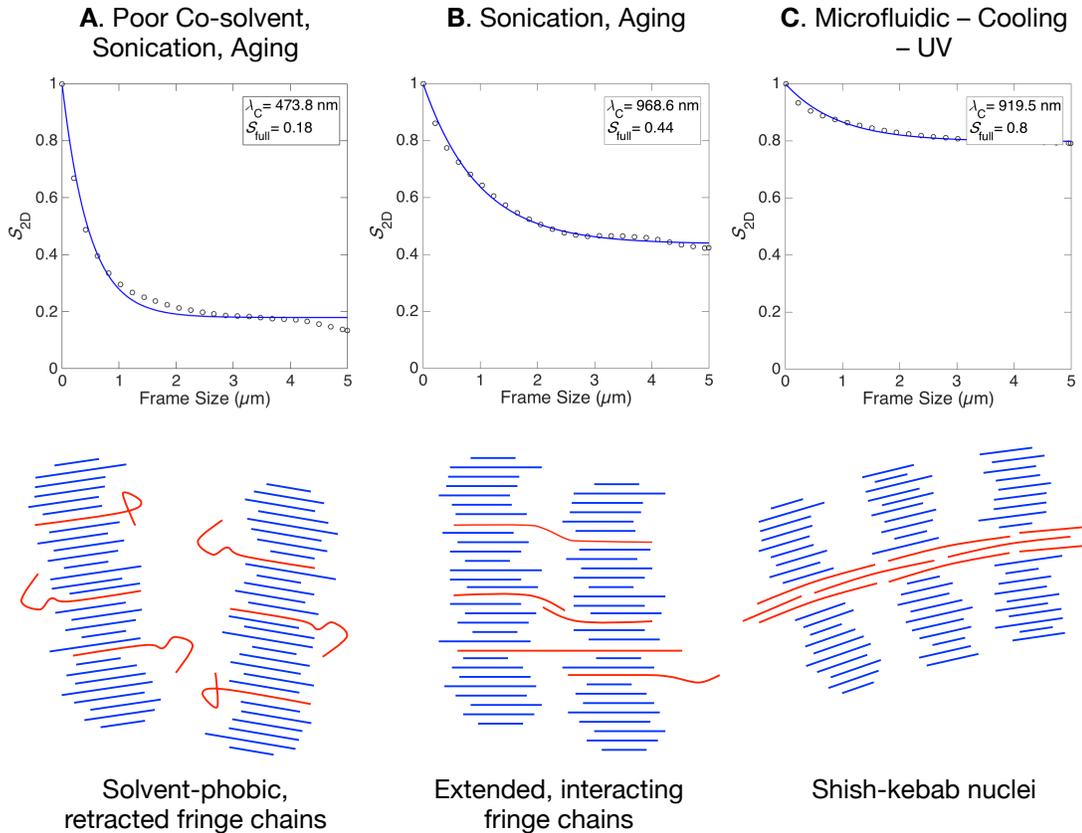


Figure 16 – Decay of orientational order and schematic of inter-fiber interaction for each of the three morphologies presented in Figure 15. Parameters used for orientational order decay calculation: frame step, 200nm; grid step, 200nm.

Using the original AFM images, it can be said qualitatively that long-range orientational order appears to increase over these three samples, from **A** to **C**. Using the structural visualizations and order parameters, we can better quantify these differences and begin to understand the mechanisms that bring them about. The Orientation Maps highlight a phenomenon termed “fiber bundling” – the presence of large bundles of neighboring fibers of the same color, *i.e.* whose backbones are parallel to one another. Note that many more isolated fibers with no parallel neighbors are present in **Figure 15A**, whereas the fibers in **Figure 15C** seem to be a part of one large bundle. The phenomenon of fiber bundling is a quantified by the decay of local orientational order, which is plotted for each

image in **Figure 16**. The combination of poor co-solvent addition, sonication and aging, **Figure 15A and Figure 16A**, shows a steep drop in S_{2D} , with a decay length of 474 nm and a final S_{full} of 0.18. S_{2D} is still decreasing at 5 μm – the last point has a value of 0.13, indicating nearly random long-range orientation. In comparison, the sonicated and aged sample and the microfluidic processed sample have decay lengths of almost 1 μm and S_{full} of 0.44 and 0.8, respectively. The increased decay length is indicative of the local fiber bundling that we observed.

The differences in local order in these samples can be attributed to differences in the self-assembly mechanism of P3HT nanofibers, as depicted in the diagrams in **Figure 16**. While it is widely known that π - π -stacking of individual P3HT chains causes the formation of these fibers, what is less well known is how fibers interact with one another, especially when considering the polydispersity of P3HT.^{33,50} The longest chains in the molecular weight distribution create fringes that extend beyond their host fibers' edges. In a poor solvent, these fringes would be expected to coil up so as to decrease their interaction with the solvent and thus their interaction with other nearby fibers. In a solvent such as chloroform, the fringe chains can extend to create strong physical interactions between parallel neighbors, such as tie chains and contact points.³⁴ In the most extreme case of microfluidic processing, the formation of shish-kebab nuclei is a well-documented crystallization mechanism that may create strongly interconnected parallel fibers, leading to the high degree of orientational order at a wide range of length scales.^{102–104} The fact that these films were all spin-coated under similar conditions indicates that order formed in solution can persist through the spin-coating process, despite the high shear rates and rapid solvent evaporation that occur.

2.3.2 *The Effect of Deposition Method on Morphology*

This is not to say that the spin-coating process has no effect on fiber orientation; to investigate, we prepared a solution of P3HT nanofibers in chloroform by sonicating and aging two days, spin-coated it onto a glass slide, and collected AFM images of the morphology at varying locations around the center of spin-coating, noting the (x,y) position of the AFM stage in the Nanoscope software. The results are depicted in **Figure 17**, in which the average orientation of each image is plotted as a line segment at the location where each image was taken, and the line segment is scaled by the full-frame value of S_{2D} for that image. There is a radial component to the orientations, which would be expected given the strong centrifugal forces present during spin coating. However, the orientations also have a bias in the direction of initial angular acceleration, which suggests that the solidification of the film occurs before a steady-state radial flow profile can develop. These results are in agreement with other studies that suggest colloidal rods align with centrifugal flow fields when they are present at a high enough volume fraction.¹⁰⁵ Given the extensive use of spin-coating in solution processing, high-throughput image analysis of the effect of its fluid dynamics could be a valuable tool for understanding and controlling deposition using this technique.

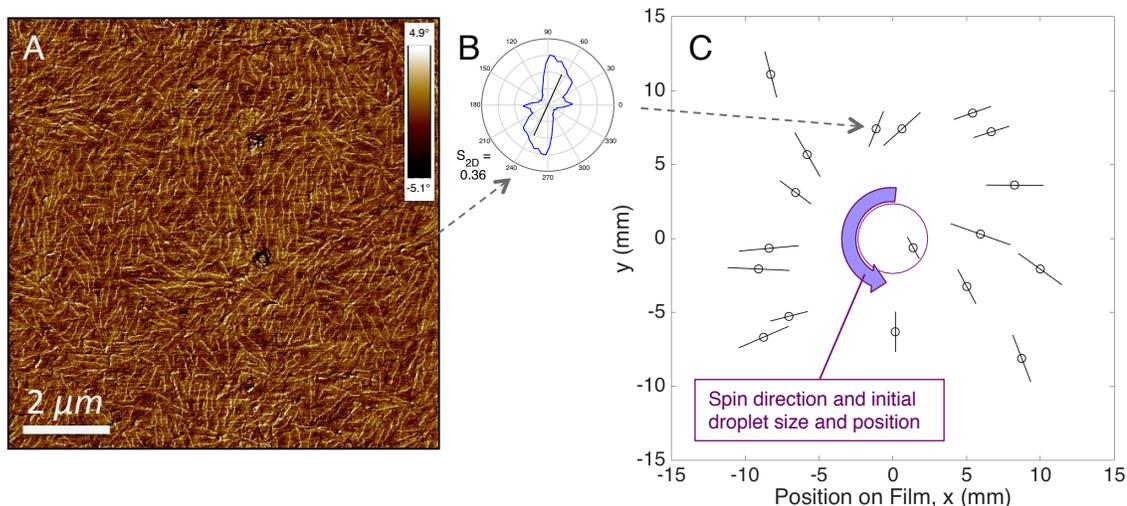


Figure 17 – The effect of spin coating on fiber alignment. (A) Original tapping mode phase image (10 x 10 μm) from one location. (B) Orientation distributions for accompanying image. (C) Map of average fiber orientation as a function of image location on final spin-coated film (30 x 30 mm). The center of the original droplet is taken to be the origin. Spin direction was counter-clockwise.

Shear-coating with micro-patterned blades has received a fair amount of attention as a mechanism to align polymer chains.^{106–108} However, when coating solutions of pre-formed nanofibers, simpler blade designs can be used to achieve macro-scale alignment and enhanced charge transport. Chang *et al.* obtained aligned films of P3HT and poly(3-butylthiophene) nanofibers by using a glass cover slip as the shearing device, and growing nanofibers using the previously described UV irradiation technique. The degree of alignment was tunable by the coating velocity, as illustrated in **Figure 18a–e**.¹⁰⁹ Exciton bandwidth, (100) *d*-spacing, fiber alignment, and mobility were optimized at intermediate coating rates between 1–2 mm/s. The coating velocity likely modulated the film thickness as well, which can play an important role in the measured charge transport.¹¹⁰

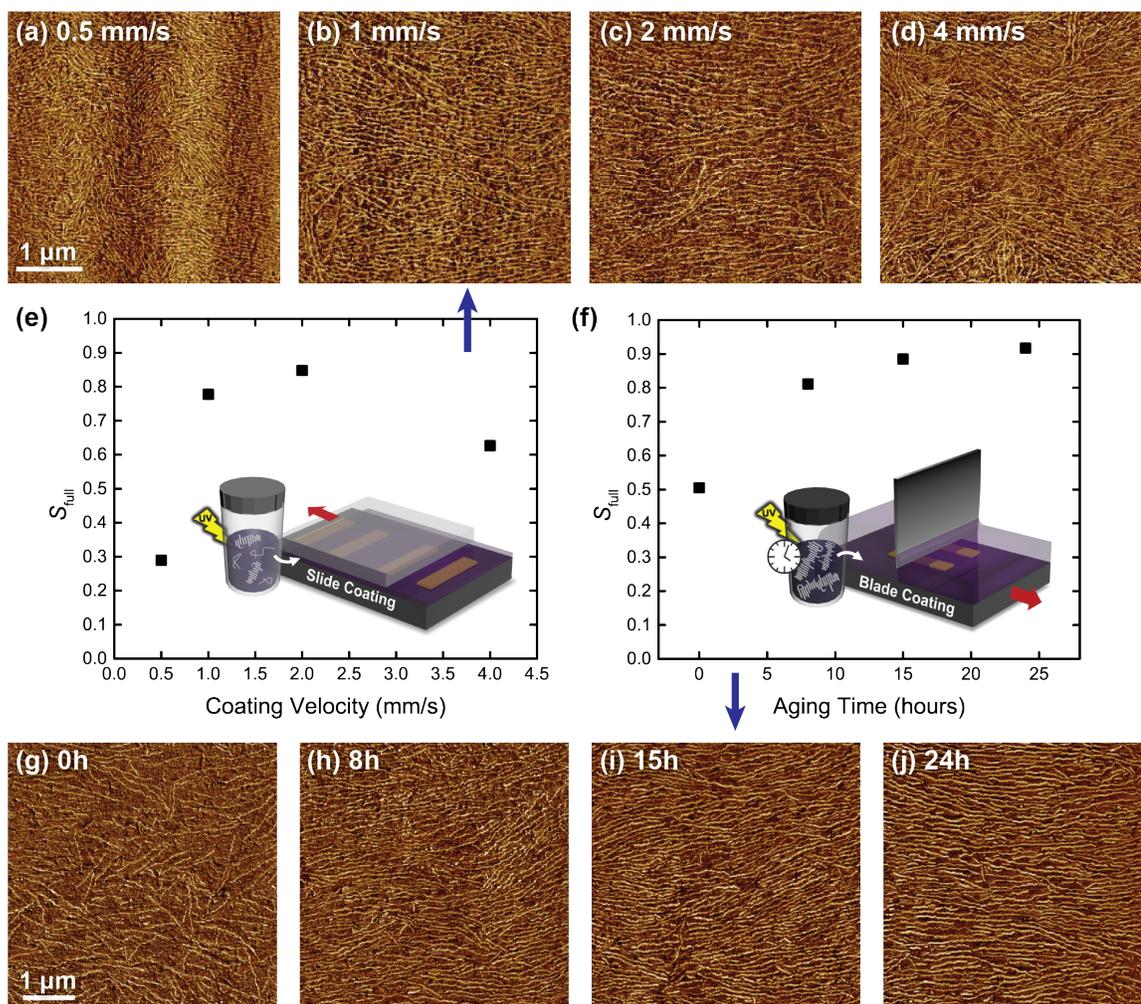


Figure 18 – The effect of blade coating parameters and solution aging on fiber alignment. (a–d) AFM tapping mode phase images (4 μm) of thin film morphology resulting from slide coating solutions of UV-irradiated P3HT nanofibers at various coating velocities. (e) Image order parameter S_{full} as a function of coating velocity for images a–d. (f) Image order parameter S_{full} as a function of solution aging time in images g–j. (g–j) AFM phase images (5 μm) of thin film morphology resulting from blade coating solutions of P3HT in which nanofibers were formed through UV irradiation and aging. Coating direction was horizontal in all images.

A similar study by Chu et al. demonstrated that mobility of up to $0.2 \text{ cm}^2/\text{Vs}$ could be obtained by combining UV irradiation, aging, and blade coating at 3 mm/s .⁴⁸ The evolution of morphology and alignment are shown in **Figure 18f–j**. As aging proceeded, alignment increased, as measured by S_{full} from AFM images and the dichroic ratio of the

0–0 transition peak from UV-vis spectra. It is possible that as fibers increased in length and number, entropic constraints forced them into greater alignment. Interestingly, anisotropic mobility was observed when measured parallel and perpendicular to the coating direction. Mobility was higher perpendicular to the coating direction, indicating that paths along P3HT chain backbones were more favorable than along π - π stacked fibers. The ability to precisely control fiber alignment should enable new developments in the theory of fibers under shear flow as well as charge transport in semi-crystalline thin films.^{59,111}

2.3.3 *Structure-Property Relationships*

Thus far, we have focused primarily on the process-structure relationships governing P3HT nanofiber formation and deposition. These experimental efforts were largely aimed at achieving higher field effect mobility. However, many factors influence the measured mobility in a given OFET beyond the thin film microstructure of P3HT, as detailed in **Appendix C**. In examining the structure-property trends observed in our OFET fabrication efforts, it is important to provide context on the processing and device architecture decisions that were made. **Table 1** compiles the process and performance information for the highest mobility devices from recent studies in the Reichmanis group.

We can infer that differences in electrical performance are due largely to differences in the thin film microstructure because our other processing decisions have remained largely consistent. All device substrates are bottom-gate, bottom-contact with an unmodified 300 nm SiO₂ dielectric, gold source and drain electrodes and a 50 μ m channel length and 2000 μ m channel width. Spin coating was performed in air at 1500 rpm for 60 seconds. Saturation regime transfer curve models were used in most cases. Almost every

study included an overnight vacuum step for solvent removal at either room temperature or 50 °C. It is important to note that the number average molecular weight (M_n) of P3HT used varied between 20 and 41 kD. While early studies indicated a strong dependence between molecular weight and mobility,³⁵ this was due in large part to the lack of inter-grain connectivity in samples with $M_n < 20$ kD.^{36,112,113} Our devices fall into the regime of interconnected aggregates, thus the differing molecular weights have a more limited impact. These are simply factors to keep in mind while comparing the performance of the processes under examination.

Table 1 – Top performing devices by processing regimen. ^aPolydispersity Index. ^bRegioregularity. ^cInitial concentration of polymer solution. ^dDeposition method: Spin = spin coated; Slide = slide coated; Blade = blade coated. ^eChannel length. ^f2-methylpentane.

Author	Year	Parameters Studied	M_n (kD)	PDI ^a	RR ^b (%)	Solvent	Init. Conc. ^c (mg/mL)	Depo. ^d	Annealing Time / Temp. (°C) / Vacuum	Transfer Curve Model	μ (cm ² /V s)
Chu ⁴⁸	2016	Aging	41	2.2	96	CHCl ₃	5	Blade	12h / 25 / Vac	Saturation	0.2
Chang ¹⁰⁹	2016	Blade Velocity	41	2.3	96	CHCl ₃	5	Slide	10m / 120	Linear	0.2
Wang ⁸³	2015	Microfluidic flow rate	20	2.2	96	CHCl ₃	5	Spin	12h / 25 / Vac	Saturation	0.16
Kleinhenz ⁵⁵	2016	Sonication, Aging	32	2.2	96	CHCl ₃	5	Spin	12h / 50 / Vac	Saturation	0.15
Chang ⁸⁸	2014	Sonication, UV	20	2.2	96	CHCl ₃	5	Spin	12h / 25 / Vac	Saturation	0.12
Choi ⁸⁰	2014	Sonication, Solvent	41	2.3	96	CHCl ₃ / 2-MP ^f	5	Spin	12h / 50 / Vac	Linear	0.1
Aiyar ⁸⁹	2011	Sonication	31	2.0	98	CHCl ₃	3.5	Spin		Saturation	0.03

Figure 19 demonstrates a linear correlation that arose between mobility and fiber alignment (S_{full}) when data from two different experiments was analyzed with GTFiber.¹¹⁴ In the first experiment, fibers were generated through sonication and aging and deposited *via* spin coating, and in the second experiment, fibers were generated through UV irradiation and aging and deposited *via* blade coating. Alignment increased with aging time in both cases, although the sonicated fibers were aged for six days, and the UV irradiated fibers were aged for only one day. Spin coating led to higher structural variance, which

was quantified by collecting three images from each device. Note that the sonicated and spin-coated films only achieve alignment up to $S_{full} = 0.5$. Even with a blade coating procedure, higher alignment was unachievable with sonicated fibers. This is likely due to their limited length; the relationship between length and alignment is explored further in **Chapter 3**.

The relationship between fiber alignment and mobility considers only fibers coated with their overall orientation perpendicular to charge transport, as illustrated in **Figure 19c**. This meant that the transistor channel was placed north of the center of spin coating, as follows from **Figure 17**, and in blade coating, the blade was drawn along the electrode width, yielding the fiber orientation shown at right in **Figure 19c**. This distinction is important for highly aligned structures because of their observed charge transport anisotropy. Aligning fibers perpendicular to the electric current allows charges to utilize chain backbones more than π - π stacks to travel across a film, requiring less use of the slower hopping mechanism. The linear nature of the correlation may be attributed to decreased bending angles in bridging chains traveling between fibers, illustrated most clearly by the theoretical results from MD + DFT simulations performed by Lan and Huang, reproduced in **Figure 19d**.³⁴ P3HT is a relatively rigid polymer because its planarizing dihedral angles are most energetically favored. In low alignment films, the presence of bridging chains is highly unlikely. However, in films with moderate alignment, or in films with a higher decay length of orientational order, bridging chains and crossing points are more likely to occur. Near-perfect alignment may reduce the bending angles of bridging chains and crossing points, which results in greater backbone planarization and thus higher mobility charge transport.

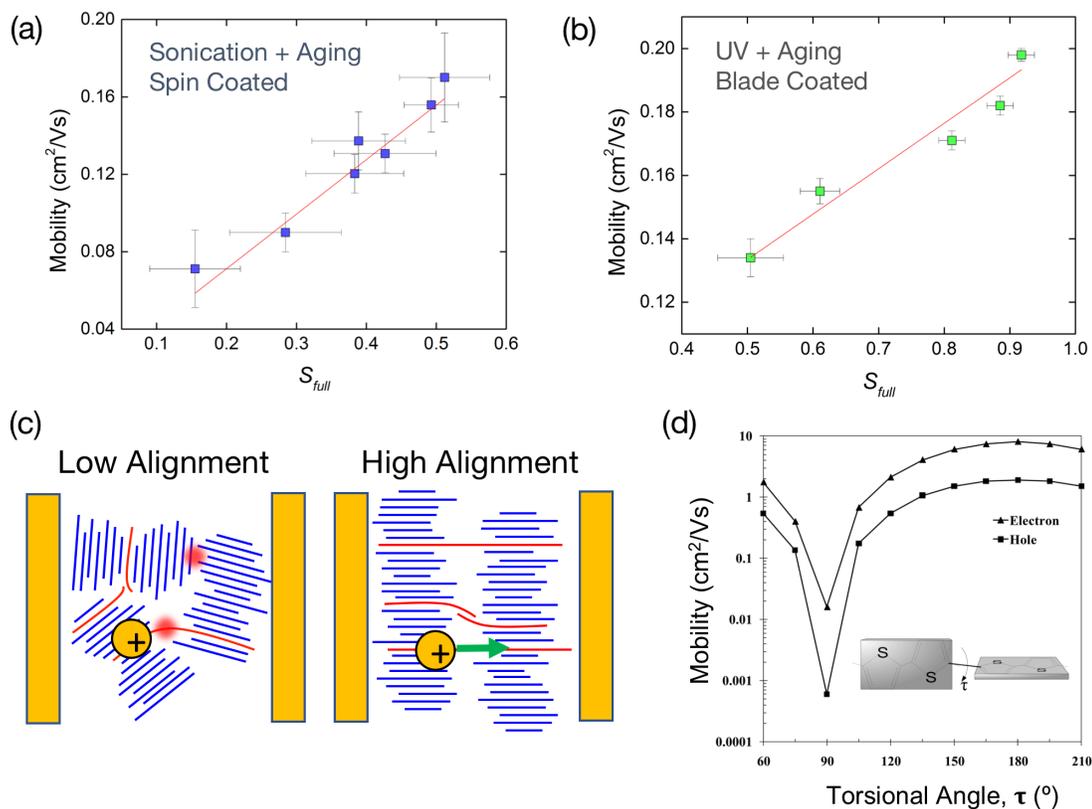


Figure 19 – Correlation between mobility and S_{full} , the image-based order parameter for in-plane fiber alignment. (a) Results from sonication, aging, and spin coating study. (b) Results from UV, aging, and blade coating study. (c) Schematic of charge transport in low and high alignment structures. (d) Relationship between P3HT backbone torsion angle and intra-chain mobility from DFT simulations by Lan and Huang. Adapted with permission from Ref. 34. Copyright 2009 American Chemical Society.

The limitations of this relationship should be noted: these are images of the thin film surface rather than the buried semiconductor/dielectric interface (although a correlation between the two is demonstrated in **Chapter 3**), and they have been collected at length scales varying from 5 – 10 μm . Nonetheless, it is an informative relationship that reflects the progress our group has made in controlling the crystallization and alignment of P3HT nanofibers, and the positive impact that has had on device performance. In general, however, an outstanding value of a single structural parameter is a necessary but not

sufficient criterion for high mobility. Performance will be maximized only when all structural properties are optimized, including a low thin film exciton bandwidth, large (100) grain size and smaller *d*-spacing, Herman's orientation factor indicating edge-on orientation, and a tightly packed and aligned fibrillar morphology.

2.3.4 Extensibility of the Relationships

There is every indication that the theory and methods introduced in this protocol are extensible to the broader class of semicrystalline conjugated polymers beyond P3HT. Analogues and copolymers of P3HT have proliferated, especially in the design of donor species for organic photovoltaics. With the intention of increasing open circuit voltage, Bronstein *et al.* added thiazole units to polythiophenes, resulting in deeper HOMO levels.^{115,116} Heeney *et al.* synthesized a selenophene-based analogue with a reduced band gap but a HOMO level similar to P3HT,¹¹⁷ as well as a selenophene-diketopyrrolopyrrole co-polymer displaying ambipolar performance in transistors.¹¹⁸ Work by the Thompson group includes the side-chain engineering of P3HT to control both open circuit voltage and surface energy.^{119,120} Some of these analogues have been shown to form the same fibrillar structures as P3HT,^{121,122} indicating that nucleation, growth and alignment strategies could be beneficial to the control of charge transport in these polymers, as well as the control of morphology in organic photovoltaics. Blending conjugated polymers with insulating polymers is another attractive strategy to reduce raw material usage while maintaining charge transport and improving barrier properties;¹²³ as such, sonication, UV irradiation and blade coating have been demonstrated on blends of P3HT with polystyrene^{124,125} and

poly(dimethylsiloxane)¹²⁶ with similar aligning effects and improvements in charge transport.

Extending these relationships beyond semi-crystalline polythiophenes and their structural analogues is trickier. More recently synthesized, higher performance conjugated polymers display only short-range aggregation, as opposed to the vivid fibers of P3HT.^{30,127–129} While this further strengthens the theory that P3HT fibers serve more as hubs that link chains rather than conduits for charge transport, it also means that the local and long-range orientational order of newer conjugated polymers is more difficult to quantify through imaging. VASE and NEXAFS measurements have been used to quantify interfacial orientation distributions of conjugated polymer backbones, but they cannot provide information on the spatial correlation of orientations.¹²⁸ For this purpose, Dark Field TEM with hyperspectral backscatter has been used effectively with PBTTT, demonstrating that a low angle between neighboring grains is beneficial to charge transport – although a quantitative metric like those presented here was not defined.¹³⁰

As process scale-up for organic electronics manufacturing becomes a stronger focus of research, precise control of electronic properties such as mobility will be necessary to meet tight performance specifications. Crucial questions to be answered include: what are the principal sources of variance in mobility for devices that were otherwise processed identically? In what stage of processing is the greatest amount of structural variability introduced, and can it be mitigated? Even for the emerging class of high-performance conjugated polymers with weaker π - π stacking interactions, long-range alignment is likely to induce charge transport anisotropy, which will be important to control in device applications.¹²⁸

2.4 Conclusions, Chapter 2

An automated image analysis protocol has been detailed for use with images of fibrillar morphologies. Each step of the processing and analysis has been characterized: anisotropic diffusion filtering, top hat filtering, thresholding, skeletonization, and orientation mapping, followed by analysis of the orientation distribution and decay of orientational order. This combination of processing and analysis builds off of previous work in other fields, however it is unique in its robustness to noise and low-contrast images, which are sometimes unavoidable with AFM. The analysis workflow was implemented as a standalone application for Mac OS or Windows, giving readers with no computational experience the ability to implement this procedure and to explore the effect of each image processing parameter on their own images. Recommendations are provided for each parameter setting corresponding to physical feature sizes: orientation smoothing should not exceed fiber width, while top hat filter size should exceed fiber width. Diffusion time should be set between one and ten “seconds.”

Quantitative analysis of morphology can be accomplished with three primary tools: average orientation, decay length, and S_{full} . Average orientation can be used to capture the direction of anisotropy of a population of fibers, which is frequently observed with flow- or strain-oriented samples. It was used in this chapter to elucidate the orienting effect of the spin-coating process on P3HT morphology. Decay length and S_{full} can be used to describe the decay of orientational order as well as its long-range asymptotic value. These parameters are used to quantify the differences in fiber bundling caused by different solution pre-processing methods: sonication and aging as well as microfluidic pre-processing create enhanced local orientational order that is maintained through the spin-

coating process, while sonication and aging in the presence of a poor co-solvent does not. Mechanistic hypotheses for these phenomena can be formulated by combining this analysis with the literature on conjugated polymer crystallization. While this technique was demonstrated on P3HT-based materials, the introduction of this software should bring a valuable level of quantification to the analysis of process-structure-property relationships in any fibrillar system. The software is available for download at [gtfiber.github.io].

Finally, a linear correlation between S_{full} and mobility was demonstrated. This correlation holds so long as overall fiber orientation is perpendicular to charge transport. It is likely due to increased inter-grain connectivity, and also the specific nature of inter-grain connectivity. As fibers come into near-perfect alignment, so too do the bending angles of tie chains between neighboring grains. The energetic favorability of this conformation reduces charge traps, allowing for freer motion of charge carriers through the film.

Chapter 3. Packing, Defects, and Growth

Mechanisms in Conjugated Polymer Nanofibers

Evaluated by High-throughput Image Analysis

In the previous chapter, the image processing strategy behind GTFiber was introduced, then used to analyze orientational order in a variety of fiber packing scenarios. However, the data presented was relatively limited and was used primarily for illustrative purposes. Additionally, fiber length was excluded from the analysis because of the difficulty of reliably extracting it. A key difference between the image analysis approach in this chapter versus that of the previous chapter is *vectorization*. In GTFiber 2.0, the fibrillar structure is still skeletonized, but the skeleton is then broken into smaller segments, fitted with vectorized contours, then reconstructed, permitting the measurement of fiber length. It is shown that fiber length is an additional crucial factor in the process-structure-property relationships for P3HT, with a strong relationship to alignment and packing defects. Some details on the orientational order parameters from the previous chapter are repeated here for completeness.

3.1 Introduction

Conjugated polymers are driving a revolution in printable electronics. These mechanically flexible, semiconducting materials are being incorporated into devices ranging from solar cells to biosensors.^{6,11,14} Conjugated-polymer-based organic field effect transistors (OFETs) are on the brink of commercial viability for use in flexible display

applications.¹⁸ Enabling these advances is a growing fundamental understanding of the process-structure-property relationships governing conjugated polymer thin films. While novel synthetic conjugated polymers continue to push the boundaries of performance, greater control of the processing of classical semi-crystalline polymers has led to a refined understanding of polymer crystallization, assembly, and long-range microstructure.^{50,128,131,132} Recent progress by our group in controlling the nucleation, growth and alignment of poly(3-hexylthiophene) (P3HT) nanofibers in OFET devices has produced a rich library of images of fibrillar morphologies containing valuable information on the microstructural evolution of conjugated polymer thin films.^{48,55,80,83,109,133} Analysis of these morphological images presents challenges in image processing and computer vision, structural order parameter identification, and data visualization. We present solutions to these problems here as part of an update to GTFiber, our open-source application for the automated segmentation and analysis of images of fibrillar materials. We then discuss the process-structure-property relationships learned from analyzing over 100 images of polymeric transistor morphology spanning a wide range of solution processing techniques.

Extraction and analysis of fibers from images has received sporadic attention over the past two decades, but is becoming more relevant due to a significant uptick in research on fibrillar materials and composites, including carbon nanotubes,⁷⁰ cellulose and other bio-based fibers,¹³⁴ self-assembling biological systems such as amyloid fibers,¹³⁵ and the present case of conjugated polymer nanofibers.^{50,136,137} Images of fibers are highly diverse and their analysis can be approached in different ways. When the number of fibers to be analyzed is low, manual tracing is the most effective method of extraction. This approach

has been greatly improved by FiberApp, a MATLAB-based software package that fits vectorized contours to fibers provided that the user clicks on multiple points along each fiber's backbone.⁷⁴ However, this approach becomes impractical as the number of images and the number of fibers per image increases. In the closely related problem of worm behavioral analysis, a machine learning approach was used effectively as part of the WormToolbox in the CellProfiler software package.⁷¹ This approach requires the generation of a large training set of manually traced worms (fibers), and assumes that worms will conform to the trained distributions of length, width, and curvature. An alternative method that yields accurate, useful results with little or no manual tracing is a heuristic skeletonization and network analysis approach, employed by both ADAblock for block copolymer analysis and FIRE, a package for collagen fiber analysis.^{66,72,73} In previous work, we introduced a similar heuristic skeletonization algorithm capable of analyzing bulk orientational order in images of fibrillar materials, but it did not extract and vectorize individual fibers, limiting its measurement capabilities.¹¹⁴

GTFiber has been updated to integrate the previous skeletonization approach with the fiber fitting and analytical algorithms from FiberApp, yielding a robust and accurate image analysis workflow, complete with a user-friendly standalone application for both Windows and Mac [gtfiber.github.io]. In the updated image processing scheme, images are de-noised and enhanced using an anisotropic diffusion filter, then thresholded and skeletonized. Isolated skeletal segments are fit with the Active Contours algorithm, then reconstructed using a scored segment-matching algorithm. The result is a software package that produces rapid, detailed measurements and helpful structural visualizations without the need for manual tracing.

In previous studies, we used our image analysis protocol to examine the relationship between thin film deposition method and orientational order, quantifying both the radial alignment inherent to the spin-coating process and the linear aligning effect of blade coating on P3HT nanofibers, as well as the relationship between fiber alignment and charge carrier mobility.^{114,138} Many techniques to induce the nucleation and growth of P3HT nanofibers have been previously reported: heating/cooling cycles, mixed solvent approaches, simple aging and shear-induced growth techniques are some notable examples.^{38,41,49,104} Here, sonication, poor solvent addition, UV irradiation, microfluidic processing and aging are all explored for their quantitative effect on fibrillar morphology, revealing a complex but highly tunable design space. Three order parameters are sufficient to capture the structural variation in this space: global alignment (S_{full}), fiber length density (ρ_{FL}), and the decay length of orientational order (λ_C). These structural parameters are illustrated through visual examples relating them to different processing techniques. Furthermore, a web-based interactive data visualization is introduced for datasets of structural imagery.

Finally, given recent progress in obtaining macroscopically aligned fibers by multiple groups,^{48,109,139} we perform a detailed analysis of fiber packing in highly aligned thin films at both the air and substrate interfaces. We show that perfect alignment is currently limited by packing defects due to shorter fibers that preferentially segregate to the substrate interface, as well as local nano-scale curvature along the backbone of P3HT nanofibers. A self-assembly mechanism is proposed that reconciles both the nanofiber curvature and the log-normal distribution of fiber lengths inherent to populations of P3HT nanofibers. We expect these software tools and results to be of interest specifically to the

organic electronics community and more broadly to researchers working on fibrillar materials and composites of any kind.

3.2 Materials and Methods

3.2.1 Materials

Regioregular poly(3-hexylthiophene) (P3HT) was obtained from Rieke Metals. Number average molecular weight (M_n) ranged from 17 kD to 40 kD, with polydispersity between 2.0 and 2.2. Regioregularity was greater than 96%. Chloroform (stabilized with amylenes) and 2-methylpentane (anhydrous) were obtained from Sigma-Aldrich.

3.2.2 Solution Processing

Solutions were prepared at 5 mg/mL by dissolving 10 mg of polymer in 2 mL of chloroform (Sigma) in 20 mL borosilicate glass vials, tightly capped, and heated to 60 °C for 25 minutes on a hot plate (Corning) to ensure complete dissolution. After cooling at ambient conditions for five minutes, the solutions were subjected to one or several of the following treatments: for sonication, the solution vial was dipped in a bath sonicator (Branson 2510, 40 kHz, 130 W) filled with tap water for two minutes. For UV irradiation, the solution vial was placed on top of a handheld UV lamp (Entela UVGL-15, 5 mW cm⁻², 254 nm) which was placed on top of a magnetic stirrer (Corning), and the solution was irradiated for 8 minutes while being stirred at 300 rpm. For poor solvent addition, 2-methylpentane was slowly added to the solution of P3HT and chloroform at concentrations ranging from 0 – 40 vol%. For aging, the treated solution was capped, wrapped with parafilm, and placed in a dark drawer for the specified amount of time. Microfluidic

processing involved pumping the prepared solution through a PTFE tube (300 μm i.d.) with a syringe pump at a flow rate of 1 mL/min; the tube passed through an ice bath (residence time 1s) and a zone under UV irradiation (residence time 10s) before exiting to a collection vial for deposition.

3.2.3 OFET Device Fabrication

OFET devices with bottom-gate, bottom-contact architecture were fabricated for the electrical characterization of thin films deposited from the as-prepared solutions. Highly n-doped silicon wafers with a thermally grown 300-nm SiO_2 dielectric layer were used as the substrate. Source and drain electrodes were patterned by a photolithography lift-off process in a cleanroom environment, and deposited by E-beam evaporation (Denton Explorer) of 50 nm of Au with 3 nm of Cr as the adhesion layer. The device substrates were rinsed with acetone, methanol, and isopropanol, sonicated for 15 minutes in acetone, and cleaned for 30 minutes in a UV-ozone cleaner (Novascan PSD-UV) to remove residual photoresist and other organic contaminants.

Thin film deposition was carried out *via* spin coating (WS-650MZ-23NPP, Laurell) of a 15 μL droplet at 1500 rpm for 60 s, or blade coating at a velocity of 3 mm/s. Blade coating was conducted on a motorized linear stage (A-LSQ150A- E01, Zaber) equipped with a vacuum chuck, with the blade perpendicular to the substrate at a gap height of 5 μm . A 4 μL volume of fluid was injected into the blade gap and allowed to wick the width of the blade fully before coating. Coated thin film devices were stored overnight under vacuum before any characterization was performed.

3.2.4 Electrical Characterization

Individual OFET channels were tested in a nitrogen-filled glovebox using an Agilent 4155C semiconductor parameter analyzer. The field-effect hole mobility (μ) was calculated in the saturation regime of transistor operation ($V_{DS} = -80$ V) by fitting the following equation to a plot of drain current (I_D) versus gate voltage (V_G):

$$I_{DS} = \frac{WC_{OX}}{2L} \mu (V_G - V_{th})^2 \quad (3)$$

where W (2000 μm) and L (50 μm) are the transistor channel width and length, respectively, V_{th} is the threshold voltage, and C_{OX} is the capacitance per unit area of the SiO_2 dielectric (1.15×10^{-8} F/cm²).

3.2.5 Atomic Force Microscopy

Thin film surface morphology was characterized with a Bruker Dimension Icon atomic force microscope operating in tapping mode with n-type silicon tips (HQ:NSC14-noAl, 5 N/m, 160 kHz, MikroMasch). Images were collected predominantly at a 5 μm scan size with 512 samples per line at 1 line/s, but some images were collected at 2, 4, 7, and 10 μm scan sizes as well. Increased drive amplitudes frequently yielded higher contrast between the fiber and amorphous phases in the phase channel. Approximately 3–5 images could be obtained per tip before loss of image quality due to tip degradation and polymer adsorption.

3.3 Image Data and Analysis

3.3.1 Solution Processing and Thin Film Deposition

The solution processing methods and raw data considered in this study are illustrated in **Figure 20**. The images in **Figure 20b-e** are AFM tapping mode phase images taken from thin films of regioregular P3HT. The fibrillar regions are π - π -stacked crystallites that are dispersed in an otherwise disordered matrix; film thickness ranges from 20–50 nm. Solutions of P3HT (M_n ranging from 20 – 41 kD) in chloroform were prepared at a concentration of 5 mg/mL in all cases. Nanofibers were nucleated using either UV irradiation or sonication, resulting in the appearance of small fiber nuclei, as seen in the blade coated film shown in **Figure 20b**. Further growth was promoted by either allowing solutions to age or by addition of a poor co-solvent, in this case 2-methylpentane. **Figure 20c** shows a thin film morphology after a sonicated solution was aged for two days and blade coated. Nucleation and growth were combined in a single microfluidic flow system through a cooling and UV-assisted growth step, with the resulting morphology shown in **Figure 20d** after spin coating. Finally, a highly aligned morphology is shown in **Figure 20e**, the result of blade coating a solution processed with UV and aged for two days. Further details on the processing methods are available in the Methods section or in the papers that introduced them,^{48,80,83,88–90,133} their nucleation and growth mechanisms were reviewed as well.¹³⁸

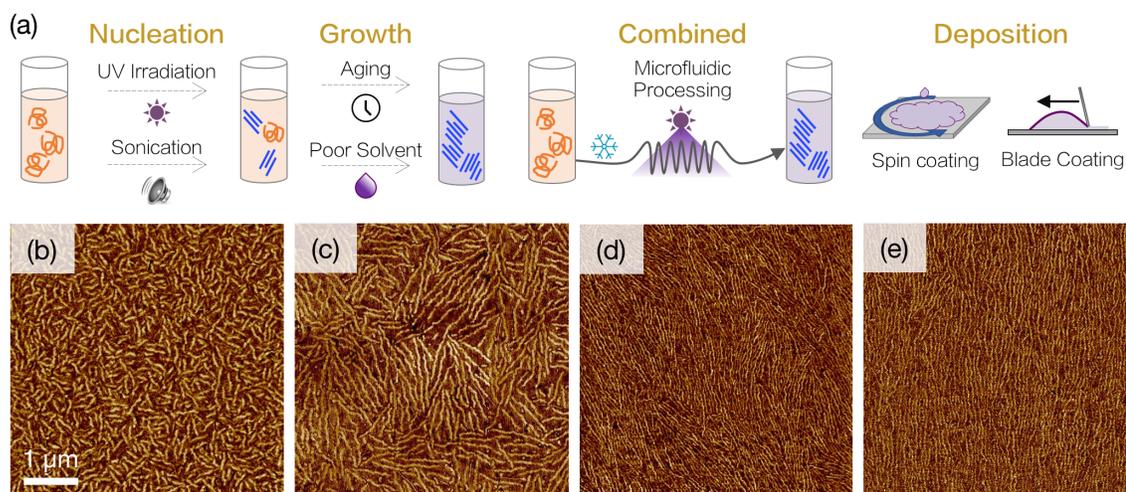


Figure 20 – Processing of P3HT nanofibers and progression of fibrillar morphologies as observed in 5 μm AFM phase images. (a) Illustrations of solution processing and thin film deposition methods considered in this study: UV irradiation and sonication induce nucleation, while aging and poor solvent addition lead to extended growth. Nucleation and growth are combined in a microfluidic-cooling-UV system. Thin film deposition methods considered include spin coating and blade coating. (b) Small fiber nuclei after sonication. (c) Longer fibers formed after aging a sonicated solution for two days. (d) Densely packed fibers formed in the microfluidic process and deposited by spin coating. (e) Highly aligned fibers formed by UV irradiation and two days of aging followed by blade coating at 3 mm/s.

The morphological space spanned by the above processing methods is highly diverse: the four images presented in **Figure 20** give some idea as to the range of fiber lengths, packing densities, and orientation order present in this system, but qualitative observations from raw images are insufficient to convey and quantify the process-structure-property relationships present in this material system. Morphological imagery was available from 137 unique samples: 33 images sourced from previously published studies and the remaining 104 consisting of either previously unpublished data or samples fabricated specifically for this study to fill out the process parameter domain as well as the morphological space. Image sizes range from 2 to 10 μm , but most of the analysis in this study is limited to 5 and 7 μm images to ensure consistency in structural measurements,

which brings the dataset down to an even 100. To obtain a satisfying analysis, it is necessary to extract fibers and quantify their length distributions, packing behavior and orientational order in terms of structural order parameters. We begin by outlining the fiber extraction and analysis procedure developed for this study and implemented in the GTFiber software package.

3.3.2 *Image Processing*

Starting from the phase channel of a tapping mode AFM image taken from a P3HT thin film, an image is processed through an anisotropic diffusion filter (**Figure 21b**), an adaptive thresholding step (**Figure 21c**), and a skeletonization step (**Figure 21d**). This sequence was used to analyze bulk orientational order, as detailed in a previous protocol.¹¹⁴ However, measuring fiber counts, lengths, and packing density, among other properties, requires the *vectorization* of individual fibers; that is, identifying which pixels belong to the same fiber, and converting those pixels into a string of vectors, referred to as a contour. This is the approach used in FiberApp,⁷⁴ in which users manually select points from which contours are initialized. Here, we provide an initial guess for our contours by breaking the skeletonized image into isolated, unbranched strings of pixels (segments) and feeding those segments to the Active Contours algorithm as illustrated in **Figure 21e**. This algorithm is discussed in detail in **Appendix B**.

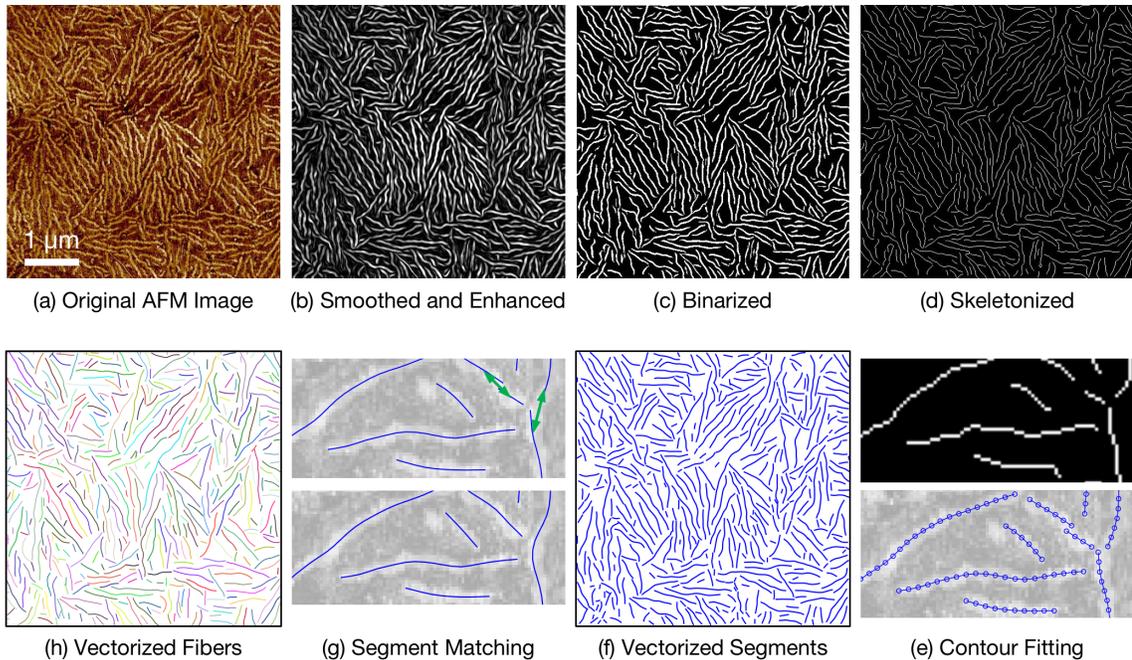


Figure 21 – Extraction and vectorization of fibers from an AFM image using GTFiber. (a) A typical AFM phase image of P3HT nanofibers, (b) Result of anisotropic diffusion and top hat filtering, (c) Result of adaptive thresholding, (d) Result of skeletonization and fringe removal, (e) Fitting vectorized contours to the unbranched skeletal segments using Active Contours at a 30 nm step length, (f) Line plot of the vectorized segments, (g) Illustration of segment matching and fiber reconstruction, (h) Line plot of vectorized fibers in which each fiber is assigned a random color to evaluate classification accuracy.

Overlapping fibers or fibers with adjacent ends tend to be combined into larger branched areas of the skeleton, so after breaking the skeleton into unbranched, vectorized segments, these segments must be reconstructed into linear fibers, illustrated in **Figure 21g**. Segments whose endpoints are close to one another are matched based on two user specified criteria: the maximum tolerable gap between the segments and the additional curvature created by stitching them together. Two segments are matched if the gap between their endpoints is less than the maximum gap, and if stitching them does not introduce significant additional curvature to their contours. Two matches are shown with green arrows in **Figure 21g**.

The result of this sequence of image processing steps is represented by **Figure 21h**, an image of the vectorized fibers in which each fiber is represented by a different random color to demonstrate the performance of the reconstruction. While reconstruction is never a perfect process, strong visual agreement between the original image in **Figure 21a** and the vectorized representation in **Figure 21h** indicates that the segmentation and vectorization were effective. To quantify the accuracy of the algorithm, manually traced images were used to train the image processing parameters of GTFiber, presented in **Appendix B**. With properly selected parameters, GTFiber yields bulk structural measurements within 10% of their true values. The initial analysis in this study is focused on broad trends across the entire dataset, which are insensitive to image processing parameter selection. The more detailed results, especially those concerning fiber length, were confirmed with manual tracing and are presented with uncertainty quantification. **Appendix B** contains a complete discussion of accuracy, sensitivity, failure modes and limitations of the software, as well as a guide to the user interface and parameter selection.

3.3.3 Structural Visualizations and Order Parameters

Extraction of vectorized fiber backbones permits the calculation of several structural order parameters, as introduced in **Chapter 2**. Besides fiber length and width distributions, fibrillar structures can be characterized by their full-image-scale alignment (S_{full}), fiber length density (ρ_{FL}), and the decay length of orientational order (λ_C). The order parameters are illustrated in **Figure 22**. Each image is plotted in terms of S_{full} and ρ_{FL} in **Figure 22b** and in terms of S_{full} and λ_C in **Figure 22c**, colored according to their solution processing method and shaped according to their deposition method. Representative

examples from each plot are shown in **Figure 22d-j** as false-colored Orientation Maps, in which the color of each fiber indicates its in-plane orientation as labeled on the attached color wheel in **Figure 22d**. Since fibers are vectorized at a 30 nm (3 pixel) step length, the orientation, and thus the color, varies along the length of each fiber. The Orientation Map makes the high-resolution AFM images readable even at small printed sizes, and helps to visually identify oriented domains of fibers.

Fiber alignment (S_{full}) and decay length (λ_C) are both extracted from a plot of orientational order calculated at increasing length scales, as illustrated in **Chapter 2 and Appendix B**. S_{full} (for “full-image alignment”) captures the alignment of fibers in an image, regardless of their orientation or packing density. **Figure 22d, e, g and i** provide examples of low alignment structures, while **Figure 22f, h and j** are examples of high alignment structures. Unlike Herman’s Orientation Factor, used for analysis of orientation in X-ray scattering, S_{full} has a minimum value of 0, indicating isotropic orientation.

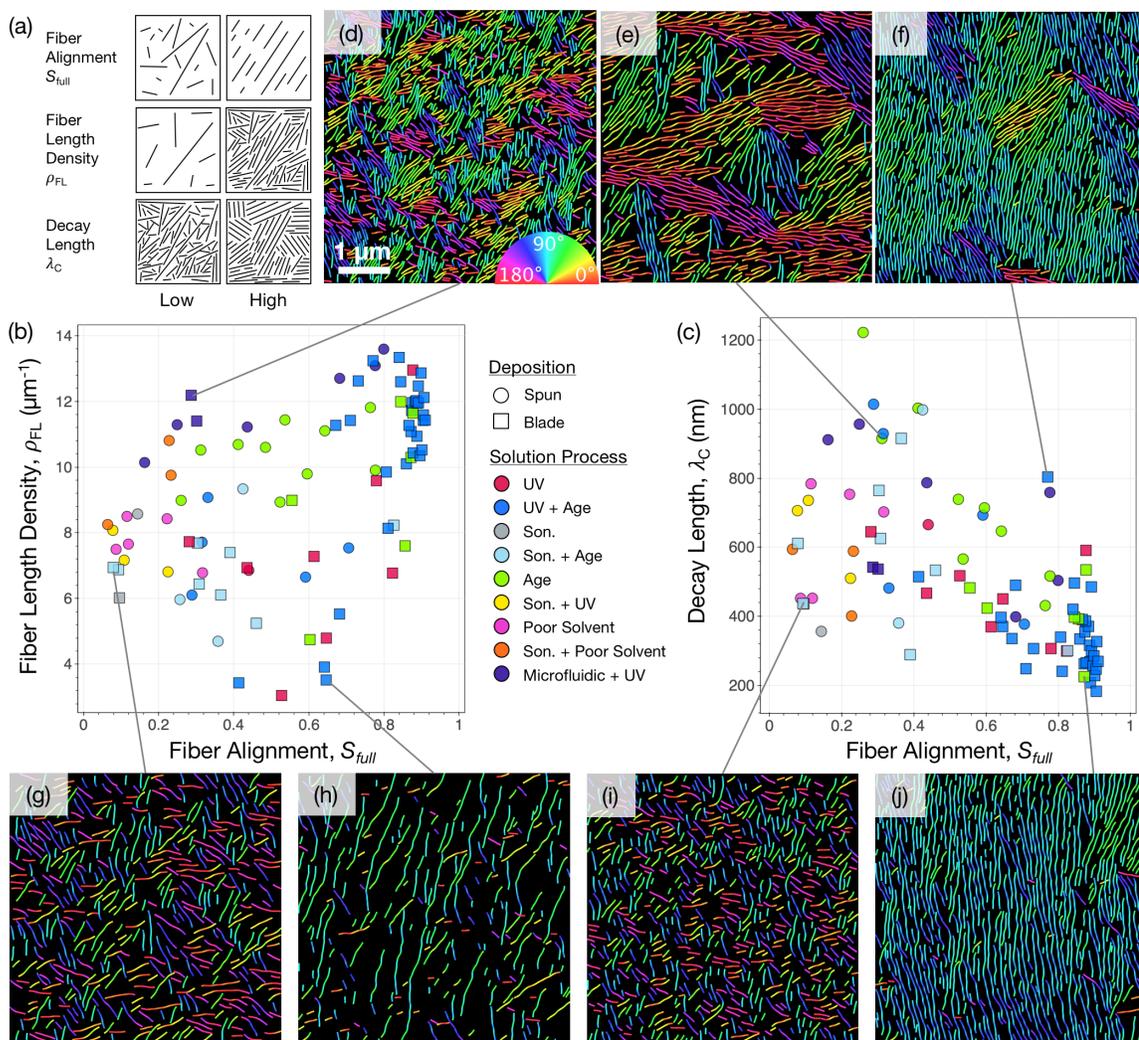


Figure 22 – Processed images of fibrillar morphologies and their coordinates in the space defined by the order parameters S_{full} , ρ_{FL} , and λ_C . (a) Schematic illustrating fiber packing arrangements typical of low and high values of the three structural order parameters. (b) The complete structural image library plotted in terms of S_{full} and ρ_{FL} , with solution processes coded by color and deposition methods coded by shape. (c) Image library plotted in terms of S_{full} and λ_C . (d-j) Orientation Maps extracted from raw images at extreme values of each order parameter, with their location in each plot indicated by a gray pointer. Each image is 5 μm , and colors correspond to the in-plane orientation of each fiber’s backbone, as indicated by the attached color wheel in d (i.e. red is horizontal, cyan is vertical).

The decay length (λ_C) captures local alignment behavior. For example, in **Figure 22e and f**, fibers appear in large aligned bundles even if the overall structure is not fully

aligned – this is characteristic of a high decay length. By contrast, **Figure 22i** and **j** represent structures with *low* decay length and low and high overall alignment, respectively. In these cases, colors are more evenly distributed across the images, meaning that fiber orientation is more evenly distributed. In an image with low decay length, fibers' orientations are not as strongly influenced by that of their neighbors. With low decay length and high alignment (**Figure 22j**), alignment is disrupted only by local defects such as the isolated horizontal red fibers, rather than by long range variations in orientation as seen in **Figure 22f**.

Fiber length density (ρ_{FL}) captures the packing density of fibers and is defined as the total length of fibers per unit area. Its calculation is illustrated in **Appendix B**. It is similar to the volume fraction of fibers but is not influenced by their width. Examples of low length density are shown in **Figure 22g** and **h** for low and high alignment, respectively. By contrast, high length density with low alignment is represented by **Figure 22d**. Achieving high length density with low overall alignment generally requires short fibers packed in tight, locally aligned bundles. High density and high alignment is well represented already by **Figure 22j**. Intuitively, images with more color than black have higher fiber length density.

Interestingly, thin films of P3HT nanofibers tend to have spatially homogeneous fiber length density – that is, films are rarely observed in which tight bundles of fibers are accompanied by large voids. This likely indicates that fiber packing in thin films is dictated by entropic constraints. When fiber length density is homogeneous, it is inversely correlated with average inter-fiber spacing; this can be useful when trying to estimate the feasibility of inter-grain tie chains. For example, in a film where $\rho_{FL} = 10 \mu\text{m}^{-1}$, the average

inter-fiber spacing would be 100 nm. For P3HT chains with $M_n = 40$ kD, an average chain has a fully stretched contour length of 92 nm (assuming repeat unit MW = 166 g/mol and length = 0.38 nm), indicating that an average chain in this example could indeed bridge the gap between neighboring fibers. This is quantified more rigorously in a later section.

Fiber length and width distributions, while not technically structural order parameters, are also calculated as part of the fiber vectorization process. It should be noted that mean fiber length as reported here is a number average, rather than a length average. Similar to polymer molecular weight distributions, short fibers are more populous by number even though more material may be contained in long fibers. Details on the calculation of all order parameters and structural metrics are contained in **Appendix B**.

3.3.4 Interactive Exploration of Materials Image Datasets

The image dataset presented in **Figure 22** is difficult to fully convey in a static figure. As part of this study, we introduce a web-based interactive version of **Figure 22b** and **c** built with the Bokeh library for Python, illustrated in **Figure 23** and available at [ZoomImgs.github.io]. When fully zoomed out, the plot looks exactly like those in **Figure 22b** and **c**. However, a user can zoom in on any data point by scrolling, similar to Google Maps, revealing the underlying image that produced each data point. The images fade into view as the zoom level increases to avoid showing too many overlapping images at once, as shown in **Figure 23a**. Furthermore, when the mouse is hovered over a data marker, a

box appears with all of that image’s structural measurements as well as its processing information, as shown in **Figure 23b**.

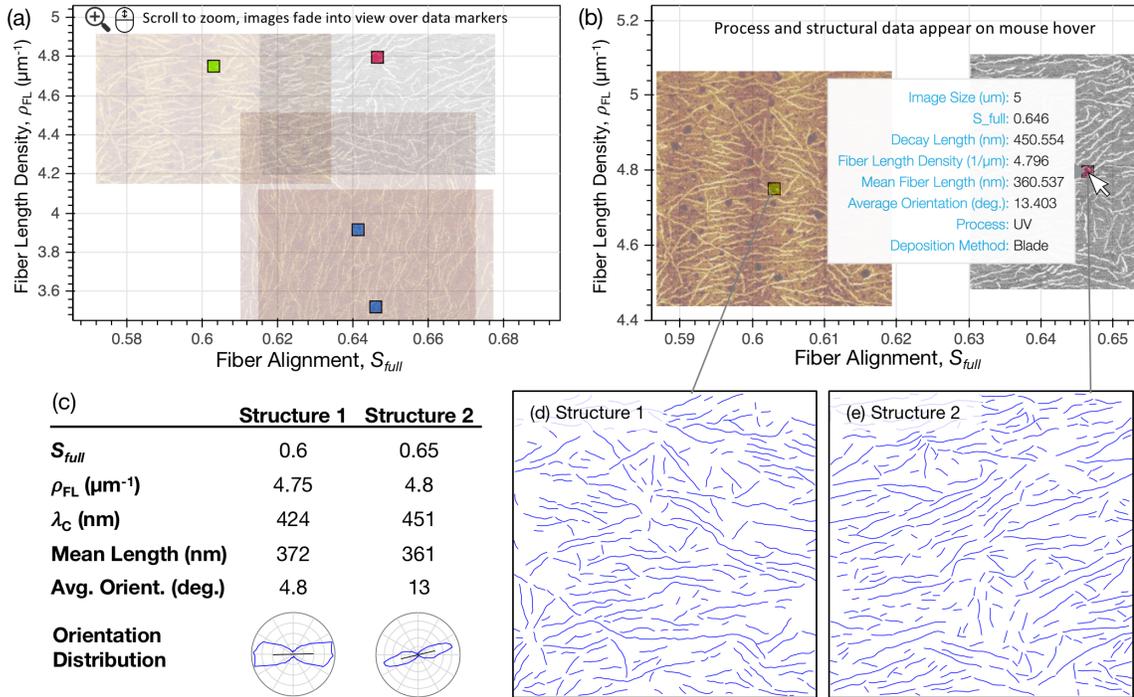


Figure 23 – Web-based interactive data visualization for materials image data. (a) Each data point represents one image; as a user zooms in on the plot by scrolling (similar to Google Maps), the underlying image for each data point fades into view. (b) Hovering the mouse over a data marker reveals the structural measurements and processing information for that image. (c) Comparison of structural order parameters and measurements for the two images shown in b. (d) Vectorized fiber plot of left structure in b. (e) Vectorized fiber plot of right structure in b.

This tool can be helpful for understanding the structural metrics used in this study. It is also useful for exploring structural libraries: for example, it can reveal similar structures obtained by different processing methods: the two images shown in **Figure 23b** have nearly identical structural metrics, but one was processed by sonication, aging, and blade coating while the other was processed by UV and blade coating. The two structures can be compared by the plots of their vectorized fibers in **Figure 23d** and **e**, as well as with the table in **Figure 23c**. In terms of their place in the broader dataset, these structures have

relatively high alignment and decay length, low fiber length density, and moderate fiber length. These parameters are now explored as they relate to solution processing, deposition, and electrical properties of thin films of conjugated polymers.

3.4 Results and Discussion

3.4.1 Process-Structure-Property Relationships

Figure 22 was intended as a demonstration of the structural variance captured by S_{full} , ρ_{FL} , and λ_C ; to investigate trends, each structural metric is plotted as a function of solution processing technique in **Figure 24**, sorted in order of decreasing values of each metric. In **Figure 24a**, it is apparent that UV, aging, and a combination of the two yield films with the highest alignment. It is no coincidence that these techniques also yield the longest fibers, as evidenced by **Figure 24d**. In fact, mean fiber length and fiber alignment are generally strongly correlated with one another, as can be seen in the correlation heat map provided in **Figure 24f**. This is not simply an artifact of the limited image size: when fibers are tracked over periods of aging, they tend to increase in both length and alignment when deposited, likely due to the entropic constraints mentioned earlier.^{48,55}

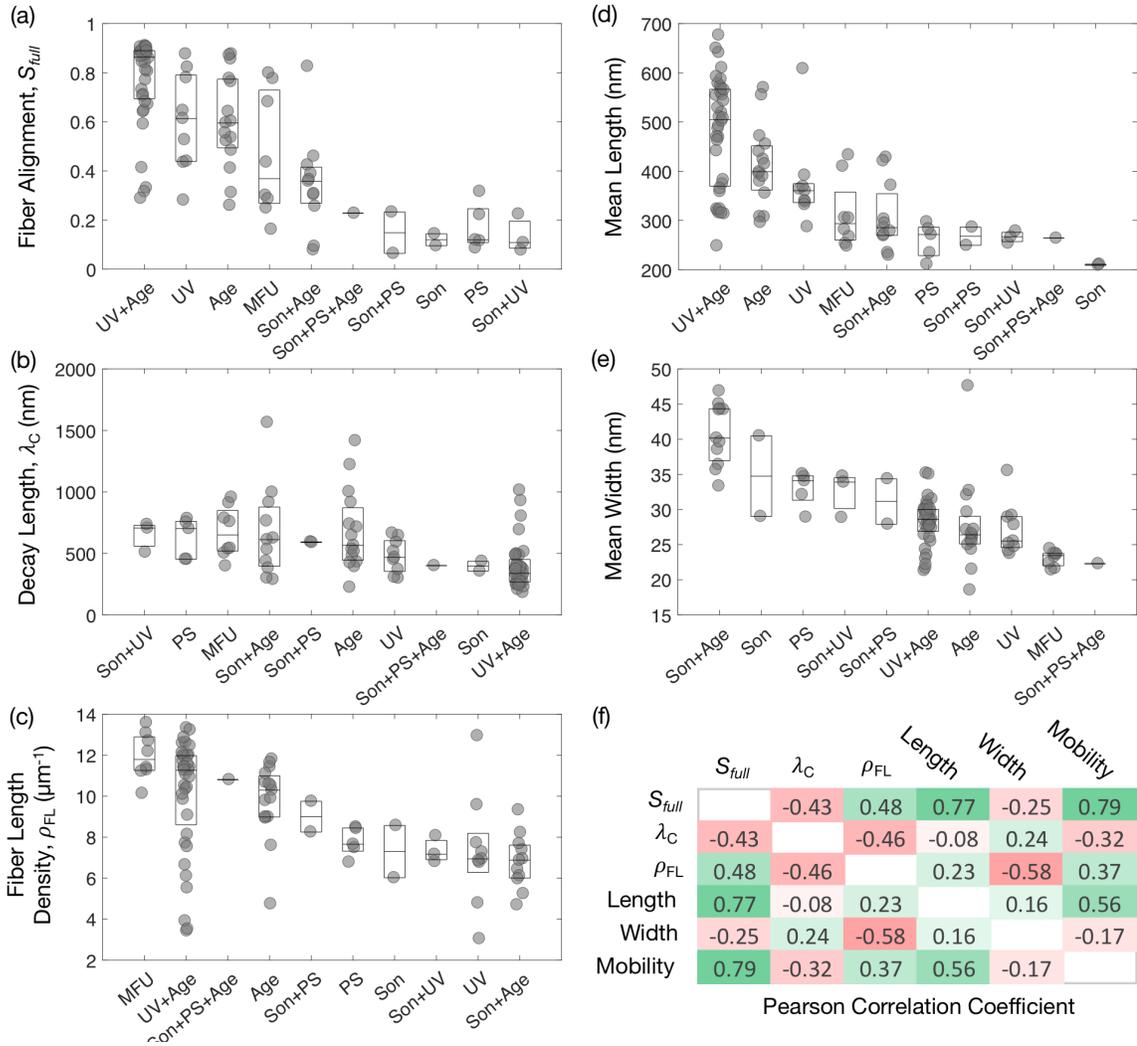


Figure 24 – Process-structure-property relationships among the image dataset, considering only 5 and 7 μm images. (a-e) Each structural metric across all solution processing conditions, sorted in order of decreasing median values of that structural metric. Each point represents one image. Slight jitter has been introduced to the x-coordinate of data points to improve readability in dense regions. Boxes span from the 25th to the 75th percentile of each point cloud, with a line at the median. (f) Correlation analysis of structural metrics with each other as well as with charge carrier mobility, where data was available. Darker green indicates strong direct correlation, while darker red indicates strong inverse correlation.

Another interesting feature of **Figure 24** is the inverse correlation between fiber width and fiber length density. Microfluidic processing yields substantially higher fiber length density than any other process, likely because fiber growth occurs in an oriented

flow environment where inter-fiber tie chains are more readily formed, ensuring that fibers remain closely packed during deposition.^{104,108,139,140} Conversely, sonicated solutions result in thin films with a consistently low ρ_{FL} , as well as observably wider fibers, even when molecular weight is controlled for (see also **Figure 27a**). This runs counter to previous reports that nanofiber width is correlated with M_n only up to about 20 kD, after which width plateaus due to chain folding.^{57,141} Besides being known as a nucleating agent, sonication has been shown to disentangle polymer solutions.⁸⁵ It is possible that while disentangling free polymers with sonication allows them to form crystals with less folds, and thus a wider fiber body, it also limits their ability to form inter-fiber tie chains, resulting in greater space between fibers and the observably low fiber length density.

Structure-property relationships are more difficult to infer from this dataset given that most images are from the top surface of the thin films, but the structure at these surfaces is at least correlated, as demonstrated by previous studies and later in **Figure 26**. As discussed in **Chapter 2** and shown again in **Figure 24f**, charge carrier mobility is most strongly correlated with fiber alignment. Perhaps counter-intuitively, this is not because of percolation through a large network of fibers; rather, fiber alignment is indicative of underlying polymer chain alignment. The fibers simply serve as hubs for charges to find the next long polymer chain to travel along.³⁶ This is confirmed by the charge transport anisotropy observed in devices with aligned fibers: charge transport is faster *perpendicular* to aligned fibers than parallel to them.⁴⁸

The observed correlation between alignment and mobility is the reason for the high concentration of highly aligned structures ($S_{full} > 0.8$) produced using UV irradiation, aging, and blade coating. These data points reflect the recent progress our group and others have

made in obtaining highly aligned nanofibrillar structures with enhanced charge transport,^{48,107,109,124,139,142,143} and are a testament to the repeatability of the process: images for this technique were collected by at least four different co-authors of the current study. UV, aged, and blade coated films that fall *below* an S_{full} of 0.8 were generally aged for less than 24 hours; aging between 24–48 hours has proven to be the optimal process condition. However, this begs the question of why alignment seems to reach a maximum around $S_{full} = 0.9$. To address this question, we turn to an aspect of polymeric semiconductors that has received relatively little attention: defects.^{58,59}

3.4.2 *The Role of Fiber Length*

Obtaining perfect nanofiber alignment is limited by two factors: packing defects and local fiber curvature. To better understand nanofiber packing during thin film deposition, we turn back to the strong correlation between fiber length and fiber alignment, as shown in **Figure 25**. It appears that beyond a mean length of 500 nm, high fiber alignment is much more likely, especially when using blade coating as a deposition method. Blade coating in and of itself is not a guarantor of alignment and consideration must be taken for the coating rate and regime of operation,¹⁴³ but it is worth noting that the low alignment outliers are mostly spin coated while the high alignment outliers are blade coated. Theoretical treatment of suspended fibers under flow is highly dependent on the specific situation. For concentrated colloidal rods at a low Peclet number (ratio of advective to diffusive transport rate), it has been observed that higher aspect ratio rods display greater alignment.¹⁰⁵ While the diffusive properties of P3HT nanofibers are not known, this

observation is in general agreement with other studies demonstrating that high-aspect-ratio objects align with the flow direction.^{70,111,144,145}

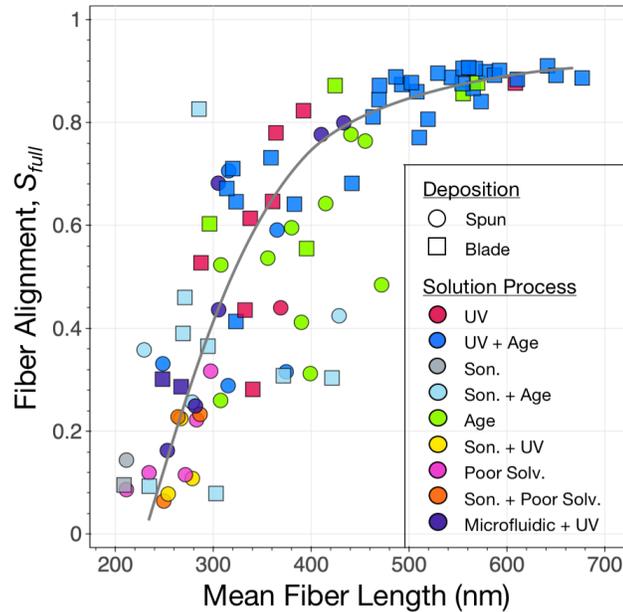


Figure 25 – The relationship between mean fiber length and fiber alignment across the entire image library. The gray line is included to enhance readability.

Since charge carrier mobility in bottom gate, bottom contact transistors is more dependent on the interfacial structure than the top surface, which is more easily captured by AFM, an analysis of each surface is presented in **Figure 26** for the UV, aged and blade coated processing method. Images of the SiO_2 interface were obtained by pressing a PDMS slab onto P3HT films, submerging the entire stack in DI water for 10 minutes, then peeling off the PDMS slab with the P3HT film attached. As tabulated in **Figure 26a**, the top and interfacial surfaces of each sample are correlated, but not equivalent. It has been observed previously that thin films composed purely of P3HT nanofibers have roughly identical-

looking morphologies on the top and interface, but the detailed differences have not been quantified.¹²⁶

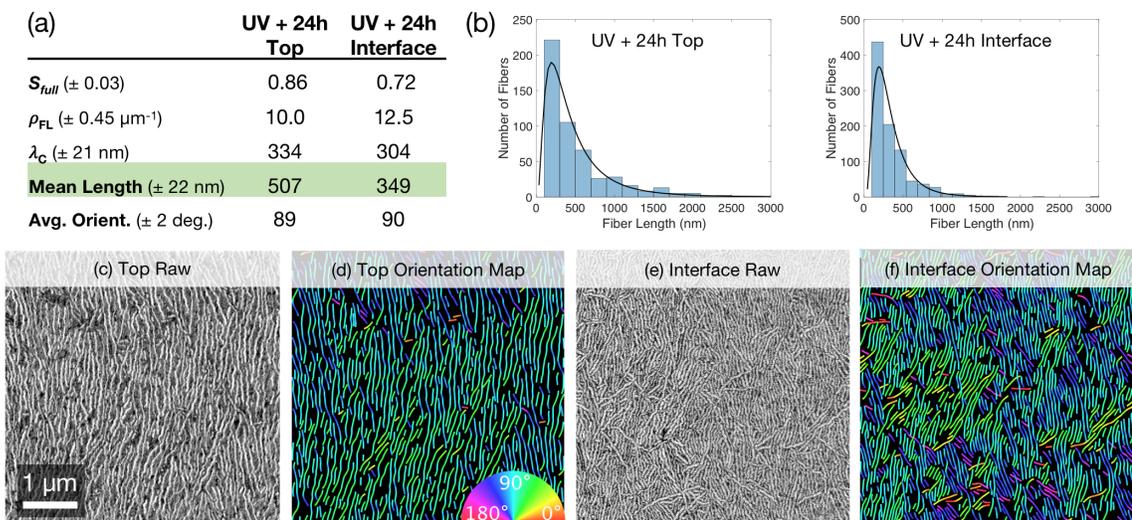


Figure 26 – (a) Image analysis results from a sample that was UV irradiated, then aged 24h and blade coated, measured at both the air at SiO₂ interface by peeling off the film. (b) Length distributions from each image, fit with log-normal probability density functions. (c-f) Raw AFM images (5 μm each) and colored Orientation Maps produced by GTFiber for each sample at each interface.

Shorter fibers tend to segregate to the SiO₂ interface (mean length of 349 vs. 507 nm), as revealed by the analysis in **Figure 26a** and **b**. This was confirmed with three AFM images of both surfaces as well as manual tracings. The difference can be rationalized by considering settling in a cereal box: the smaller items can create a more dense phase, and thus sink to the bottom. In this case, the smaller fibers are likely more mobile as well, facilitating their transport to the interface. Unfortunately, short fibers tend to cause defects in fiber alignment, evidenced by (a) the defects in the aligned structure in **Figure 22j**, (b) the decreased alignment with shorter fibers at the interface in **Figure 26f**, and (c) the general correlation between fiber length and alignment. Further improvement in mobility for pure, regioregular P3HT may be obtained by removing shorter fibers from the

population. Simply aging the solution for longer times is not an option, however, because macro-scale phase separation and gelation begins to occur after 48h.^{137,146} Targeting crystals of a controlled size is a well-known challenge in crystallization, and will be the subject of further investigation.¹⁴⁷

3.4.3 *The Mechanistic Origins of Fiber Curvature*

While short fibers disrupt alignment by creating meso-scale packing defects, an additional factor limiting alignment is the simple fact that P3HT nanofibers are not straight, rigid crystals. We conclude our analysis by investigating potential self-assembly mechanisms that reconcile both the observed length distribution and nano-scale curvature of P3HT nanofibers. In **Figure 27**, cropped sections of 2 μm images (500 nm total width) taken from a sonicated and aged sample (**Figure 27a**) and a UV and aged sample (**Figure 27b**) are presented alongside images in which the extracted fiber contours are colored according to their local curvature. Calculation of curvature is detailed in **Appendix B**. While P3HT nanofibers appear to be relatively rigid in larger 5 μm images, they have significant curvature on a length scale of ~ 100 nm, sometimes making up to 90 degree “turns.” Given the 0.38 nm π - π stacking distance along the fiber backbone, such a turn is energetically unfavorable and likely impossible simply by introducing an angle between

each chain, or “bending” the fiber as it were. Instead, the growth mechanism of P3HT nanofibers offers an alternative explanation.

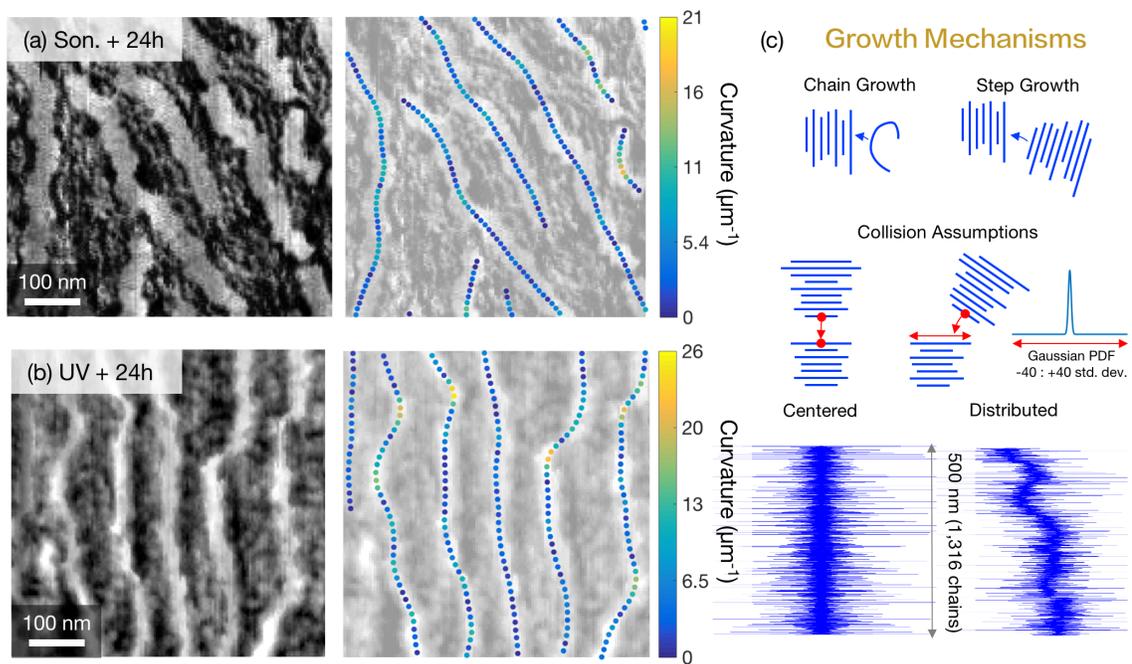


Figure 27 – Analysis of P3HT nanofiber curvature and its mechanistic origin. (a) Cropped section of a 2 μm image taken from the same sonicated, aged and blade coated sample as Figure 26c, and a plot of vectorized fiber contours with vertices colored according to their curvature. (b) Cropped section of a 2 μm image taken from the same UV, aged and blade coated sample as Figure 26e, with the same accompanying curvature heat map. (c) Proposed nanofiber growth mechanism and accompanying simulation results. Growth is assumed to occur via a polycondensation-like route. Collisions between growing segments are assumed to occur at a narrow Gaussian distribution of locations along the growth front. Results of nanofiber growth simulations under both centered and distributed collision assumptions, plotted to scale with chain lengths sampled from a log normal distribution corresponding to an M_n of 40 kD and PDI of 2.25.

As demonstrated by the histograms and fits in **Figure 26b**, P3HT nanofibers follow a log-normal distribution of lengths. The log-normal distribution is typical of a polycondensation-like mechanism (minus the release of water here), allowing both chain growth (the addition of one chain to a stack), and step growth (the combination of stacks), as illustrated in **Figure 27c**.¹³⁵ In P3HT nanofiber assembly, the driving force for growth

is the physical π - π stacking interaction.¹⁴⁸ Perfectly centered growth would be the naïve assumption in this case, with each additional chain or stack overlapping the rest of the fiber as much as possible (left, **Figure 27c**). However, this need not be the case, partially because π - π stacking acts at the repeat unit level, and partially because of the wide variety of chain lengths in a typical polymer molecular weight distribution. Varied chain lengths conceivably present a growth front of varied width. If we assume that collisions between new chains and stacks are not perfectly centered, and instead follow a narrow Gaussian distribution of locations around the center of the fiber's growth front – so narrow that the ends of the fiber are 40 standard deviations from the center – a to-scale simulation of an assembled fiber replicates the observed curvature quite accurately.

In this Monte Carlo simulation, a starting population of chains are sampled from a log-normal distribution of chain lengths ($M_n=40$ kD, PDI=2.25). One-third of these chains are immediately combined into nuclei of ten chains each, reflecting the fact that the aggregate fraction of P3HT chains measured by UV-Vis absorbance multiplies by a factor of ~ 3 during solution-based fiber growth.^{48,55} Further nucleation is not considered because it is a rare event, as evidenced by the fact that nucleation must be induced in the vast majority of processing methods. Each isolated chain and assembled unit of chains is then assigned an integer, and at each step of the simulation, two random integers are chosen to collide and merge. When a collision occurs, the incoming chain or stack is stacked on top of the receiving chain or stack. The center of the unit being added is placed at a location relative to the receiving center sampled from a Gaussian distribution along the width of the growth front. The far ends of the growth front are set at 40 standard deviations (regardless of its width), and collisions beyond the ends are not allowed. Growth proceeds until all

chains are part of one fiber. The resulting structure is presented to-scale in **Figure 27c**, with observably similar curvature features to that of the imaged nanofibers.

3.5 Meso-Scale Simulation of Polymer Packing

The simulation of fiber growth with chain-level resolution leads to another interesting question: why not simulate the packing of polymer chains throughout an entire morphological image? It could yield an indirect approach to the analysis of grain boundaries and inter-grain connectivity. Inter-grain connectivity is frequently cited as a thin film structural feature necessary for high field effect mobility in polymeric transistors, as shown in **Chapter 2**. Discussion of this concept centers around the presence of “tie chains”: polymer chains long enough to be incorporated into multiple crystalline grains, providing both a physical link and a charge transport pathway between grains.³⁶ Indeed, tie chains have been directly observed *via* high-resolution TEM,¹⁴⁹ and indirect evidence for their existence has also been presented by purposefully breaking polymers’ conjugation at regular intervals along their backbone.¹⁵⁰

While the mere presence of tie chains is acknowledged as an important structural feature for charge transport, quantitative metrics describing their contribution to thin film structure have not been developed – a necessary first step in relating inter-grain connectivity to processing and properties. It has been shown that higher molecular weight conjugated polymers, or similarly, mixtures of high and low molecular weight polymers, show improved charge transport in field effect transistors, apparently due to the connectivity imbued by the longest chains in the molecular weight distribution.^{26,52,82,112} Mechanical property simulations indicate that the presence of tie chains between

conjugated polymer crystallites improve their resistance to inelastic deformation.¹⁵¹ However, because of the difficulty of experimentally characterizing tie chains, discussion of their quantitative impact on bulk material properties has been limited to molecular-scale simulations and, in the case of experimental studies, cartoons. Due to these limitations, important questions regarding the influence of tie chains remain unanswered: what is the quantitative relationship between a polymer's molecular weight distribution and the number of tie chains it can form? How do tie chains influence the grain structure of a conjugated polymer thin film and *vice versa*?

3.5.1 Simulation Details

Once all of the fibers in an image have been vectorized, P3HT chains can be placed along the backbone, as illustrated in **Figure 28**. Starting from one end of a fiber, chains are randomly sampled from a log-normal chain length distribution specified by a number average molecular weight, M_n , and a polydispersity, PDI. A chain's length runs perpendicular to the orientation of its parent segment's backbone. Each subsequent chain is placed one π - π stacking distance further along the segment backbone; in this case, a distance of 0.38 nm. Because this simulation is targeted more at understanding long-range structure than individual fiber structure, it is assumed that each chain is centered on the fiber backbone (the so-called "naïve" assumption in **Figure 27**), and it is also assumed that chains lie flat in the plane of the image, extended to their full length, and do not bend or fold. As mentioned above, these assumptions are not realistic. However, the simulation can

provide an *upper bound* on tie chain behavior, through which one can statistically analyze the potential for tie chain formation in a given film.

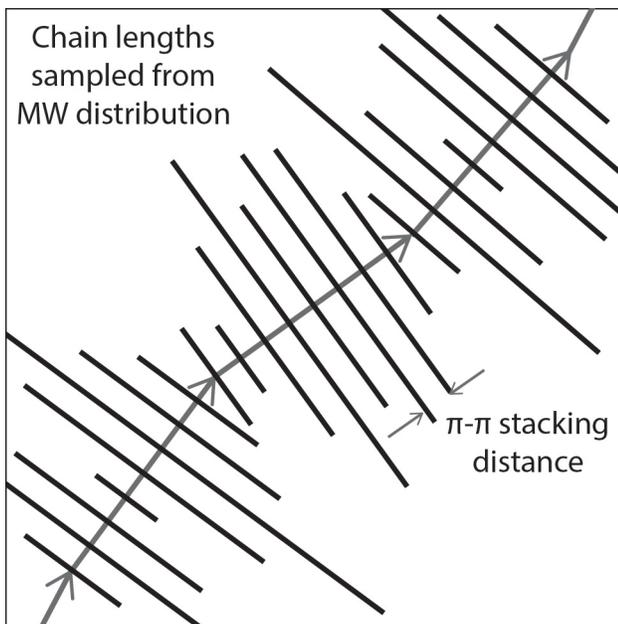


Figure 28 – Simulation of polymer chain packing along a vectorized fiber backbone. Chain lengths are perpendicular to their parent segment’s orientation, and are randomly sampled from a specified molecular weight distribution. Chains are spaced according to the π - π stacking distance of P3HT.

The simulation results are illustrated in **Figure 29**. The Orientation Map of an image, **Figure 29b**, serves as the initialization for the simulation. **Figure 29c** shows a to-scale plot of all 501,331 polymer chains simulated in one realization, using an M_n of 12 kD so that there is visible space between fibers. A zoomed-in windows of the simulation is provided in **Figure 29d**. Only the chains from the crystalline fibers are simulated here – disordered chains in between grains are omitted due to the lack of information on their conformations.

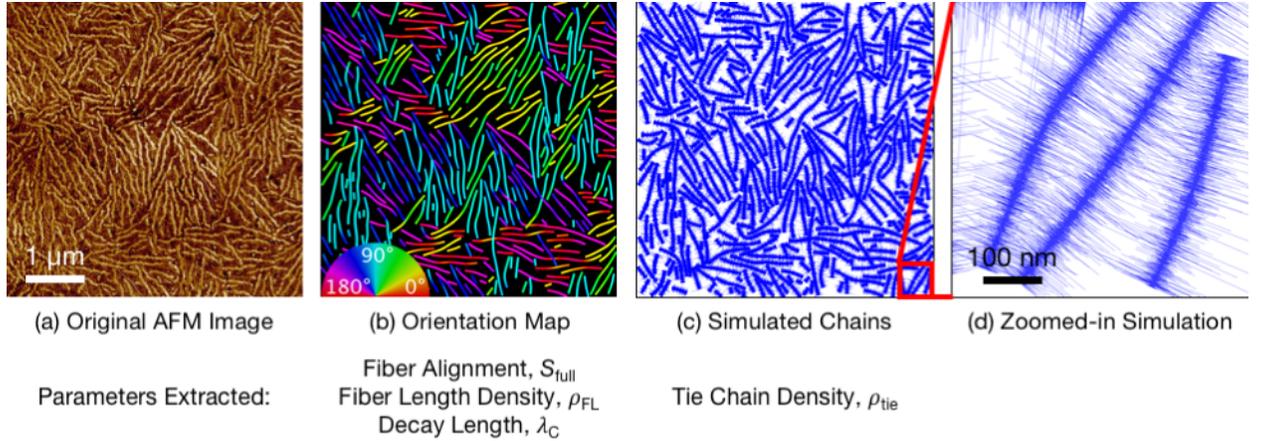


Figure 29 – Visualization of image processing and simulation results. (a) Original 5x5 μm AFM image, phase channel. (b) Orientation Map, in which fibers have been extracted, vectorized, and each segment colored according to its in-plane orientation. (c) Full-scale plot of all simulated polymer chains (501,331) in a 5x5 μm window. (d) Zoomed-in view of (c) with a 500 nm window.

To quantify the presence of tie chains in a simulated microstructure, we introduce the “tie chain density” metric, referred to here as ρ_{tie} , and defined for a single vectorized fiber as:

$$\begin{aligned} \text{Tie Chain Density, } \rho_{tie} (\text{nm}^{-1}) \\ = \frac{\text{Number of tie chains intersecting fiber backbone}}{\text{Length of fiber backbone (nm)}} \end{aligned} \quad (4)$$

It can be seen above that ρ_{tie} quantifies the number of tie chains *per unit length* of a fiber. The process of counting the number of tie chains intersecting the fiber backbone is illustrated in **Figure 30**. First, the main fiber’s chains are removed from consideration to avoid counting self-intersections, although the underlying backbone is still drawn as a dashed black line. Next, every polymer chain in the entire simulated microstructure (~500,000 for a $5 \times 5 \mu\text{m}$ image) is tested for intersection with every segment vector along the main fiber’s backbone using a line segment intersection analysis algorithm optimized

for speed. The coordinates of any intersection points are marked as red circles along the main fiber's backbone in **Figure 30**. Finally, these intersection points are counted and divided by the total length of the fiber in nanometers, yielding the tie chain density, as defined above. This is calculated for every fiber and averaged across the image and across multiple realizations of chain packing.

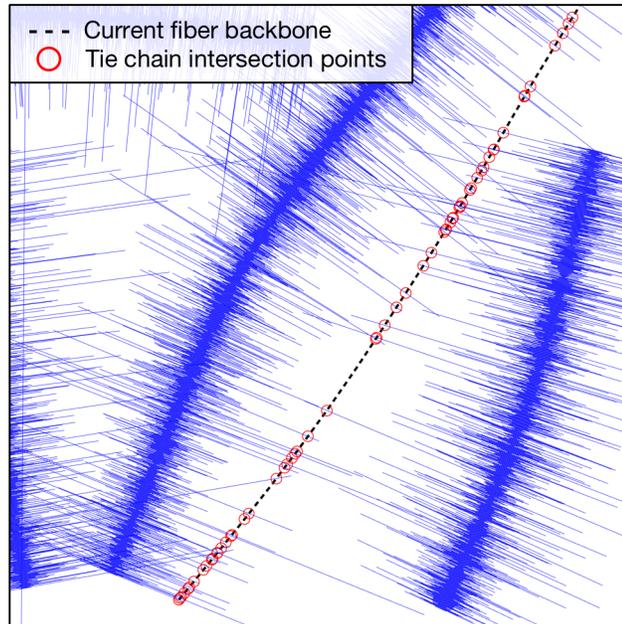


Figure 30 – Quantifying tie chain density for a single fiber. Starting from Figure 29b, the central fiber is made invisible. That fiber's backbone is represented by a black dashed line, while red circles indicate the intersection point along its backbone of any chains from any other fiber in the entire microstructure (notice that even chains from the extreme left fiber are counted).

3.5.2 Molecular Weight, Fiber Length Density, and Tie Chains

Since we have access to AFM images of a wide variety of fibrillar structures from the current study, we can use these as realistic starting points for polymer packing simulations, and subsequently obtain a relationship between the maximum achievable tie chain density and any of the structural order parameters extracted by GTFiber with any simulated molecular weight distribution of our choosing. The relationship between fiber length density and tie chain density is plotted in **Figure 31a & b** for three different assumed molecular weight distributions: $M_n = 10, 25,$ and 40 kD at a PDI of 2.25. It can be seen here that tie chain density increases with increasing fiber length density for all molecular weights. In other words, as fibers get closer to one another, their chains are more likely to overlap, which is not altogether unexpected.

What may be more surprising is the change in the slope of the relationship as molecular weight increases. For 40 kD, not only does tie chain density increase with increasing fiber length density, it increases twice as fast as that of 25 kD and over 10 times as fast as 10 kD, surpassing a value of $1/\text{nm}$ at a fiber length density of $10/\mu\text{m}$. This means that in a typical fiber in a structure with fiber length density of $10/\mu\text{m}$ and M_n of 40 kD, there is the potential for a tie chain to occur every nanometer, or approximately 1 in every 2.5 chains, although it should be stressed again that this is an upper limit due to the assumption of no chain folding.

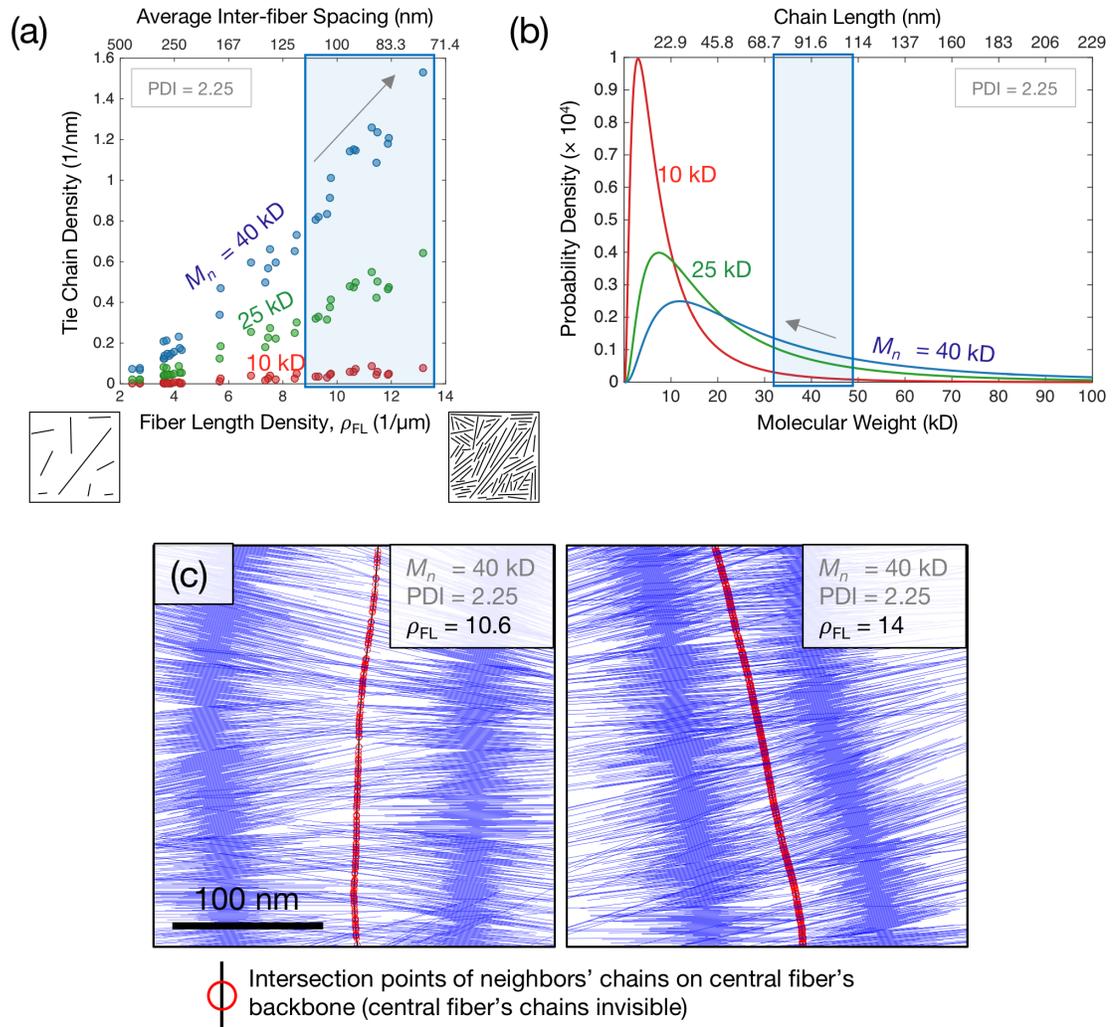


Figure 31 – The potential for fibers to form tie chains given molecular weight distribution and fiber length density. (a) Tie chain density as a function of fiber length density, simulated for three different number-average molecular weights. Blue: $M_n = 40$ kD, green: $M_n = 25$ kD, red: $M_n = 10$ kD. PDI = 2.25. Fiber cartoons are provided on the x-axis as a visual descriptor of fiber length density. (b) Probability density functions of the three molecular weight distributions considered at a PDI of 2.25. Highlighted in light blue is the range of chain lengths corresponding to the range of inter-fiber spacings similarly highlighted in (a).

The reason for the difference in slopes between the different molecular weight distributions is illustrated in **Figure 31c** and **d**. As fibers approach each other, shorter chains in the molecular weight distribution can bridge the gap between fibers. Average inter-fiber spacing is simply the inverse of fiber length density, plotted as a second x-axis

in **Figure 31a**. Between the values of 10 and 14 μm^{-1} , inter-fiber spacing decreases from approximately 115 to 75 nm, highlighted by a blue box. The corresponding chain lengths are highlighted in **Figure 31b**, in which the probability density functions of the three molecular weight distributions under consideration are plotted. In this range of inter-fiber distances, the chains lengths being accessed are more highly represented in the 40 kD distribution than the 25 kD by a factor of nearly 1.5, and the 25 kD is more populous than the 10 kD by nearly an order of magnitude, which matches the observed differences in the slopes of tie chain density *versus* fiber length density. This is illustrated with frames from actual simulations in **Figure 31d**. A single fiber in the center of each frame is shown as a black line, overlaid with red circles where chains from the neighboring two fibers (blue lines) intersect it. The chains from the central fiber are not plotted so that the intersection points are visible. As the fibers are packed more closely, the density of red intersection points visibly increases, confirming the trends quantified in the previous plots.

The density of polymer chains in this depiction may appear visually unrealistic after having seen many studies with cartoons of polymer packing. It is stressed that this is the first time that conjugated polymer packing has been simulated at this length scale, and is arguably the most accurate depiction of the actual underlying structure presented to date. It is anticipated that these results will pair well with past simulations of internal nanofiber structure from coarse-grained MD simulations.¹⁵²

3.5.3 Transferability of Results and Future Work

The analysis and simulation tools presented here represent an important step forward in the quantification of conjugated polymer microstructure. While P3HT is not the state of the art in terms of performance, control of orientation and alignment in any conjugated polymer thin film is recognized as one of the most important factors in determining charge transport, thermal, and mechanical properties. Since the characterization of soft materials is generally quite challenging, and imaging plays a substantial role in structural characterization, any information that can be quantitatively extracted from images is potentially useful. It is our hope that insights from the packing and defects in aligned P3HT thin films will help researchers better understand related polymer systems.

Computer vision tools for materials science are also rapidly advancing as imaging technologies improve and structures become more complex. Many strategies that have been successful in biology and medicine will be applicable to materials systems as well, the present system being one example. An ideal materials imaging ecosystem would include a centralized database of structural imagery, through which one could upload their images as a “search term” and be directed to past studies that produced similar images as well as the code used to analyze those images.¹⁵³ Such an approach would facilitate greater standardization in measurements from images, as well as produce a large image database that would enable the application of more advanced computer vision techniques for image segmentation, such as generative adversarial convolutional neural nets.⁶⁰ This is the subject of the next chapter.

3.6 Conclusions, Chapter 3

In conclusion, we have demonstrated both a powerful method for the analysis of fibrillar structures and the important process-structure-property relationships that fall out of this analysis for the P3HT nanofiber system. Our image analysis workflow is packaged as an open-source standalone application to be used by other researchers, allowing the calculation of their image-scale alignment (S_{full}), fiber length density (ρ_{FL}), and the decay length of orientational order (λ_C), as well as estimation of fiber length and width distributions with reasonable accuracy. At best, the availability of this software will enable the rapid analysis of fibrillar images right out of the box, and at minimum will provide a code base on top of which further image analysis tools can be developed. Additionally, a new interactive data visualization was introduced for structural imagery, enabling intuitive navigation of image libraries with quantitative order parameters.

The results of our analysis extend the already rich knowledge base on P3HT nanofiber growth and deposition to include detailed data on meso-scale packing, alignment, and defects. Microfluidic crystallization was shown to produce narrower fibers with higher fiber length density when deposited, while sonication was shown to produce wider fibers with lower fiber length density and alignment. A strong correlation between fiber length and alignment was identified, as well as evidence that shorter fibers preferentially segregate to the buried dielectric interface, causing packing defects that potentially limit alignment and charge carrier mobility in otherwise highly aligned structures. A simulation of fiber growth was introduced that reconciles both the observed log-normal distribution of fiber lengths and the nano-scale curvature inherent to P3HT nanofibers. Furthermore, a meso-scale Monte Carlo simulation of polymer packing was used to illustrate to-scale

relationships between fiber length density, molecular weight, and the potential for tie chain formation. These results should support the effort to bring about a more quantitative understanding of conjugated polymer self-assembly, structural ordering and charge transport, while providing foundational tools for informatics-enabled materials experimentation and the general class of fibrillar materials.

Chapter 4. Structural Similarity Metrics for Materials Image Repositories

4.1 Introduction

The Materials Genome Initiative (MGI) is a far-reaching effort to halve the development time and cost of new functional materials.⁷⁶ One of the main aims of the MGI is to modernize the management of increasingly large amounts of materials characterization data by providing cloud-based data hosting, version control systems and integration with supercomputing resources. While there are a vast number of materials characterization techniques, imaging, specifically microscopy, arguably stands the most to gain from the construction of cloud-based repositories. Imaging is undergoing rapid technological advancement, with new applications emerging in high-throughput biological analysis,⁷¹ *in situ* monitoring of crystallization,⁶⁷ and *in vivo* evaluation of biomaterials.¹⁵⁴ It is used heavily in organic electronics research, revealing molecular-scale crystal packing motifs,^{131,155} meso-scale morphology and phase separation,¹⁵⁶⁻¹⁵⁸ and is increasingly used for spatially-resolved functional measurements.^{142,159}

The number of microscopy images from materials characterization contained on laboratory hard drives around the world is likely enormous, but most of these datasets are uncatalogued and are therefore not searchable or discoverable. This is unfortunate because recent advances in computer vision such as Google's Cloud Vision API¹⁶⁰ and Facebook's DeepMask¹⁶¹ demonstrate that computers can reliably analyze scenes and identify objects when trained with a large enough (> 1 million images) database of tagged images. If

modern computer vision techniques were applied to materials science, it could revolutionize how structural imagery is searched, discovered, and analyzed.

Image analysis in materials science (and biology, and medicine) is currently approached on a case-by-case basis, and generally involves a segmentation step (identification of pixels belonging to a phase or object of interest) and the extraction of physically relevant quantities from the segmented image. Even when the microstructural feature of interest is something simple, such as the volume fraction of a particular phase, quantitative feature extraction requires at least some degree of algorithmic tailoring – and this is assuming that the interested researcher can find an existing code library that deals with their specific task. Oftentimes, due to the idiosyncrasies of an imaging technique or material system, algorithms must be developed from scratch.¹⁶²

In a more ideal materials imaging ecosystem, a researcher would use one of their own images as a query to a central image database, which would return the most similar images, the publications they appear in, and the code library used for their analysis. This would be superior to traditional text-based searches because the performance of image analysis algorithms is driven more by the appearance of the image being fed to them than by the description of the feature they purport to extract or analyze. For example, the analysis of *C. Elegans* worms in biology shares many similarities with the analysis of fibers in materials, but a materials researcher would never think to search for “worms”.^{71,114} Given that the biological and medical communities are somewhat more advanced than the materials community in their approach to image analysis, this problem is likely widespread.^{66,100,163}

To realize this vision for materials image analysis infrastructure, robust descriptors of image similarity must be developed. A descriptor, in this case, refers to an n -dimensional feature vector calculated for an image, ideally possessing invariance to affine transformations such as translation, scaling and rotation. Image similarity can then be defined as the Euclidean distance between two images' feature vectors, usually after simplifying the features with a dimensionality reduction algorithm. Such descriptors must also be calculable on a raw image, rather than a segmented image, because of the aforementioned difficulty of designing segmentation algorithms for different material classes.

Recent studies have begun to address this problem: DeCost *et al.* demonstrated the use of SIFT (Scale-Invariant Feature Transform) as a robust feature space for materials images and later extended their work to include features from intermediate layers of convolutional neural networks (CNNs), using both Bag of Words and Vector of Locally Aggregated Descriptors (VLAD) for feature encoding.^{153,164} Chowdhury *et al.* and Lubbers *et al.* investigated the use of a wide variety of features, including a CNN layer, in combination with a wide variety of feature selection methods.^{165,166} In these studies, performance was evaluated on a microstructure classification task using datasets containing up to 1,000 labeled images, mostly from metallurgical systems. The best performance in both studies was obtained with CNN features classified using a support vector machine with a linear kernel. Chowdhury did not see a significant improvement in performance through feature encoding, and DeCost did not consider any cases *without* feature encoding. In either case, both were able to obtain ~95% accuracy in some classification tasks, while still experiencing difficulty with others. In general, it seems that

the combination of feature generation with SIFT or CNN features, dimensionality reduction, and classification with a SVM is a robust strategy for the classification and comparison of images of materials structures.

In this study, we aim to extend these previous results to a larger, more diverse image database: the set of every image contained on our group's atomic force microscope. Over 16,000 images were available – we selected 3,000 for tagging and analysis. This dataset contains images from three different imaging modes and ten different users, likely encompassing hundreds of materials experiments and dozens of unique structural classes. It also contains many noisy, artifact-laden, or otherwise discarded imaging results that traditionally never see publication, but here illustrate the additional need for automated artifact recognition in imaging systems. The above analysis strategies are evaluated for two classification tasks: the automated identification of images containing fibers, and the automated identification of images containing artifacts. The former is intended as a demonstration of automated dispatching to tailored image analysis programs (in this case, GTFiber¹¹⁴), while the latter is intended to demonstrate a path toward automated acquisition of microscopic imagery.

4.2 Methods

4.2.1 Database Assembly, Cleaning, and Tagging

Raw AFM images are stored in a file format containing multiple channels of image data as well as metadata about the acquisition process such as scan size, resolution, and instrument settings. The “image” data is technically a 2-D spatial map of the electrical response of a scanning probe as it traverses a sample, like a record player. Each channel of

an AFM image file represents a different aspect of the tip's response to its drive signal as influenced by the local physical and electrostatic environment of the sample. For example, the "intensity" of the Height channel indicates how the tip was shifted vertically at each pixel to maintain the specified separation from the sample. The Amplitude Error channel indicates the deflection of the tip at each pixel, which corresponds to the slope of the sample surface. The Phase channel indicates the phase shift of the tip's response relative to its drive signal (typically a sine wave in the hundreds of kHz, the resonant frequency of the tip material), and is known to qualitatively capture differences in hardness, crystallinity, and chemical composition. While these channels are not technically *images* in the strictest sense of the word, mapping the acquired matrices to an image through a color map generally yields images that have a visually digestible appearance and correspond to recognizable structural features such as particles, grains, fibers, and other objects.

The 3,000 images under consideration are thus generated from 1,000 original raw files, each with three channels. The images range in acquisition size from 0.5 – 30 μm , and in resolution from 256 – 1024 pixels/line. Each image is traditionally pre-processed by a flattening operation. This involves the subtraction of a third-order polynomial fit to each row of pixels from itself, which removes long-range variations in tip response. To enable the extraction of consistently-sized feature vectors across the entire database, all images were resized to 512×512 . While this removes some of the detail contained in higher resolution 1024×1024 images, only 30 images were taken at this resolution. Upsampling of the 256×256 images is not expected to introduce previously unseen features, especially given that most feature extractors perform scale-invariant operations on the image gradient.

Images were labeled by hand by scrolling through image previews and entering class labels in Excel, but further handling of the image database was performed using the Pandas package in Python. One set of labels were generated for whether an image contained fibers and could be meaningfully processed using GTFiber. The other set of labels contained information on the various types of artifacts present in AFM images, as illustrated in **Figure 32**. Automation of AFM to reduce image artifacts is an area of consistent research, and is usually approached from a controls perspective targeting the instrument's electronic systems, rather than from a computer vision perspective.¹⁶⁷⁻¹⁶⁹ Seven categories of artifacts were observed in our labeling process: **(1)** horizontal and vertical line defects, characterized by a row or consecutive rows of pixels visibly disrupting the surrounding structure; **(2)** horizontal and vertical gradients in image quality (sharpness), indicative of tip degradation; **(3-5)** short-range horizontal artifacts of three different types: streaking, periodic noise, and short "bars" of dark pixels; **(6)** long-range sinusoidal waves in image intensity, usually caused by interference between the laser's reflection off the tip and its reflection off of any more reflective parts of a sample, such as metal electrodes; and **(7)** homogeneous, static-like noise, possibly due to improperly tuned gain or drive amplitude. Surprisingly, images containing noise or artifacts of some kind represented 38% of the dataset – although the severity was highly variable and in many cases may not have precluded the image from being presentable. While classification of individual noise classes would ultimately be desirable, we limit classification experiments here to simply deciding whether or not an image contains an artifact, because not all classes of noise are well-represented in the dataset.

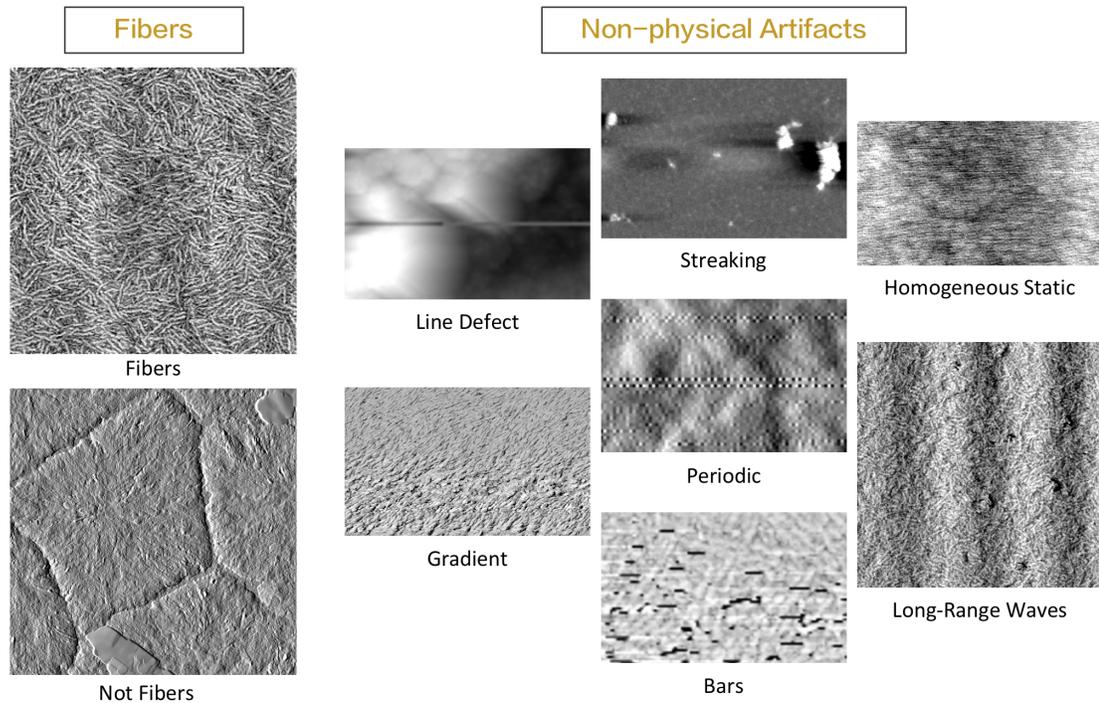


Figure 32 – Structural classes and artifact classes labeled in the dataset with examples.

4.2.2 Feature Generation: Spatial Statistics

A discussion of feature generation for materials microstructures would be incomplete without first describing 2-point spatial correlations, illustrated in **Figure 33**. Spatial statistics enable the embedding of spatially correlated data in a translation-invariant feature matrix, and have found use in geographical information systems as well as in the statistical representation of microstructure.^{65,95,170} They are computationally efficient, taking advantage of the Fast Fourier Transform algorithm. To use spatial statistics, distinct microstructural states must first be identified and segmented. In **Figure 33**, there are only two states: fiber or not fiber. The heat map at right indicates the probability that a vector with coordinates (t_x, t_y) will have both of its ends contained in a fibrillar pixel when placed

randomly in the microstructure. This is known as an autocorrelation function – a cross-correlation would quantify the probability that the same vector had its ends in *different* phases. When only two phases are present, the autocorrelation is the only linearly independent feature.

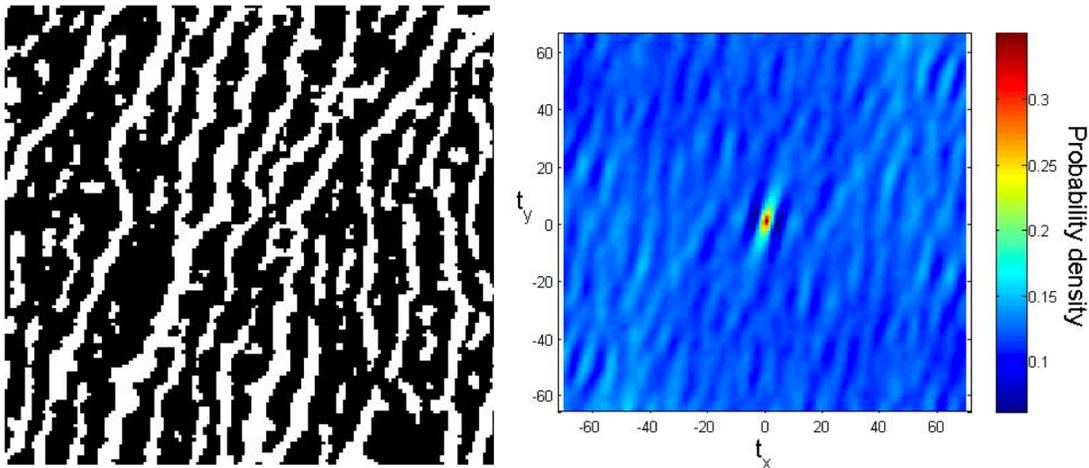


Figure 33 – Illustration of 2-point autocorrelation matrix (right) for the segmented fibrillar microstructure at left. In the segmented microstructure, a white pixel indicates a fibrillar phase. In the autocorrelation heat map, the color indicates the probability that a vector with coordinates (t_x, t_y) will have both of its ends contained in a fibrillar phase when placed randomly in the microstructure.

Note that the origin of the autocorrelation is in the center, which corresponds to the zero vector, or the probability that any single point is in a fibrillar phase – *i.e.* the volume fraction. The top end of the scale bar indicates that the volume fraction of fibers in this microstructure is 0.35. Probabilities decrease quickly to the left and the right: if one starts in a fiber and moves horizontally, they will quickly end up in a background region. Probabilities decrease more slowly if one looks at an angle of $\approx 80^\circ$, corresponding to the average orientation of the fibers in the example microstructure. This feature space is advantageous because it is fundamentally related to physical laws governing material

mechanics. However, it is severely limited in two aspects: (1) it is not rotationally invariant, and (2) it requires an accurate segmentation. Rotational invariance could be dealt with by using a polar autocorrelation function: however, fast and direct *inverse* polar FT algorithms have only recently been developed.^{171,172} As an aside, the implementation of polar FTs could be potentially revolutionary for computer vision and perhaps even molecular simulations to the same extent that the original FFT algorithm revolutionized signal processing many decades ago.

4.2.3 SIFT Features

Spatial statistics serve as an intuitive warm-up to more complex image representations such as SIFT and CNNs. SIFT features are illustrated in **Figure 34**, borrowing from DeCost *et al.*¹⁵³ For each image, a set of keypoints are identified, illustrated as circles of varying size with radial lines indicating their orientation. Keypoints are identified by subtracting Gaussian-blurred versions of the image from itself at increasing Gaussian widths and evaluating gradient magnitudes. For each identified keypoint, a 4×4 grid is overlaid on the image, and the image gradient magnitude is calculated in 8 different directions for each cell of the grid, resulting in a 128-dimensional feature for each keypoint. Each image is thus described by a $k \times 128$ array of features, where k is the number of keypoints identified. Further information on SIFT features can be found in its original publication.¹⁷³

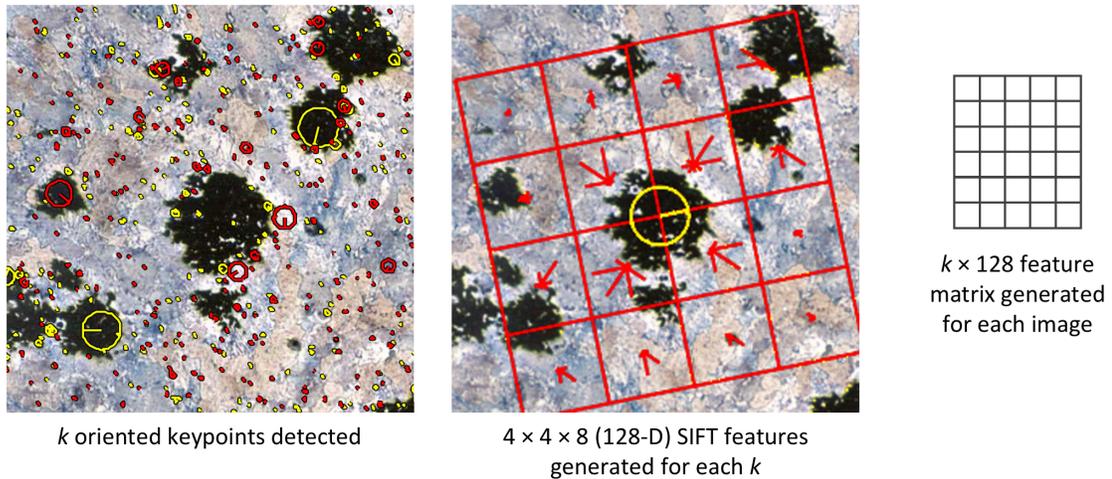


Figure 34 – SIFT Feature Generation. Keypoints are identified by circles with radius corresponding to their scale and radial lines indicating their orientation. For each keypoint, an oriented and scaled 4×4 grid of gradients in 8 directions is calculated, yielding a 128-dimensional feature for each keypoint. Adapted from Ref. 153.

4.2.4 VLAD Encoding

An immediately obvious problem with SIFT features is that they generate a different number of features per image. This is solved by performing a k-means clustering of all of the identified features across the entire image database, which generates a dictionary of image features. In a Bag of Words representation, an image is described by the histogram of occurrences of each visual word (cluster centroid) in the dictionary throughout the image.¹⁵³ However, features that are close to cluster boundaries are not well represented. The Vector of Locally Aggregated Descriptors (VLAD) approach mitigates this by representing each cluster's contribution to an image through the normalized, aggregated difference between the features assigned to that cluster and its centroid. A VLAD representation of an image thus contains a unit vector for every cluster centroid in the visual dictionary that points from the centroid to the locus of features (from that image)

assigned to the centroid. Here, we use 100 feature clusters, yielding a 12,800-dimensional VLAD vector for each image. VLAD has been shown to outperform Bag of Words representations as well as other encoding methods, and more details can be found in the provided references.^{174,175}

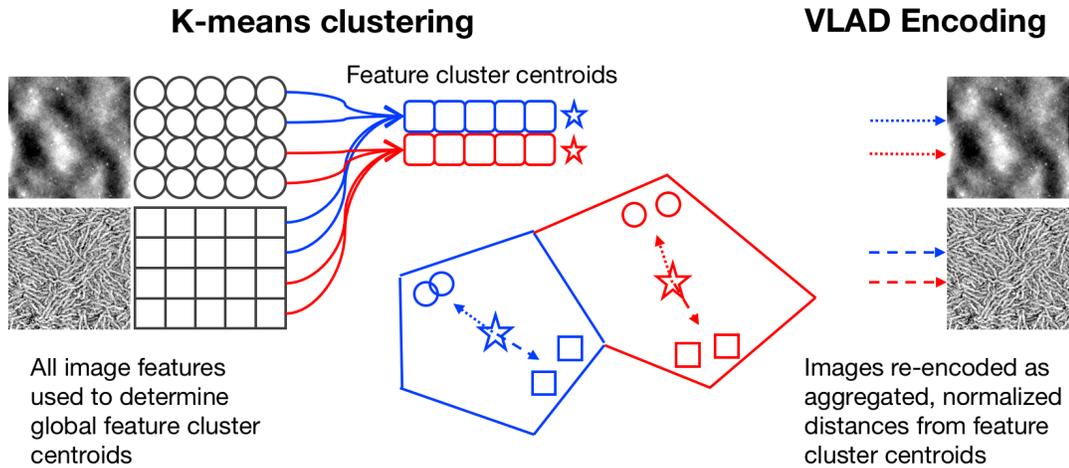


Figure 35 – VLAD encoding after k-means clustering of all identified image features. Colors represent distinct feature clusters.

4.2.5 Convolutional Neural Net Representations

Research on CNNs for computer vision applications has skyrocketed in recent years. A CNN, illustrated in **Figure 36**, is sequence of convolutional filter banks connected by artificial neurons. The filter and neuron weights are trained for object recognition and localization on a dataset of millions of images, usually of natural scenes containing everyday objects.⁶⁰ While not explicitly rotationally invariant, due to the extreme variety of these training sets, the trained neural net must inherently account for such transformations. While designing and training a CNN from data is an extremely challenging task that can take weeks to months, it has been shown that pre-trained CNNs

can perform admirably on datasets that fall entirely out of their training set, referred to as “transfer learning.”¹⁷⁶ Here, we use the pre-trained weights of the VGG16 CNN architecture trained on the ImageNet ILSVRC-2014 dataset.¹⁷⁷ While the full network ultimately reduces down to 1,000 object category labels, for transfer learning it is generally best to use the mid-level features as an image descriptor, then train a classifier on that feature space. Others have shown that for microstructural imagery, mid-level convolution layers are the most effective, so we have selected the 3rd convolution layer from the 4th block of the VGG16 model as our feature set.

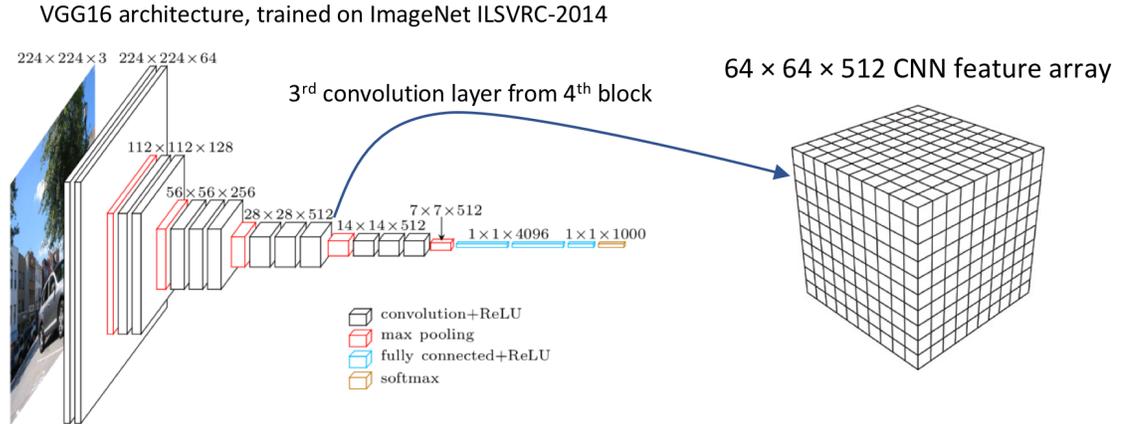


Figure 36 – Convolutional neural net architecture and feature representation.

Each block of a CNN is reduced in size by a factor of 4 through a max pooling layer – since VGG16 was trained on images with 224×224 resolution, the 4th block has channels of size 28×28 . Since our images are 512×512 , the 4th block features have 64×64 512-channel features. We run these features through a 100-cluster VLAD encoding, similar to the SIFT features, resulting in features with $512 \times 100 = 51,200$ dimensions.

4.2.6 Classification, Dimensionality Reduction, and Reverse Image Search

The full algorithm stack used for this study is outlined in **Figure 37**. First, the feature generation and encoding methods are applied to the image database. Classification is then performed using a Support Vector Machine (SVM) with a linear kernel, margin parameter $C=1$, and L2-normalized feature vectors. Classification performance is evaluated by performing 10 runs of 10-fold cross validation on randomly chosen balanced slices of the dataset containing 200 images per class. Standard deviation of classification accuracy is computed across the 100 validation sets that were tested. This cross validation procedure ensures that over-fitting is not occurring, a legitimate concern given the high dimensionality of the feature vectors under consideration.

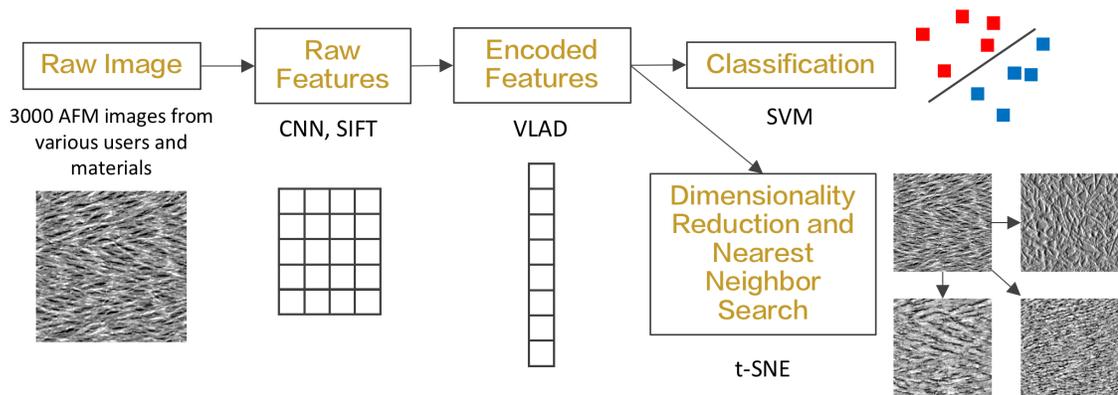


Figure 37 – Outline of feature extraction, classification and nearest neighbor search.

Another way to verify that the features are meaningful is through the next step of our workflow: dimensionality reduction and nearest neighbor search. For this, we use t-Distributed Stochastic Neighbor Embedding (t-SNE), a dimensionality reduction technique that was developed for visualizing high-dimensional datasets in two dimensions.¹⁷⁸ Its objective function is tailored to preserve local similarity, but does not heavily penalize

neighboring mapped points whose high-dimensional Euclidean distance was large. So, images that were close to each other in high-D space are guaranteed to be close to each other in 2-D, but images that were far from each other in high-D could end up close to each other in 2-D.

Using this reduced representation, we can define nearest neighbors as those with the minimal Euclidean distance from a query image. This type of search functionality could be useful for many applications: researchers who have found a unique-looking structure could find other work, perhaps in completely different fields, that obtained the same structure. Or, researchers could search for images similar to their own to identify papers with more detailed image analysis protocols for their specific features. Interestingly, this concept of “reverse image search” can be extended to include almost any kind of data. Mass spectrometry utilizes this to match a collected mass spectrum to an existing library of known mass spectra for various compounds. If all research data were compiled in a central database, any image, spectrum, or spreadsheet could be used to link researchers together based simply on the similarity of their raw data. This could lead to a revolutionary shift in how research is performed.

4.3 Results and Discussion

4.3.1 Classification Performance

The results of SVM classification cross validation are presented in **Table 2**. Using 100 VLAD clusters, SIFT and CNN features show remarkably similar performance on both the artifact and fiber classification tasks – within their respective margins of error. Increasing the images per class (N Per Class) did not have a significant effect on

classification accuracy. Increasing the number of VLAD features is certainly an avenue for exploration – Chowdhury *et al.* tested a wide variety of feature encoding methods on CNN features for microstructure classification, including principal components and Fisher vectors, and found that encoding actually does not significantly enhance classification accuracy versus unencoded features. While encoding is necessary for SIFT features because of their variable length, it is technically not required for CNN features when images are of equivalent size, although the current dimensionality of the raw CNN features is $64 \times 64 \times 512 = 2,097,152$, which seems unnecessarily high.

Table 2 – Classification performance on artifacts and fibers with SIFT and CNN features.

LABELS	FEATURE	ENCODING	SVM KERNEL	N PER CLASS	ACCURACY	STD. DEV.
ARTIFACTS	SIFT	VLAD100	linear	200	74.9%	± 5.9%
ARTIFACTS	CNN	VLAD100	linear	200	78.8%	± 6.2%
FIBERS	SIFT	VLAD100	linear	200	95.5%	± 3.2%
FIBERS	CNN	VLAD100	linear	200	96.8%	± 2.7%

Classification accuracy is much higher for the fiber classification task than the artifact classification task. Fiber classification is over 95% accurate for both feature sets, with low variance. This matches the best performance from previous studies of microstructure classification and further validates the generalizability of this method. Since our database is larger than those that have been studied before and has good representation of this particular structural class, this was expected.

Artifact classification, on the contrary, is only around 75% accurate during validation. There are a couple factors that may contribute to the observed drop in performance. First, the types of artifacts that were labeled vary substantially in both scale

and severity. Long-range horizontal waves likely have a very different CNN representation than that of single-line defects or static. Training accuracies were generally in the range of 95%, however, indicating that the feature space can be discriminative. It is possible that increasing the size of the training set would further improve performance by providing more examples of the rarer artifact classes. Another complication with CNN features is that the VGG16 net was trained on natural, artifact-free images, and thus does not have weights that would be expected to discriminate among artifacts. It may be the case that a CNN needs to be completely re-trained for effective artifact classification. On the other hand, SIFT feature extraction relies heavily on image gradients and keypoints. Some artifacts are characterized by the absence of a gradient, or have extremely complex local gradient patterns that may get washed out by the gradient calculation kernels used by SIFT. Artifact classification certainly warrants further study, as it is a problem that frequently goes unreported.

4.3.2 2-D Embedding and Library Visualization

To visualize the feature embedding of the image library, we mapped the VLAD100-encoded SIFT features to a 2-D representation using t-SNE, pursuant to previous reports. The result is plotted with image thumbnails in **Figure 38**. Each image has been labeled with a blue border if it contains fibers, and a red border otherwise. The database seems to separate into three distinct regions, with the furthest right holding most of the fiber images. These three regions are not the three AFM channels, as may have been expected: attempting to train an SVM to classify the different image channels yields poor results, indicating that they are not easily distinguished. Closer inspection reveals that the bottom-

left region contains mostly smooth surfaces, free of strong edges, while the top region contains images with higher contrast, indicating either distinct phases or rougher surfaces. The region on the right contains most of the images labeled as fibers, but also contains images with ridge-like features as well as oriented artifacts. In the zoomed-in view at right, we can see that images with fibers cluster very closely, and that one of the images that was not manually labeled as having fibers still contains some type of oriented anisotropic objects that could be construed as fibers. This highlights the limitations of text-based, experimentalist-provided class labels when describing microstructure, and further emphasizes that the similarity metric is ultimately what matters.

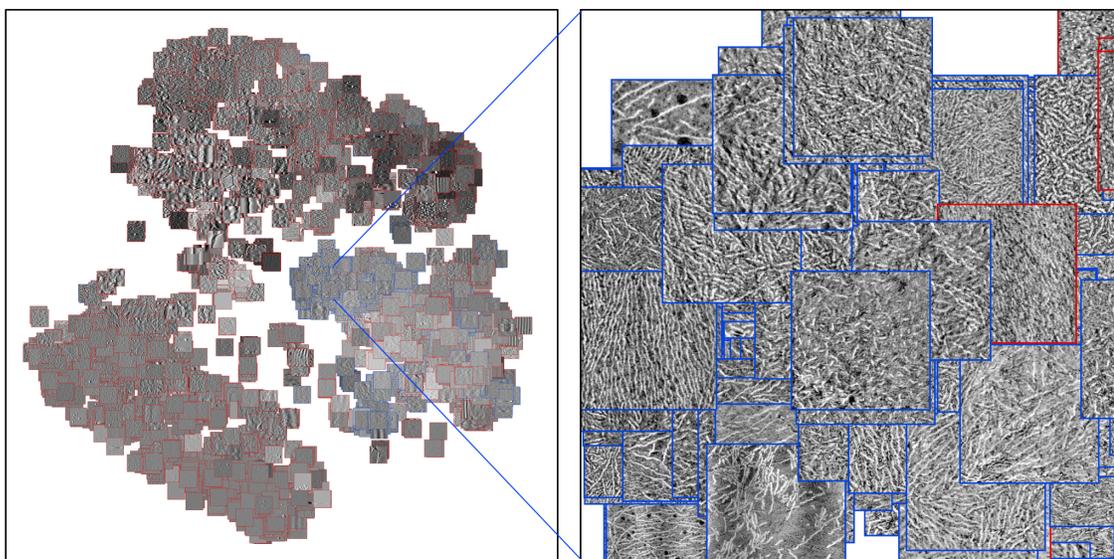


Figure 38 – 2-D embedding of the image database using t-SNE. (Left) Full dataset, (Right) Zoomed-in view of the neighborhood containing most of the fiber images. Blue outlines indicate a label of “fiber” while red indicate “not fiber”.

It should be noted that the intensity scales in these images were selected by the AFM users at the time of image acquisition and have not been normalized or standardized in any way. Most users are primarily interested in a single image channel and attempt to

scale that channel's intensity to increase its contrast. Other channels are scaled automatically by the acquisition software. At least for the Height channel, it may make sense to standardize the intensity scale (which should be possible when pulling images from the raw files) so that all images share a common scale. Then again, the range of height values is likely quite large, and re-scaling them across the whole database may make many images appear essentially featureless. Furthermore, the height profile can be influenced by objects *below* the surface being measured. In the present study, we are mostly concerned with the performance of the image similarity metrics, but these practical issues are important to consider as larger databases come online. One would likely pre-filter the metadata of the image database to ensure results are on a similar length and height scale before proceeding to structural evaluation.

4.3.3 *Reverse Image Search*

To demonstrate the usefulness of structural similarity metrics, we provide examples of reverse image search results in **Figure 39** for four representative query images with highly varied features. The results shown are the first three images with the lowest Euclidean distance from the query image in t-SNE space. In the first row, we consider a structure that could best be described as “terraced,” again raising the issue of describing structures with words. The first two results return other images that appear to be the same material system. The third result is from a different user and is therefore from a different material system despite its similar appearance. In a real-world application of reverse image search, one would exclude their own data from consideration, but here, being returned data from the same system is simply an indication that the similarity metrics are performing

well. In the second row, the query image is a structure of networked globules and returns other networked globules. The third row consists mainly of short fibers with no preferential orientation, and the fourth row contains what appear to be porous or dendritic structures. The third and fourth row of results indicate that the image search algorithm is robust to noise, as some results contain long-range waves or large particulate artifacts that are not present in the query image. From a search and discovery standpoint, this is good, but it is also an indication of why the artifact classification problem is difficult.

Implementation of reverse image search as a service to researchers runs into complexity *vis-à-vis* intellectual property concerns. Researchers must release the copyright to their data and images for any website to display that data. In most cases, these copyrights may actually be held by the institution employing the researcher, rather than the researcher themselves. A potential workaround would be to provide search results only from *published work*. That is, scrape every online scientific journal to download their images, compute the representations of those images with SIFT or CNN features, then direct a researcher to the DOI of the paper containing each result image, rather than showing the image itself.

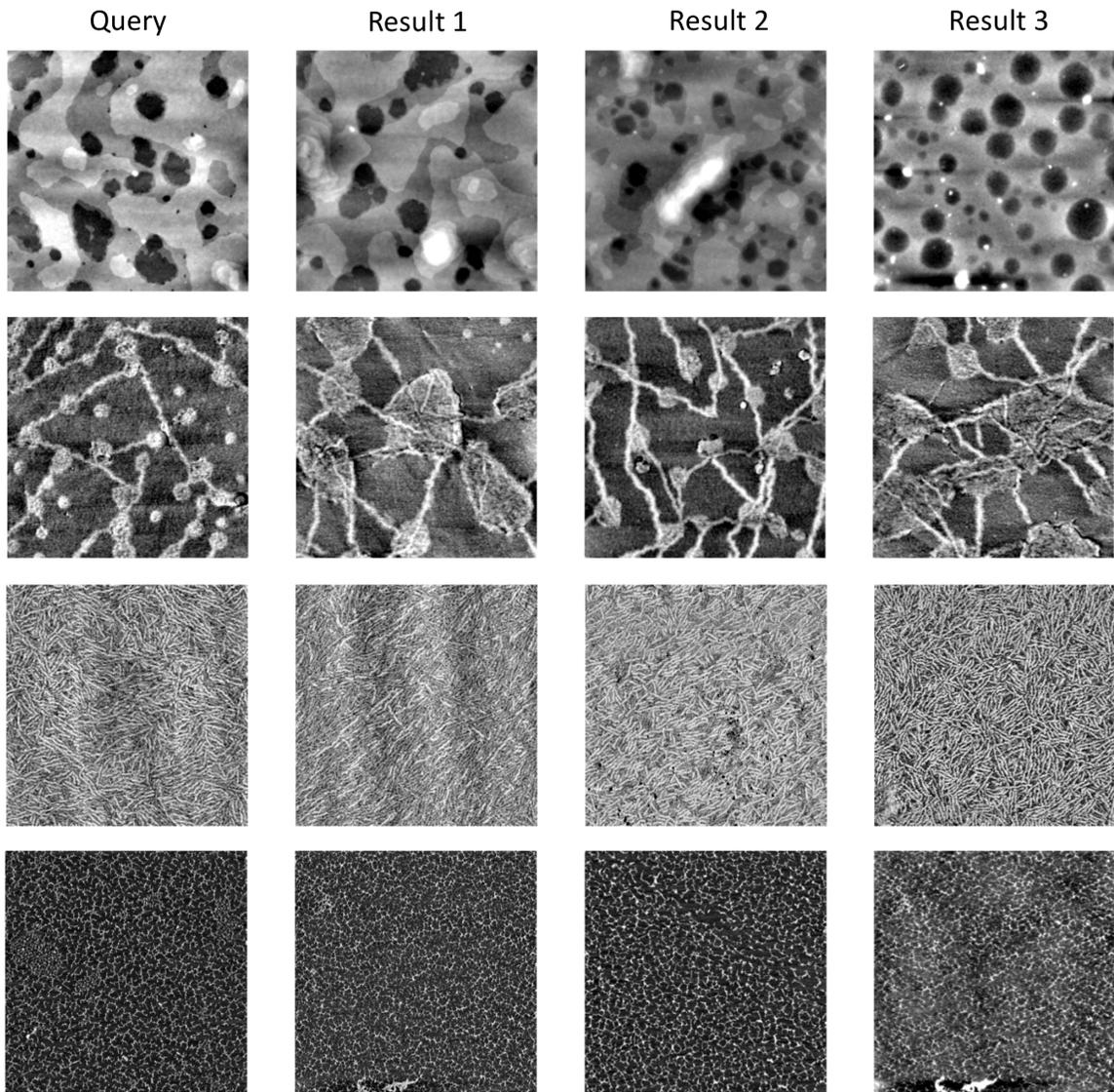


Figure 39 – Performance of reverse image search evaluated on a variety of microstructures, displaying first three results sorted by shortest Euclidean distance in VLAD100-encoded, t-SNE-reduced SIFT space.

4.3.4 Automated Image Analysis Dispatching

In **Figure 40**, we illustrate how an automated structure classifier could be used to dispatch image search results to lower-level image analysis codes and software. GTFiber is an open source software package developed by our group for the automated extraction and analysis of densely packed fibrillar structures, generating both length and width distributions as well as several orientational order parameters. Although it is designed specifically to analyze fibers, it is also somewhat limited in the range of fibrillar images that it can reasonably analyze – and this is true of most heuristic or model-based image analysis pipelines.^{66,71,73} They are designed for a specific dataset and may be effective on a range of other data, but generally have limitations when transferred to foreign material systems or imaging technologies. A central imaging database with reverse image search functionality and robust similarity metrics could therefore help direct researchers to appropriate low-level image analysis code and software.

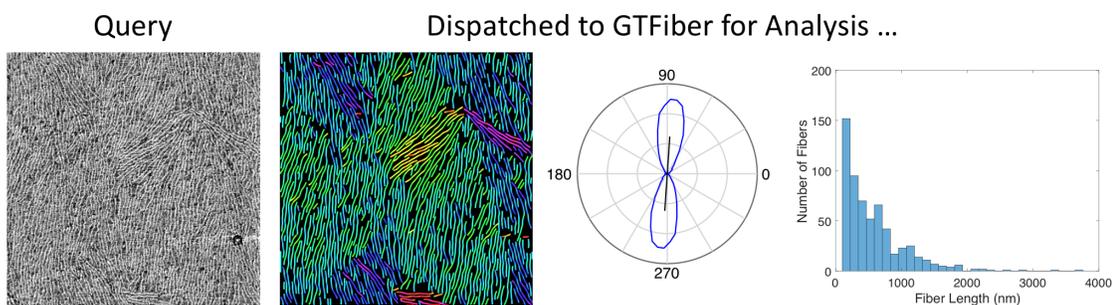


Figure 40 – Example of dispatched image analysis search result: a query image is uploaded, classified as containing fibers that can be analyzed by GTFiber, and is dispatched for further analysis, including the extraction of orientation and fiber length distributions.

The concept of using raw data as a search query to direct researchers to appropriate analytical code can be generalized to nearly any type of data, *provided that raw scientific*

data is publicly available online. Beyond promoting transparency and reproducibility, this is the single biggest argument for an “Open Science” ecosystem.⁷⁷ Any absorbance spectrum, X-ray diffraction pattern, calorimetry curve, gas chromatogram, or even Excel spreadsheet could be uploaded to the central research cloud and return the most appropriate analytical code or protocol based simply upon data similarity, potentially saving thousands of hours of labor spent re-writing models and analysis codes. If we take this thought experiment even further, raw data could be automatically analyzed and converted into a presentable figure with a standardized format for that data source. From there, it is only another step to the automated generation of research publications from raw data. While the storytelling aspect of paper writing is an important facet of research, the implementation of at least some of this functionality could save researchers a significant amount of time and effort.

4.4 Conclusions, Chapter 4

We have demonstrated the application of well-studied computer vision algorithms to the task of automatically classifying AFM images according to their structural class as well as the presence of artifacts. VLAD-encoded SIFT and CNN features were both shown to identify fibers in images with >95% accuracy, while artifacts were identified with 75% accuracy. Additionally, reverse image search functionality was demonstrated using dimensionality reduction and Euclidean distance similarity metrics, with a wide variety of microstructures returning intuitively similar results, robust to the presence of image artifacts. This is the first time that an AFM image database of this scale has been subjected to such an analytical treatment, spanning many different experimental systems and instrument users, as well as a range of imaging modes. This study was intended to demonstrate the value of centralized, open image repositories for materials science, which could be used to facilitate the discovery of low-level image analysis codes and software as well as research publications containing structures matching those present in an image-based search query. Ultimately, such an ecosystem would enable more efficient, informatics-driven data analysis as well as increased connections between researchers in disparate fields.

Chapter 5. Conclusions and Future Work

5.1 Conclusions

The main topics addressed in the methods and results presented here fall into three categories: process-structure-property relationships in P3HT-based OFETs, successful computational strategies for image analysis of fibrillar materials, and informatics methodologies for handling big, open-access materials data. In the case of process-structure-property relationships, many relationships were discovered through collaboration with other experimentalists and through follow-up experiments to fill out previously explored design spaces.

5.1.1 Process-Structure-Property Relationships in P3HT-based OFETs

The Structure-Property Relationship for P3HT

Several molecular and bulk relationships were previously known: a planarized polymer backbone enhances mobility through increased charge delocalization, higher molecular weight polymers are necessary for inter-grain connectivity, and an edge-on bulk grain orientation is necessary to ensure that both polymer backbones and π - π stacks are contained in the quasi-2-D plane relevant to charge transport.

Here, these guidelines were refined through a quantitative analysis of the meso-scale grain structure of P3HT thin films. Of all the fibrillar structural metrics considered, fiber alignment was most strongly correlated with charge carrier mobility when overall fiber orientation was nominally perpendicular to charge transport. This indicated that the

function of P3HT nanofibers was two-fold: (1) to planarize the backbones of as many P3HT chains as possible, and (2) to provide charges access to tie chains that can carry them through the next grain boundary on their way to the opposite electrode. Alignment is correlated with mobility because higher alignment guarantees lower angles between neighboring fibers, which allows tie chain molecules to be more planarized as well.

Factors Limiting Alignment

Despite achieving thin films with $S_{full} > 0.9$, perfect meso-scale fiber alignment is still limited by defects. Fibers formed in solution tend to follow a log-normal length distribution, indicative of a polycondensation-like growth mechanism. Since π - π stacking is non-specific and does not fix the registry of neighboring polymer chains, small shifts in stacking location during step- and chain-growth lead to nano-scale kinks that prevent P3HT from forming perfectly straight fibers. In addition, the shorter fibers in the population tend to segregate to the buried dielectric interface and interrupt packing, which causes alignment at the interface to be slightly lower than at the top surface.

Process-Structure Relationships

Fibers can be nucleated through sonication, UV irradiation, a brief residence time at low temperature, or stochastically over time. Further growth can be induced by aging, poor solvent addition, or application of UV irradiation in a microfluidic flow environment. However, these nucleation and growth methods do not follow equivalent pathways. Sonication leads to the formation of wider fibers that grow more slowly and thus do not align as well. Crystallization in a microfluidic environment leads to significantly higher fiber length density, but is more difficult to accurately control. Poor solvent addition leads

to rapid, uncontrolled growth that limits the formation of tie chains. UV irradiation combined with aging from 1–2 days yields optimal nucleation and growth conditions for long fibers that align well when blade coated. A more controlled experimental system for microfluidic crystallization may also yield similar results.

In thin film deposition, spin coating was shown to induce partial radial alignment of nanofibers with a slight rotational bias, indicating that film drying occurs so quickly that the transient rotational acceleration of the substrate influences the resulting thin film microstructure. Blade coating was shown to induce linear alignment of fibers with mean length greater than ~ 500 nm at a blade velocity of 2–3 mm/s. Lower and higher blade velocities were less effective. A full investigation of the effect of blade coating parameters on morphology is certainly warranted as future work, given its relevance to roll-to-roll coating processes.

5.1.2 Image Analysis Strategies

Robust, accurate extraction and analysis of P3HT fibers from AFM images was demonstrated through a combined heuristic and machine learning approach. Coherence-enhancing anisotropic diffusion filtering and top hat filtering were demonstrated as necessary pre-processing steps to enhance segmentation performance in the presence of the significant noise and artifacts inherent to the imaging of soft materials with AFM. Subsequently, adaptive thresholding and skeletonization yield a representation of fibrillar structure that permits the extraction of fiber alignment and the decay length of orientational order.

To proceed further in analysis, fibers must be vectorized. It was demonstrated that breaking the skeletonized image into unbranched segments, fitting each segment with the Active Contours algorithm, and reconstructing fiber segments through a matching algorithm parameterized by maximum stitch gap and maximum curvature yielded accurate and computationally efficient fiber extraction and measurement. Parameter tuning by visual agreement yielded rapid and meaningful estimates of fiber alignment, orientational order decay length, fiber length density, and fiber length and width distributions. Structural measurement accuracy within 10% of ground truth values was demonstrated by optimizing image processing parameters through training on a set of manually traced images.

The resulting software package, GTFiber, is provided as open source software and is available as a standalone application for Windows or Mac.

5.1.3 Materials Informatics Infrastructure

Handling Process-Property Data

Significant insight was gained through the construction of a structured, open-access database for OFET processing and property information. This revealed early on the effect of molecular weight on device performance, as well as the importance of controlling device architecture and characterization decisions such as channel length, electrode material, and surface treatment. Through a rigorous search of the database, a standardized device was identified that had been reported with little variation in processing across multiple studies in the literature. This revealed subtle changes in process, architecture and characterization that still led to significant variance in reported performance. While literature reviews are sometimes carried out through database curation, the assembled database is infrequently

published with a searchable interface. This study additionally demonstrated the value of such an approach.

Material Image Databases

While the image analysis protocol presented in this work represents a step forward in a specialized case of microstructural analysis, there is potential for disruptive innovation in the broader ecosystem of materials imaging. This was demonstrated through the analysis of a large, diverse materials image database: all of the images from a single Atomic Force Microscope. Feature extraction through a pre-trained CNN or SIFT, combined with feature encoding, dimensionality reduction, and support vector machines, were shown to be highly effective at identifying whether images contained fibers, while image artifacts were classified with relatively lower accuracy. In addition, reverse image search was demonstrated, through which an image could be used as a search query to find other material systems that produced similar microstructures. The combination of these functionalities suggests that the assembly of large scale, open-access materials image repositories could improve the discovery of both low-level image analysis protocols and bring about unexpected connections between researchers, promoting greater interdisciplinary collaboration.

5.2 Suggested Future Work

5.2.1 *Leveraging IBM Watson for Initial Experimental Design*

A problem that frequently arises in academia and industry is that of initial experimental design on a new material system. Development teams generally have a specific performance target and a limited range of scalable processing strategies.^{179,180} Text-based search and discovery of promising literature results can often be an incomplete process, especially if the development team is making a foray into an area with which they have limited expertise. If the scientific literature were curated such that all data on performance metrics and process protocols were structured and stored under standard metadata tags, dynamic databases on materials experiments such as the OFET database could be constructed instantly, leading to simplified selection of initial experimental design spaces, as illustrated in **Figure 41**. A by-product of this infrastructure would be semi-automated generation of literature review articles.

IBM Watson was designed essentially for this task, and has been used with varying success for performing research across large volumes of literature data, mostly in the life sciences.¹⁸¹ Materials research stands to benefit equally from such an approach, especially given that physical properties are defined much more rigorously and quantitatively for materials than in biology.

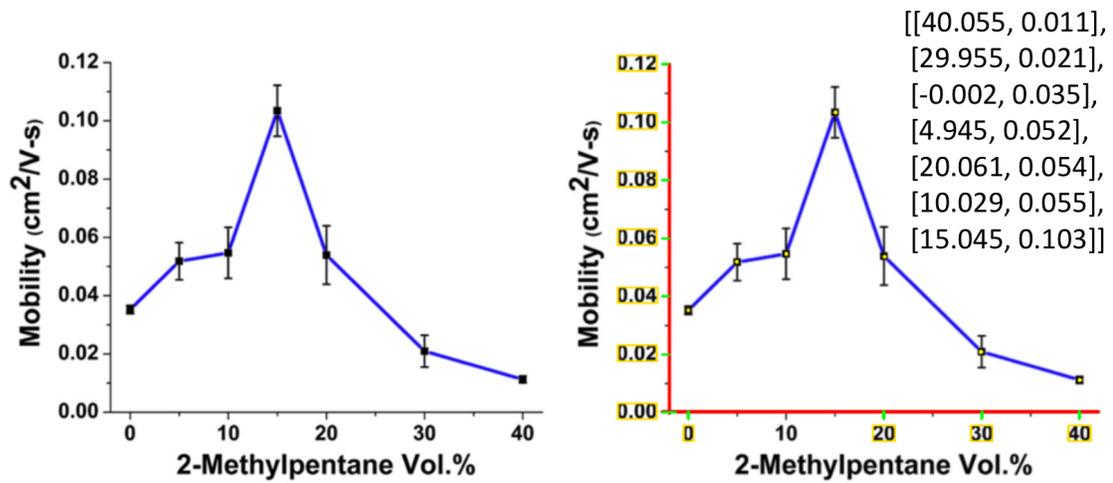
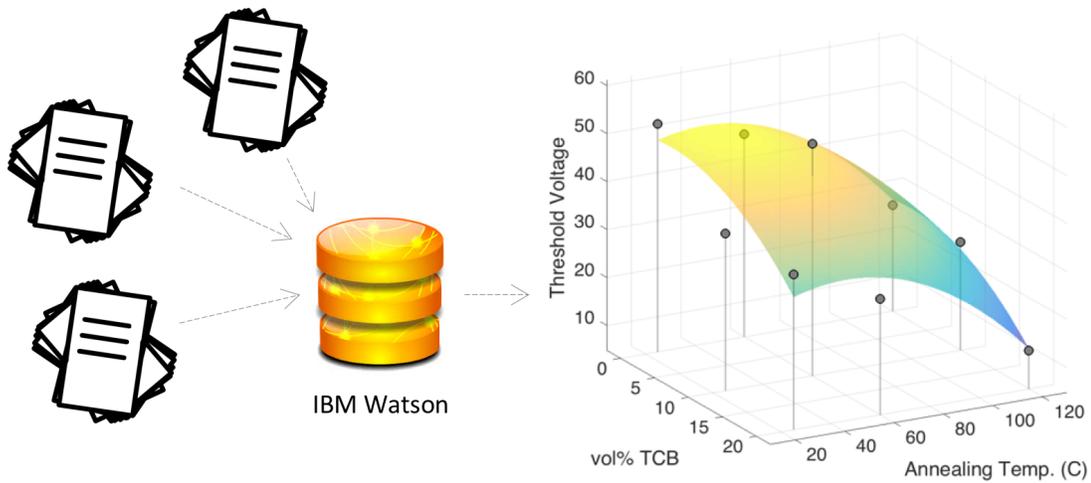


Figure 41 – (upper) Dynamic curation of process-property databases with IBM Watson, (lower) Automated extraction of raw data from published figures.

Extraction of Raw Data from Published Figures

The main limitation of IBM Watson is that it cannot read data from figures, which probably contain the majority of relevant scientific data. Automated extraction of data from published figures is perhaps one of the grand challenges in computer vision. State-of-the-art neural networks are currently limited to classifying what type of plot is presented in a given figure (bar, line, scatter, etc.).^{182,183} This problem has many parallels with fiber

extraction and measurement: line plots are essentially semi-flexible contours that require highly accurate object localization and measurement. As part of an NSF FLAMEL project, myself and two colleagues designed a proof-of-concept algorithm pipeline to identify axis labels and extract raw data from single- and multi- series scatter plots, shown in **Figure 41**. It combines Hough transforms, optical character recognition, some morphological transformations, and k -means clustering to accomplish this task. There is one similar approach to this problem in the broader literature for bar charts.¹⁸⁴ The solution to this problem in the general case could be fundamentally transformative for scientific research.

5.2.2 Generative adversarial neural nets for microstructure segmentation

Solving object localization and measurement problems for materials in the general case will require using state-of-the-art approaches from deep learning. Generative Adversarial Neural Networks (GANs) are currently making great strides in producing high-performance artificial intelligence models, and have even found use in some microstructural simulation tasks.¹⁸⁵ These function by having one neural net generate training examples while another neural net attempts to classify them. The generator steadily increases difficulty (through the addition of noise, other artifacts and affine transformations) in an attempt to break the classifier. Through this iterative approach, manual labelling of training sets becomes less necessary.

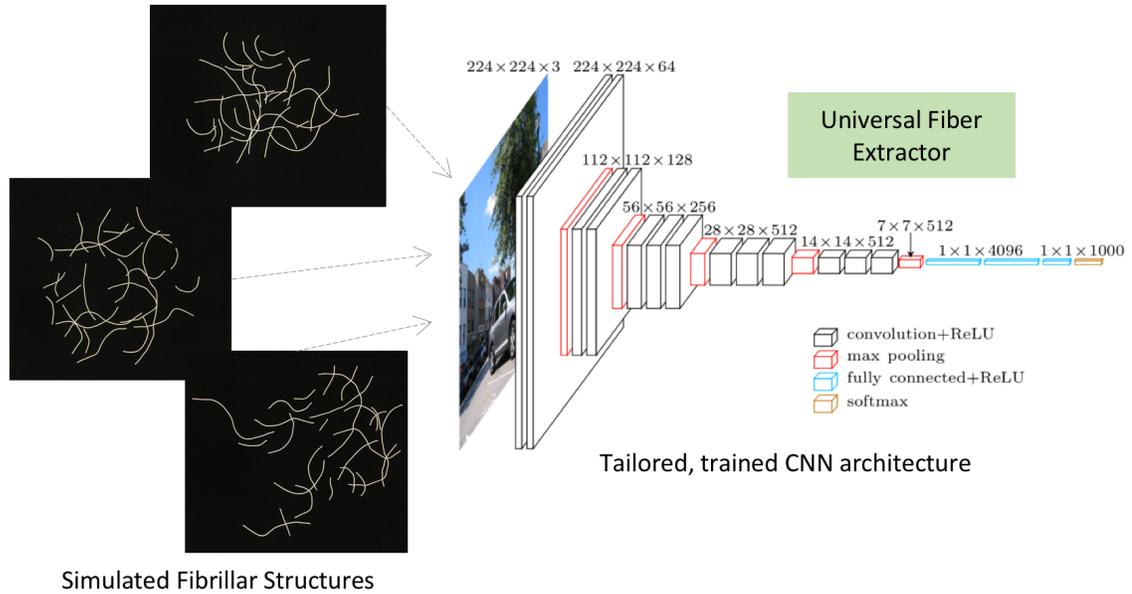


Figure 42 – Generative Adversarial Neural Net approach to universal object localization and measurement.

In this proposed project, a GANN would be trained by simulating packed fibrillar structures in increasingly complex scenarios, similar to the approach used in **Appendix A** for evaluating GTFiber. By simulating fibers with varied properties, packings, and levels of added image distortions and noise, a CNN could likely be trained to extract fibers in the general case. The difficulty in this approach is defining an objective function for the CNN. Simply producing a vector of class labels is not enough: the objective function must confirm that the *population* of fibers extracted matches that of the training image. GTFiber utilized average structural measurements for accuracy training. Vectorization algorithms can match these averages by playing tricks that generate unrealistic fiber populations. While this was accounted for in this work, a one-to-one population match is the best test of accuracy.

5.2.3 *In situ* characterization of OFET processing and properties

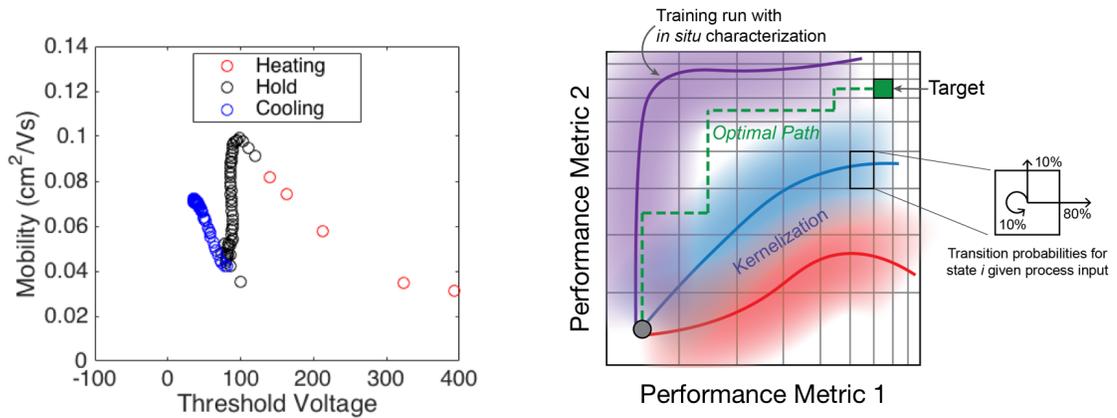


Figure 43 – *In situ* monitoring of device performance during processing.

Finally, this section would be incomplete without a recommendation for further studies of OFET processing. While explorations of microstructure are intellectually stimulating and provide significant insight into the physical mechanisms at play during processing and under device operation, sometimes microstructural considerations must be ignored in favor of rapidly optimizing performance. Incidentally, this was the stated goal at the outset of this project, but existing OFET processing methods were not well-controlled enough to enable such an approach, and the number of unresolved aspects relating to microstructure were too large to ignore. Nonetheless, attempts were made at establishing quantitative control of OFET performance through process control, with one example shown in **Figure 43**.

Here, both mobility and threshold voltage were measured *in situ* on an OFET device that was being annealed. Unfortunately, annealing was not an especially effective processing technique for our devices: as can be seen, while holding the annealing temperature, mobility dropped substantially, although threshold voltage was reduced (a

value closer to 0 is better). Since P3HT is no longer considered a state-of-the-art conjugated polymer, controlling performance by reducing mobility was not likely to be publishable. However, for newer conjugated polymers with more of a mobility cushion, such a control strategy could help to resolve the frequently observed variance in mobility. Indeed, for real world applications, mobility will need to be controlled within very tight tolerances.

This *in situ* approach happens to pair well with the Markov State Model + dynamic programming approach developed by the Grover group for *in situ* control of crystallization and colloidal self-assembly.^{64,147} Training runs are used to fill out the process parameter space and train the weights of a Markov State Model, indicating the effect of each process variable on performance at each discretized performance state (pair of mobility / threshold voltage). Kernelization of training run results can be used to fill out the model parameters for states that were not accessed during training runs. From here, dynamic programming is used to decide how to alter the process *in situ* to reach the desired performance target. The main difference between an OFET processing system and the crystallization systems considered previously is the lack of reversibility. Regardless, this approach could represent a unique approach for the optimization of OFET processing using *in situ* measurements.

APPENDIX A. GTFiber 1.0

GTFiber was developed in two stages. In version 1, the bulk fiber structure was extracted as a “skeleton” – an image in which fibers were thinned to single pixel width, and each pixel had an associated orientation. From this data structure, fiber alignment and orientational order decay length could be reliably estimated, but fiber length and width were not measurable. The results still retained some sensitivity to image processing parameters, as will be shown. Even at this stage, however, the program proved to be useful.

A.1. How to Spatially Average Orientations

The structure tensor maps angles and vectors into a space in which linear averaging is valid, and thus filtering techniques such as Gaussian filtering can be effective. Consider, for example, the set: $\{89^\circ, 90, 91, 269, 271, 180\}$. These angles all point more or less vertically, except for 180° , which is horizontal. Their numerical average is 165° and their median is 135.5° . These values are clearly not meaningful statistics for this set of angles – the average and median are both significantly horizontally biased. Statistics for angles are much more rationally computed in terms of the structure tensor, defined as:⁷⁹

$$J = \begin{bmatrix} \cos(\theta)^2 & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin(\theta)^2 \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \quad (5)$$

For two-component vectors, this is the same as taking the outer product. For the above set of angles, we end up with the following set of structure tensors (omitting the 3rd entry because it is identical to the 2nd):

Table 3 – Calculating average orientation with the structure tensor.

Angle	J_{xx}	J_{xy}	J_{yy}
89°	0.0003	0.0174	0.999695
90	0	0	1
91	0.0003	-0.0174	0.999695
269	0.0003	0.0174	0.999695
271	0.0003	-0.0174	0.999695
180	1	0	0
Average	0.1669	0	0.8331

J_{xx} describes the overall x-component of an orientation, i.e. “how horizontal”, J_{yy} describes the overall vertical component of an angle, and J_{xy} describes “how positively diagonal.” An angle of 45° would have the highest possible value of J_{xy} , whereas an angle of -45° would be the lowest.

From the averages of each element of the J matrices, we re-form a matrix $J_{average}$, and take its eigenvalues and eigenvectors:

$$eig(J_{average}) = eig\left(\begin{bmatrix} 0.1669 & 0 \\ 0 & 0.8331 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0.1669 & 0 \\ 0 & 0.8331 \end{bmatrix} \quad (6)$$

The eigenvector with the highest eigenvalue (0.8331) indicates the orientation of $J_{average}$, so [0; 1] is the average, a straight vertical vector, which makes perfect sense in the context of our hypothetical set. The above procedure is applied to find the black lines that are plotted at the average of each orientation distribution. When using the structure tensor for diffusion filtering, the scheme illustrated in **Figure 44** applies.

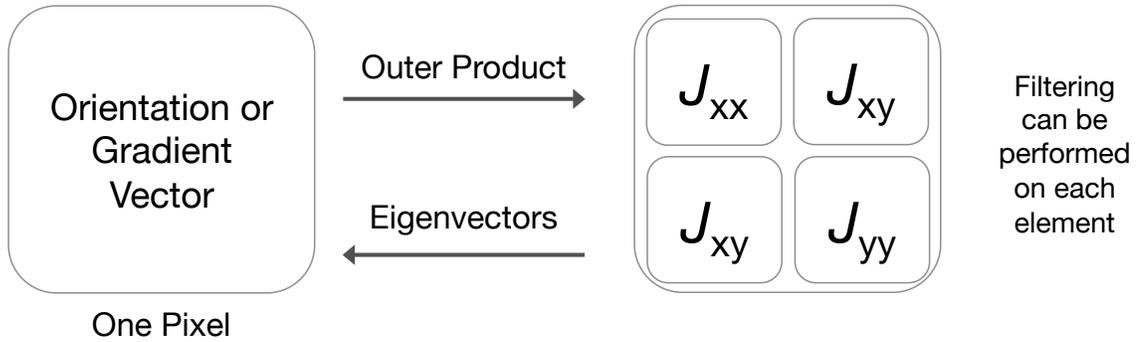


Figure 44 – Diagram of the use of the structure tensor. An $m \times n$ image is converted into an $m \times n \times 4$ image, where each pixel has a corresponding J_{xx} , J_{xy} , and J_{yy} (J_{xy} is repeated and thus redundant). Gaussian filtering can be applied to each layer of the $m \times n \times 4$ image, and then each pixel’s structure tensor is reformed, eigenvectors are computed, and a filtered orientation is returned.

A.2. Coherence-enhancing Anisotropic Diffusion Filtering

A.2.1 Algorithmic Details

Anisotropic diffusion filtering applies the principles of two-dimensional Fickian diffusion to the enhancement of images. Gray levels are the species undergoing diffusion, where “concentrations” can range from 0 (black) to 1 (white). The average gray level of the image is maintained, adherent to the equivalent of a mass balance in particulate diffusion. Isotropic diffusion would simply spread gray levels from bright regions to dark regions until a homogeneous gray level is obtained over the entire image – this is effectively the same thing as a recursively applied Gaussian filter. Anisotropic diffusion implies that the diffusivity is represented as a tensor, rather than a scalar, at each pixel.

The $m \times n$ image u is modeled as gray values between 0 and 1 as a function of position and time:

$$u(\mathbf{x}, t), \mathbf{x} \in \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} m \\ n \end{bmatrix} \right] \quad (7)$$

The diffusion equation to be integrated is the following, with a no-flux boundary condition at the image edges:

$$\frac{\partial u}{\partial t} = \text{div}(\mathbf{D} \nabla u), \quad t \in (0, \infty) \quad (8)$$

$$-\mathbf{D} \nabla u = 0, \quad x = 0, m; y = 0, n \quad (9)$$

where ∇u is the gradient of the image intensity and \mathbf{D} is a 2×2 positive definite diffusion tensor.

We would like \mathbf{D} to strengthen diffusion along the fiber directions, so that gray levels diffuse along fiber backbones, but not perpendicular to them. Furthermore, we would like regions of the image with high directional confidence to diffuse more strongly. These directions and their confidences can be obtained from the structure tensor \mathbf{J} described above in Section S1, by taking the outer product of the image gradient:

$$\mathbf{J}_0(\nabla u) = \nabla u \nabla u^T \quad (10)$$

A fiber can be visualized as a sharp ridge in the context of this analysis. The image gradient will be strongest at fibers' edges, and will be pointed *perpendicular* to fibers' backbone orientations. However, at the backbone of each fiber the image gradient will have a very low magnitude, and thus a weakly oriented structure tensor. To account for this, and to smooth noise in the locally calculated gradients, each element of \mathbf{J}_0 is convoluted with a Gaussian filter, K_ρ :

$$\mathbf{J}_\rho(\nabla u) = K_\rho * (\nabla u \nabla u^T) \quad (11)$$

where ρ is the standard deviation of the Gaussian filter. Due to the unique properties of the structure tensor described above, this accomplishes a local averaging of orientations. The

more strongly oriented pixels will influence their neighbors; so the center of a fiber between two sharp edges will adopt a strong orientation pursuant to that of the edges.

Now the diffusion tensor can be constructed. It carries the same eigenvectors as \mathbf{J}_ρ , but its eigenvalues are replaced with values that enhance diffusion along the *coherence orientation*, or the fiber direction. The eigenvalues of \mathbf{J}_ρ are sorted so that:

$$\lambda_1 \geq \lambda_2 \quad (12)$$

and the eigenvector \mathbf{v}_1 therefore corresponds to the locally averaged gradient vector, and \mathbf{v}_2 corresponds with the desired coherence orientation. Coherence is a measure of the anisotropy of the structure tensor, calculated as:

$$\kappa := (\lambda_1 - \lambda_2)^2 \quad (13)$$

The coherence is used to formulate new eigenvalues for the diffusion tensor:

$$\lambda_{D1} := \alpha \quad (14)$$

$$\lambda_{D2} := \begin{cases} \alpha, & \kappa = 0 \\ \alpha + (1 - \alpha)\exp(-C/\kappa), & \kappa \neq 0 \end{cases} \quad (15)$$

In both of these equations, $\alpha \in (0,1)$ is a small positive parameter that ensures finite isotropic diffusion even when no coherent structures are present. An α of 1 results in perfectly isotropic diffusion, equivalent to a sequence of Gaussian filters. The anisotropic nature of the diffusion tensor comes from the second term of λ_{D2} . The parameter C acts as a threshold for κ above which the diffusion tensor will become strongly anisotropic, and below which diffusion will tend toward the standard isotropic case.

To summarize, large local gradients in an image leads to a structure tensor with strongly differing eigenvalues, which in turn lead to a diffusion tensor with strongly differing eigenvalues. The only difference between the two tensors is that the structure tensor's largest eigenvalue corresponds to the gradient direction, while the diffusion

tensor's largest eigenvalue corresponds to the fiber direction, or coherence orientation. To mathematically construct the diffusion tensor \mathbf{D} , we formulate a new matrix using the eigenvectors of \mathbf{J}_ρ and the new eigenvalues calculated in Eqns. 10 and 11:

$$\mathbf{D} = \mathbf{V} \begin{bmatrix} \lambda_{D1} & 0 \\ 0 & \lambda_{D2} \end{bmatrix} \mathbf{V}^T \quad (16)$$

where \mathbf{V} is the matrix containing the original eigenvectors of \mathbf{J}_ρ . Finally, the gray scale image is updated according to Eqns. 4 and 5 using an explicit Euler integration scheme and rotationally invariant kernels for spatial derivatives.¹⁸⁶ The default values of α , C , and the integration time step are 0.001, 1×10^{-10} , and 0.15, respectively. The results of filtering are invariant to deviations in these parameters, so they are not explored in depth here.

A.2.2 Effect of Orientation Smoothing

At the end of the day, only two parameters are truly adjustable in anisotropic diffusion filtering: the size of the Gaussian filter on the gradient orientations, ρ , and the diffusion time, which dictates how many time steps are taken when integrating. Heuristically, the value of ρ should not exceed the typical fiber width, otherwise orientations from high-intensity fibers will influence that of their neighbors. This is illustrated in **Figure 45**. Stronger structure tensor filtering leads to stronger measured local alignment, so this parameter is held as low as possible while an accurate segmentation is still obtained. As long as image scales and parameters are applied consistently, however, the trends in structural metrics will always hold. This source of variance is eliminated by vectorization, as shown in **Appendix B**.

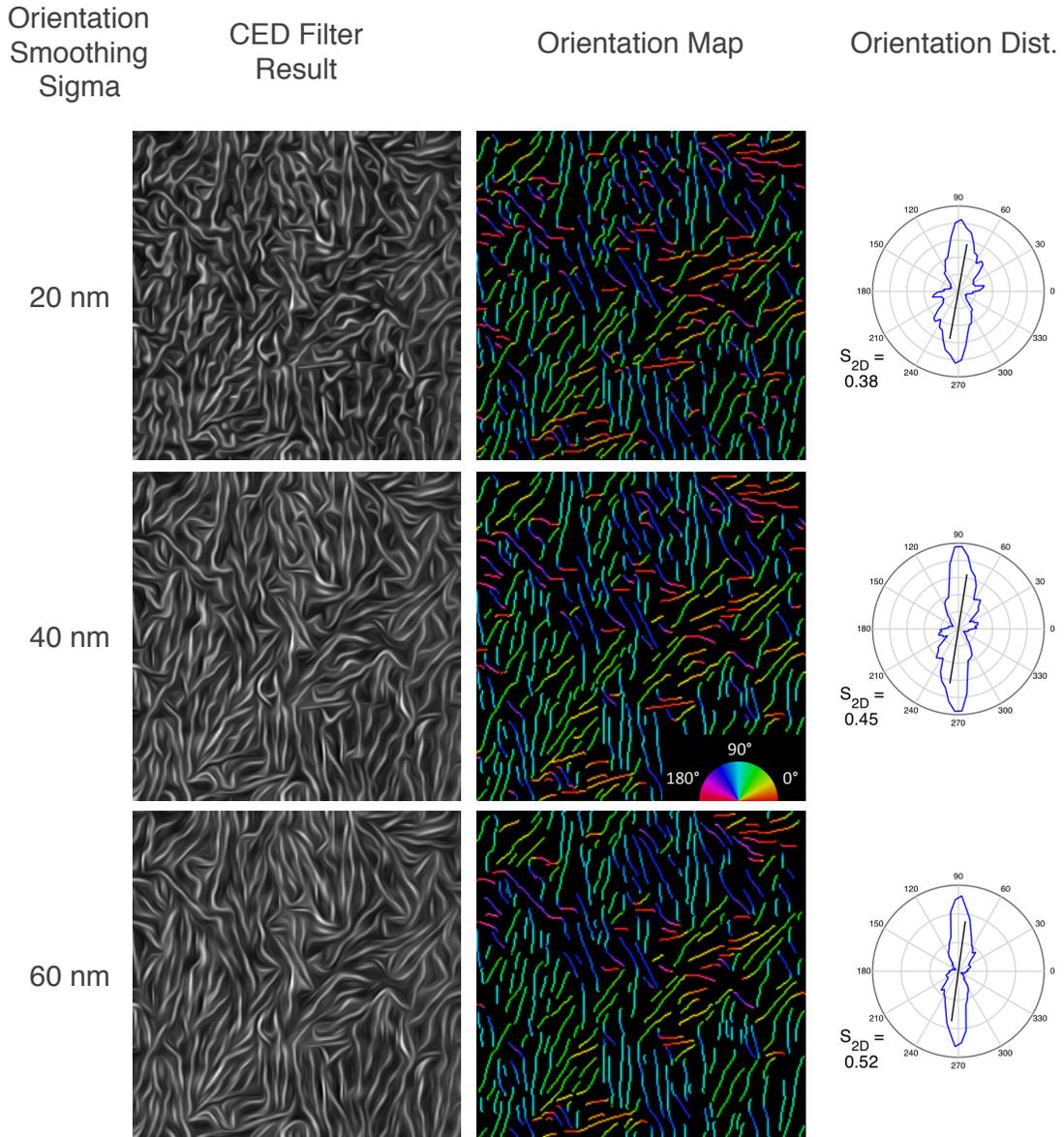


Figure 45 – The effect of structure tensor smoothing. Standard deviation of orientation smoothing kernel in nm indicated at left, followed by the gray scale result of diffusion filtering, the Orientation Map, and an orientation distribution. Other parameters used for this filter: initial Gaussian smoothing, 10nm; diffusion time, 8s; top hat size, 30nm; adaptive thresholding; noise removal area, 3500 nm²; skeleton fringe removal, 40nm.

A.2.3 Effect of Diffusion Time

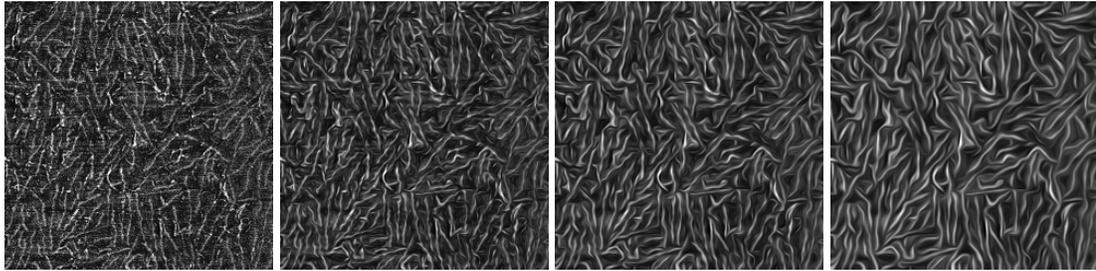


Figure 46 – The effect of the “diffusion time” parameter on diffusion filtering results, from left to right: original gray scale, 1s diffusion time, 3s, 8s. Constant parameter values: initial Gaussian smoothing, 10nm; orientation smoothing, 30nm.

Diffusion time is illustrated in **Figure 46**. Gray values diffuse by a default time step of 0.15s until a time specified by the user in the field “diffusion time”. The effect of time step is negligible, as long as it is fractions of a second. The only consideration for choosing a different time step would be for computational efficiency. However, as diffusion is allowed to proceed, one observes a pronounced smoothing of fibrillar regions. Wrinkles and kinks in fibers tend to be straightened, and segments that may have appeared to be separate may connect. CED filtering preserves the average intensity of the image, so no wash-out will occur. While the 8s diffusion time appears over-smoothed in the gray image, it yields a more accurate skeleton than shorter diffusion times. The orientation distribution and S_{2D} are not strongly affected by diffusion time, so it is more of an aesthetic choice. The user is encouraged to explore this parameter through the GUI to determine what works best for their application.

A.3. Top Hat Filtering, Thresholding, and Skeletonization

A.3.1 Top Hat Filtering

Top hat filtering enhances sharp details and flattens the background in images. Its name is derived from the shape of its structuring element, which looks similar to a top hat. The transformation consists of three steps: an erosion, a dilation, and a subtraction. The combination of erosion and dilation is called an image opening, so the top hat filter, in essence, subtracts from the original image the result of its opening. This is illustrated in **Figure 47**, using a structuring element of a disk of radius two pixels.

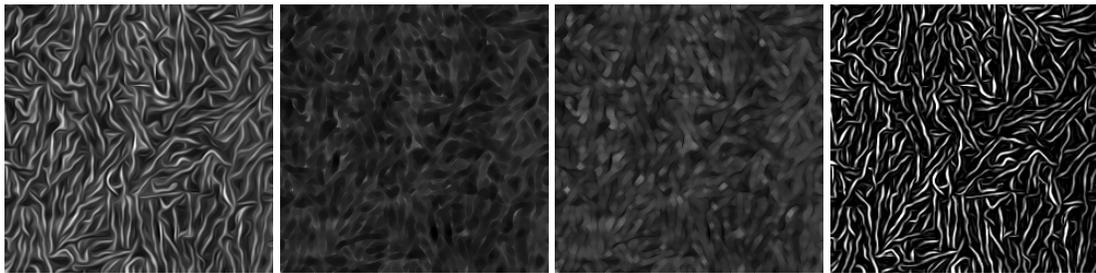


Figure 47 – The top hat filtering process, from left to right: result of anisotropic diffusion filtering, result of image erosion, result of image dilation, and result of subtracting the dilated image from the result of anisotropic diffusion filtering.

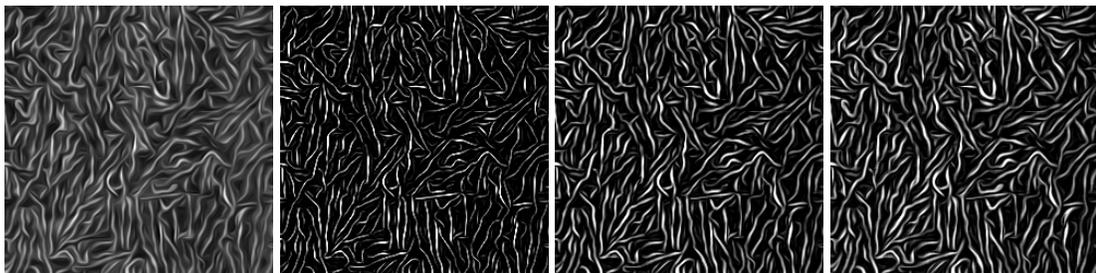


Figure 48 – The effect of the “top hat size” parameter on contrast enhancement, from left to right: initial result of diffusion filtering, result of 10nm top hat filter, 30nm, 50nm. Constant parameter values: initial Gaussian smoothing, 10nm; orientation smoothing, 30nm; diffusion time, 8s.

The top hat filter is a robust contrast-enhancing filter that is incredibly effective after diffusion filtering has been applied. It enhances sharp peaks in an image and suppresses the background. The images in **Figure 48** are all gray scale – still no thresholding has been applied. The top hat size parameter determines the radius of the disk-

shaped structuring element used for the erosion and dilation operations performed during top hat filtering. It is apparent from these images that a top hat size greater than or equal to the smallest fiber width should be used; very small top hat size excessively erode the thinnest, lightest fibers. It should be noted that the actual top hat disk radius used in the program is converted from nm to pixels and rounded up to the nearest pixel size – only integers are allowed for this morphological operation.

A.3.2 Thresholding Choices

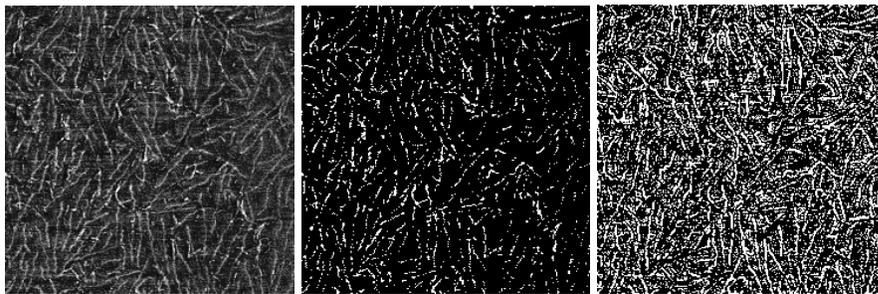


Figure 49 – Thresholding of original image. Left: original gray scale. Center: Automatic global threshold. Right: Adaptive threshold surface using Yanowitz-Bruckstein method.¹⁸⁷

Figure 49 shows traditional thresholding approaches applied to the original gray scale image. Clearly the smoothing and contrast enhancing filters were necessary before thresholding. Two thresholding options are offered to convert the result of top hat filtering into a binary image: adaptive thresholding using the method introduced by Yanowitz and Bruckstein,¹⁸⁷ and a simple global threshold. Noise removal is also offered to clean the resulting binarized image: connected components (contiguous white regions) of area less than the specified maximum noise area are removed. The effect of these operations are shown in **Figure 50** and **Figure 51**.

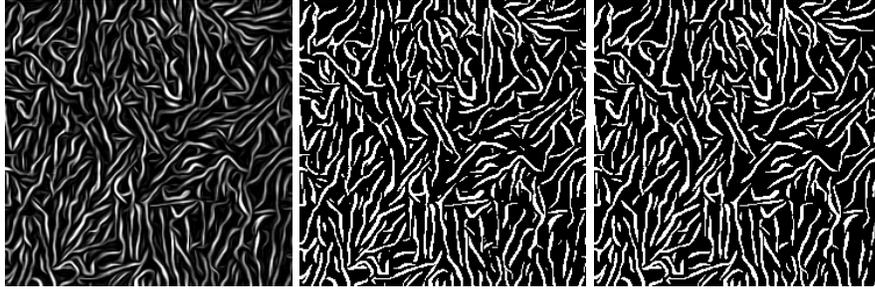


Figure 50 – Thresholding of image after diffusion and top hat filtering. Left: gray scale top hat filter result. Center: global threshold at gray value of 0.3. Right: Noise removed if area < 3500 nm².

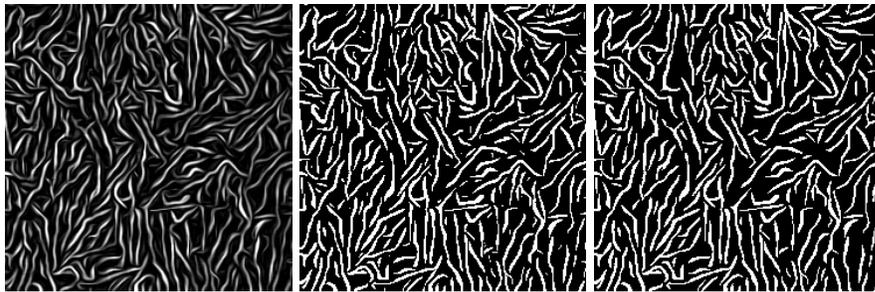


Figure 51 – Thresholding of image after diffusion and top hat filtering. Left: gray scale top hat filter result. Center: result of adaptive thresholding. Right: Noise removed if area < 3500 nm².

Since the contrast-enhanced diffusion and top hat filtered image is already so clear, thresholding is now trivially simple. Even a global threshold of 0.3 yields nearly the same result as a more advanced adaptive thresholding operation. Additionally, the binarized structure is very smooth, which aids in the skeletonization process.

A.3.3 Skeletonization

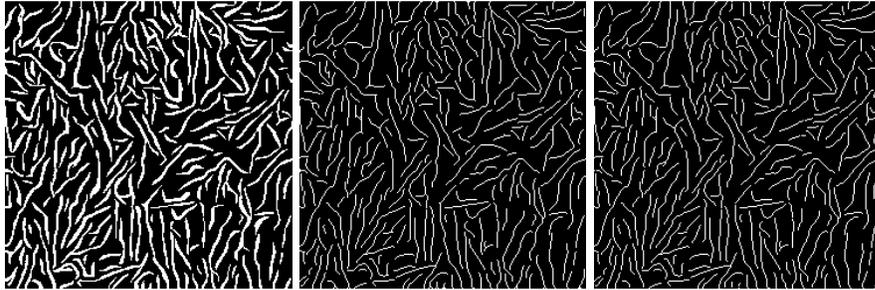


Figure 52 – Skeletonization and trimming. Left: result of adaptive thresholding and noise removal. Center: result of skeletonization. Right: skeleton after removal of branches less than 40nm long.

Skeletonization is the final processing step to be performed. It is a common morphological operation with a built-in MATLAB function that thins connected white components to single-pixel width. Bumps along the edges of the white components can cause the skeleton to produce small fringe branches along the main backbone, so a feature has been included that removes fringe branches of a specified length or less. Very few fringe branches can be seen in **Figure 52**, so a better example is provided in **Figure 53**. Essentially, this algorithm removes pixels if the sum of their geodesic distances from their nearest branch point and end point is less than the specified length, and that pixel lies *between* its nearest branch and end points (indicating that it is actually on a branch).



Figure 53 – Skeleton fringe trimming, before and after.

A.3.4 Other Imaging Systems

As long as an image contains fibers whose brightness contrasts substantially with the image background, contours can be extracted and analyzed, as demonstrated in **Figure 54**. With different imaging systems, however, length scales can vary drastically. For the 10 mm micro-CT image in **Figure 54A**, it is impractical to enter the image size in nanometers. In this case, simply enter 5000 nm as the image width, and use the default filter parameters that are pre-filled in the app window. The default parameters are optimized for a 5000 nm image and will generally yield a good first result. The results in **Figure 54 A-C** were obtained using this strategy. In general, the units of the filtering parameters should be determined relative to the specified image size.

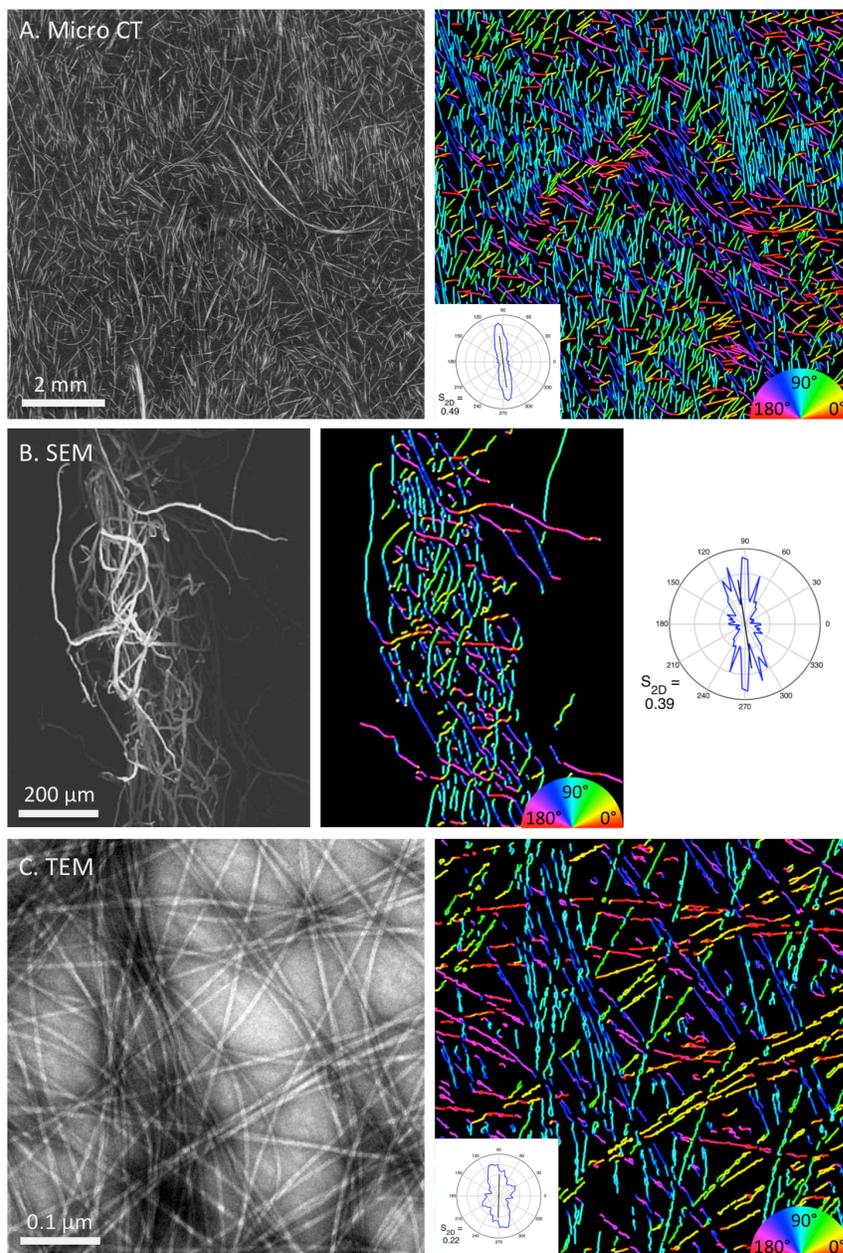


Figure 54 – GTFiber applied to: (A) Micro-CT imaging (top) of a fiber composite, (B) SEM imaging of cellulose fibrils (reproduced from Ref. 4 with permission from the Royal Society of Chemistry),¹³⁴ (C) TEM imaging of fibrillar protein assemblies.

APPENDIX B. GTFiber 2.0

In version 2 of GTFiber, length and width measurement are introduced through a vectorization and segment matching and reconstruction process. This has the added benefit of removing the sensitivity to orientation smoothing. It also permits the measurement of fiber curvature, as well as the generation of a number of striking data visualizations.

At its core, GTFiber is a collection of image processing and fiber modeling algorithms that enable the segmentation, vectorization and analysis of two-dimensional images of dense populations of fibers. Segmentation refers to the classification of which pixels are in fibers. Vectorization refers to the identification and grouping of pixels that belong to the same fiber, followed by their conversion to a sequence of vectors, called a contour. Segmentation enables the extraction of orientational order parameters, while vectorization further enables the estimation of fiber length, width, and curvature distributions, as well as the juxtaposition of higher-resolution structural simulations. The segmentation process encompasses algorithm steps 1–5 (**Figure 55**), while steps 6 and 7 make up the vectorization process.

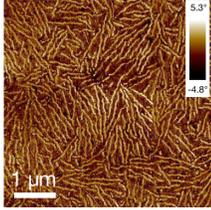
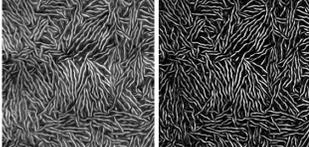
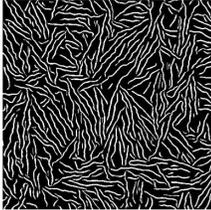
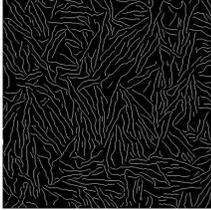
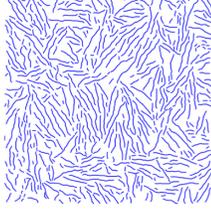
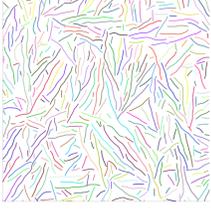
Algorithm Step	Underlying Data Structure	Data Visualization
1. Original AFM Image	$M \times N \times 1$ image M, N = height, width in pixels Intensity corresponds to height or phase data	
2. Diffusion Filter 3. Top Hat Filter	$M \times N \times 1$ image Intensity has been smoothed and contrast enhanced	
4. Threshold & Clean	$M \times N \times 1$ binary image Each pixel stored as 0 (background) or 1 (fiber)	
5. Skeletonization	$M \times N \times 1$ binary image White regions from previous step are thinned to single-pixel width	
6. Segment Vectorization	$S \times 1$ structure array S = # of fiber segments Each element s contains: - list of pixels in segment, ordered from one endpoint to the other; - list of vectors fitted to segment using Active Contours	
7. Fiber Vectorization	$F \times 1$ structure array F = # of Fibers Each element f contains: - list of segments s in fiber, ordered from one endpoint to the other; - list of vectors fitted to fiber using Active Contours	

Figure 55 – Overview of the GTFiber algorithm stack and data structures.

B.1. Fiber Vectorization

While segmentation was detailed in **Appendix A**,¹¹⁴ we review the main steps as well as new additions in **Figure 55**. The first goal is to obtain an image in which pixels in fibers are white and the background is black – referred to as binarization. The result in **Figure 55b** is obtained after applying an anisotropic diffusion filter (for oriented smoothing), a top hat filter (to enhance ridge-like objects), an adaptive threshold (accounts for long-range variations in pixel brightness), and noise removal (removes connected objects less than a specified area). To vectorized fibers, however, we require a list of points along the backbone of each fiber *in linear order*, *i.e.* ordered from the start to the end of the fiber. The binarized image contains branched regions which, while physically possible from a self-assembly standpoint, make it significantly more complicated to obtain an ordered list of points for each fiber.

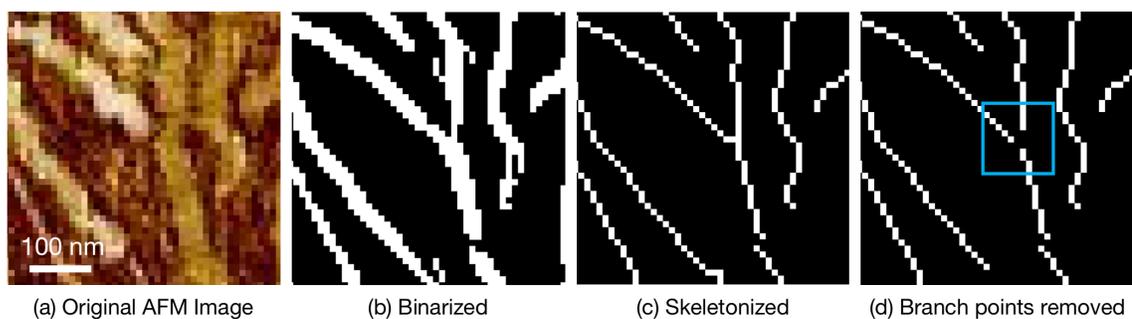


Figure 56 – Segmentation, skeletonization, and segment isolation in GTFiber.

Skeletonization solves this problem by thinning a black and white image so that the white regions have single-pixel width, as shown in **Figure 55c**. This step includes a pruning algorithm that removes branches of the skeleton less than a specified length – usually the width of the fibers in question. In the skeletonized image, branch points are trivial to

identify and remove, yielding an image of unbranched fiber segments, shown in **Figure 55d**. In addition to branch point removal, segments with extremely high curvature (over the specified maximum, described below) are broken as well. However, in removing branch points, it is likely that some fibers are broken into smaller segments, and need to be reconstructed. To do this, we first vectorized each of these isolated, unbranched segments.

B.1.1 Segment Vectorization

Vectorization utilizes the Active Contour model to convert a list of pixels into a list of vectors. Active contours can be described as a discrete contour (a sequence of equally-sized segment vectors) with an internal energy term penalizing bending between neighboring segments, and an external energy term derived from the surrounding image gradient, a vector field which points toward the bright central line of a fiber-like object. The result is a list of vectors which obey structural constraints such as limits on curvature, while conforming to the region of an image that appears the most fiber-like. It is the core algorithm in FiberApp, by Usov *et al.*, in which a user identifies fibers of interest by manually clicking on points along their backbone. In the high-throughput fiber extraction algorithm introduced in GTFiber, fibers are first broken down into isolated skeletal segments, which can be represented as a list of pixels. The Active Contour Model is used here to convert that list of pixels, which have integer coordinates, to a list of vectors, which have continuous x-y coordinates in units of nanometers. For further reading see Usov *et al.* and Smith *et al.*^{74,188}

In 2D, let $\mathbf{r}(s) = (x(s), y(s))$, $s \in [0, L]$ be the parametric representation of an open contour (**Figure 57**), where s represents arc length along the open curve, and L is the

length of the active contour. The starting and the ending points of the active contour are $s = 0$ and $s = L$ respectively. A set of N discrete sampling points $\mathbf{r}_i = (x_i, y_i)$, $i = 1, 2, \dots, N$, is sampled at equally spaced intervals from the active contour to represent it.

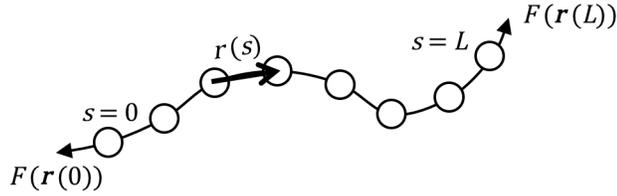


Figure 57 – Discretization of a fiber, where s is the path length coordinate and r is the tangent vector at s . F is the stretching force applied to the ends of the contour.

The internal energy of an active contour penalizes its curvature, smoothing the abrupt noise influence from the image. The internal energy term, E_{int} , is defined as:

$$E_{int} = \int_0^L \beta |\mathbf{r}_{ss}(s)|^2 ds \quad (17)$$

where $\mathbf{r}_{ss}(s)$ is the second derivative of the tangent vector, and β is a coefficient that penalizes local bending. The external energy consists of two terms: the image term E_{img} and the stretching term E_{str} :

$$E_{ext} = \int_0^L k_1 E_{img}(\mathbf{r}(s)) + k_2 E_{str}(\mathbf{r}(s)) ds \quad (18)$$

where two additional weighting constants are applied, k_1 , which determines how strongly the contour seeks the brightest areas of the image, and k_2 , which determines how strongly the contour will stretch at its ends. The image energy, E_{img} , is simply the inverse intensity of the image at the point s , which is interpolated between pixels. Thus, minimizing E_{img} equates with moving the points of the contour along the image gradient (illustrated in **Figure 58b**).

E_{str} is derived from a force applied to the ends of the contour ($s=0$ and $s=L$):

$$F(\mathbf{r}(s)) = \frac{I(\mathbf{r}(s))}{I_f} \quad (19)$$

$$\nabla E_{str}(\mathbf{r}(s)) = \begin{cases} -F(\mathbf{r}(s)) \mathbf{t}(s) & s = 0 \\ F(\mathbf{r}(s)) \mathbf{t}(s) & s = L \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

The force, F , is applied only to the endpoints, tangential to the terminating unit vectors, $\mathbf{t}(s)$ at each end. Its magnitude is proportional to the image intensity at each tip relative to fiber intensity. The contour is stretched if its ends are still in bright regions, but it is contracted if its ends lie in dark regions.

Figure 58 provides a visual overview of the active contours algorithm. The contour is initialized with some set of points (red circles, **Figure 58a**), from which equally spaced initialization points are interpolated (blue circles, **Figure 58a**) at a specified step length. These points are equilibrated by iteratively shifting their x-y coordinates to minimize the sum of E_{int} and E_{ext} using an implicit Euler integration scheme for a specified number of iterations – 10 is sufficient in this case (magenta dots and circles, **Figure 58b**). However, being a local energy minimization, a very good initialization is essential to obtaining a good contour. The natural choice for the initialization is the set of pixels from each isolated, unbranched skeletal segment, shown in **Figure 58d**. For each segment found in an image, its skeletal pixels are fed to the active contours algorithm and the contour's energy is minimized on the original grayscale image. The result of this is shown in **Figure 58e**.

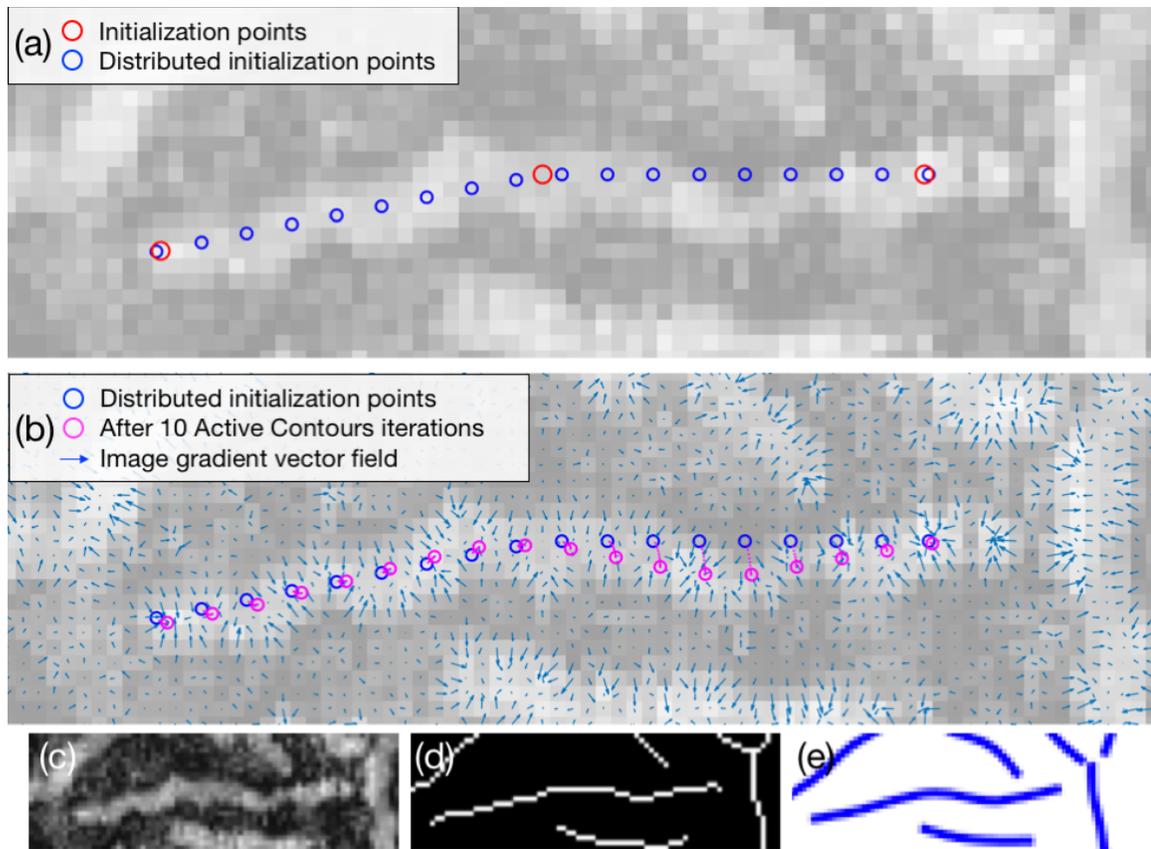


Figure 58 – Energy minimization of an active contour and visualizations of the first three stages of GTFiber. (a) Example initialization of an active contour. Red circles: three initialization points. Blue circles: distributed initialization points interpolated between the red initialization points, but with 3-pixel spacing. (b) Energy minimization of the active contour (10 iterations), with $\beta = 5$, $k_1 = 20$, and $k_2 = 10$. Image gradient vector field overlaid for illustrative purposes. (c) Original grayscale image. (d) Skeletonized image extracted using GTFiber’s image processing operations. (e) Plot of the vectorized version of the skeleton in (d) using Matlab’s *line* plotting function (with anti-aliasing).

Since the skeletal segment pixels are an extremely good initial guess for Active Contours, a set of default parameters work in almost every case here. The stretching parameter, k_2 , is set to zero, the image gradient parameter, k_1 , is set to 20, and the curvature penalty, β , is set to 5. These parameters ensure that the initial segment vectorization is not significantly different than the skeleton – however the Active Contours algorithm does smooth the more abrupt features of the skeleton. The most important parameter to select

here is the step length, which determines how many points are used to vectorize each segment. The step length should be short enough to capture the local curvature of the fibers, but not shorter than two pixels – shorter step lengths increase processing time. Here, a 30 nm (3 pixel) step length is used.

While **Figure 58d** and **Figure 57e** look nominally similar (which they should), the underlying data structure of the vectorized segments is more powerful than the skeleton in that mathematically defined measures of length, curvature and orientation can be extracted from it. First, segments that were broken apart in **Figure 56d** must be reconstructed.

B.1.2 Fiber Reconstruction

The current state of the art in contour reconstruction involves minimizing the cost of path ensembles through a graph representation of the contour network.^{71,72} This approach is reliant on a clean pixel-wise segmentation, which is highly dependent on the level of image noise. The smoothing filter and adaptive thresholding algorithm in GTFiber alleviate many problems due to noise and inconsistent illumination, but an image with hundreds of fibers presents many cases of fibers incorrectly broken into smaller segments, or overlapping fibers fused into branched structures. A graph theory approach requires all fibers to be fully connected by nodes of the skeleton, which is not always the case. It is also a computationally expensive approach, requiring the enumeration of many paths through a complex network of nodes and edges.

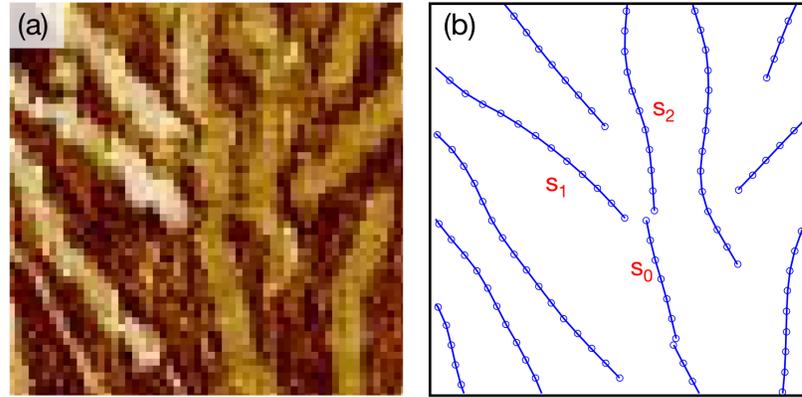


Figure 59 – (a) A portion of a raw AFM image. (b) Line plot of the broken, vectorized segments.

We found that the most robust approach to reconstruction in our case was to break the entire image skeleton into unbranched segments, then rebuild or “stitch” the fibers from this new starting point. Once all segments have been broken down and vectorized, they must be stitched together into fibers. Stitching of neighboring segments is decided through a two step process: identification of candidate matches and match scoring. The example shown in Figure 59 was chosen because it represents an extremely difficult reconstruction decision: even from a human perspective, an argument could be made for stitching s_0 with either s_1 or s_2 . While most cases are much more straightforward and do not require a complex reconstruction algorithm, this case is used to illustrate the level of control provided by GTFiber, if necessary.

The scoring and matching algorithm is laid out in **Figure 60**. First, candidate matches are identified by projecting a search kernel out from the end of s_0 in the direction of its terminating vector. Endpoints of other segments falling within this search kernel are scored for matching. While it appears in **Figure 60a** that three segments may be candidates, we consider only s_1 and s_2 in this example. One could in theory generate a score between

each segment and each other segment in the entire image, but the search kernel approach drastically reduces the computational burden of match scoring.

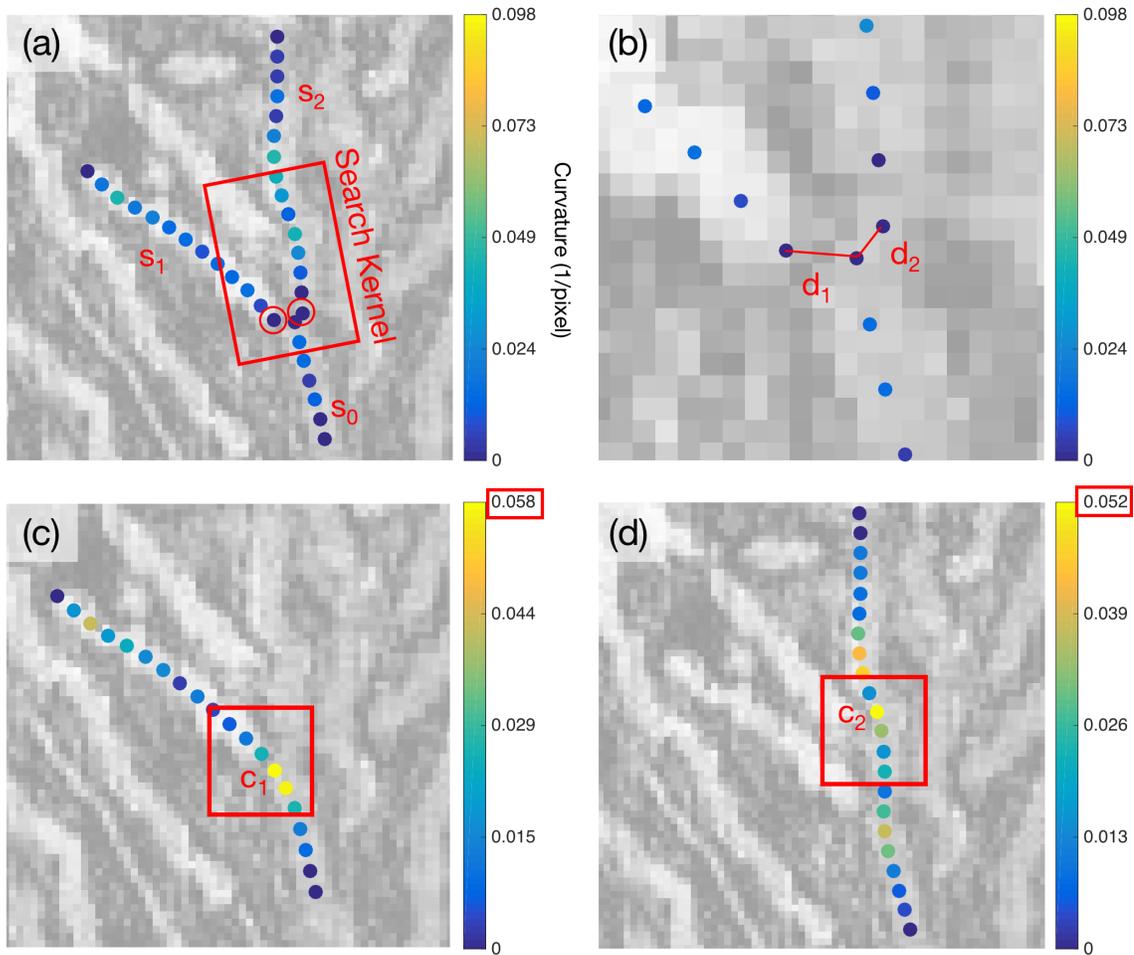


Figure 60 – Illustration of segment matching and scoring. (a) Candidate endpoints are identified through a projected search kernel. (b) Distances are measured between the searching endpoint and its match endpoints. (c) A test fit of an active contour combining segments s_0 and s_1 . Curvature at each point is indicated by the color scale at right in units of 1/pixel. Red boxes indicate the points considered for curvature scoring. (d) Test fit of an active contour between segments s_0 and s_2 .

The score for a match penalizes the distance between endpoints and the curvature added by stitching the two segments together, using the following formula:

$$(Total\ Score)_i = \begin{cases} disqualified & \text{if } d_i > d_{max} \text{ or } c_i > c_{max} \\ \frac{d_i}{d_{max}} + \frac{c_i}{c_{max}} & \text{otherwise} \end{cases} \quad (21)$$

where d_i is the distance between the searching endpoint and the candidate match's endpoint, and c_i is the maximum curvature of all contour points added between the segments during stitching. Each of these components of the score are scaled by the user-specified parameters d_{max} and c_{max} , the maximum allowable endpoint separation and curvature, respectively. If *either component* exceeds its specified maximum, the match is disqualified. Matches for every endpoint of every segment are ranked and mutual top matches are awarded first, followed by top choice / second choice matches if neither endpoint is already involved in a top / top match; the guiding principle being that each endpoint can only match once.

The distance score is trivial to calculate, as illustrated in **Figure 60b**, but the curvature score is slightly more involved. The goal of the curvature score is to determine how much curvature is added by stitching two segments together. The curvature, κ , at each point in a contour is defined at each point as:

$$\kappa = \frac{d\phi}{ds} \quad (22)$$

where ϕ is the angle of the contour's tangent vector in radians and s is arc length along the contour. Since each active contour is parameterized as $\mathbf{r}(s) = (x(s), y(s))$, the following formula for curvature in a two-dimensional curve applies¹⁸⁹:

$$\kappa = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}} \quad (23)$$

where derivatives are computed using centered finite differences along the point list of an active contour. Curvature at each point is plotted in **Figure 60** using a color scale.

To score curvature between two segments, a new active contour is fit to the image using the points from the matched segments as an initialization. These test fits are illustrated for segments $s1$ and $s2$ in **Figure 60c & d**, respectively. Note that the color scale is narrower in these images to highlight the differences. Only new points falling in the gap between the matched endpoints are considered for the curvature score, indicated by red boxes. The maximum curvature from among these points is used as the score c_i , which is then scaled by the chosen c_{max} .

Once all segment endpoints have found their matches, if any, the final vectorized fibers are determined by fitting an active contour using the set of all points in all of a fiber's constituent segments as an initialization. For the region considered in this example, the result of this can be seen in **Figure 61b**, including a curvature color overlay. Note that the segment at the other end of s_0 in **Figure 59b** was also correctly stitched.

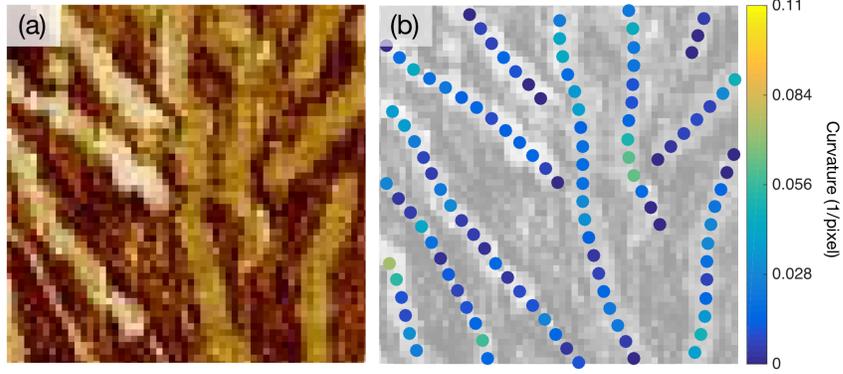


Figure 61 – (a) Original AFM image. (b) Vectorized fibers with curvature indicated by a color scale at each point.

B.2. Order Parameter Calculation

Having vectorized fiber contours permits the calculation of several structural order parameters. First, we go over the calculation of S_{full} and λ_C , (fiber alignment and orientational order decay length) followed by ρ_{FL} (fiber length density). S_{full} and λ_C are orientational order parameters derived from a plot of S_{2D} calculated at expanding frame sizes averaged across the entire image, as illustrated in **Figure 62**. S_{2D} is a standard orientational order parameter that calculates the alignment of a population of vectors relative to their director. Here, it varies from 0 to 1 because the director is defined as the average orientation of the vectors. At very small frame sizes, the contour vectors of fibers are highly aligned because only vectors from a single fiber are considered. As the frame expands, however, alignment decreases. Small fluctuations in this curve are normal, but large fluctuations indicate that a larger image size should be used to quantify orientational order. The data are well fit by an exponential decay curve, modified with an asymptote. The characteristic length of exponential decay is defined as λ_C (thus “decay length”), while the height of the asymptote is defined as S_{full} , for “full image orientational order.” In

practice, S_{full} is a useful metric to quantify overall fiber alignment, while λ_C can be used to quantify various types of local disorder, as described in the main text.

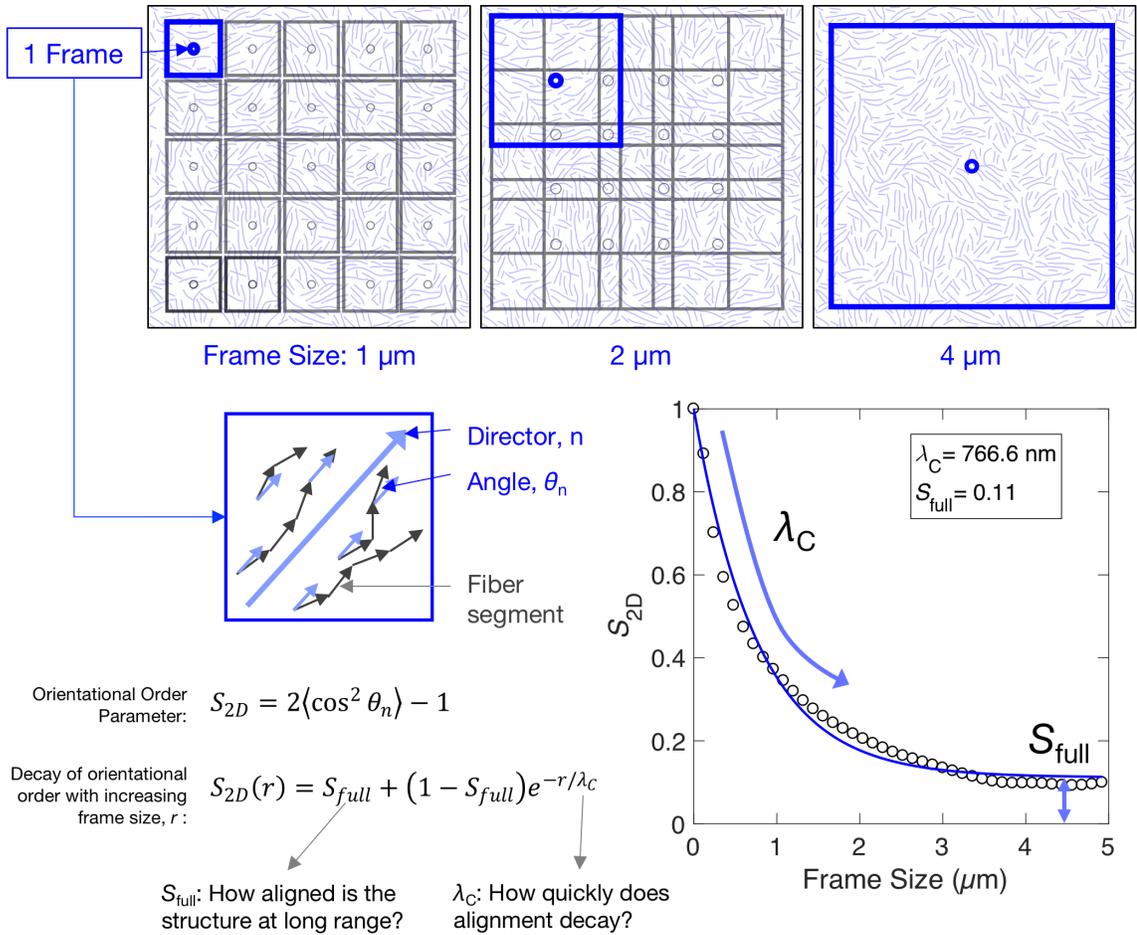
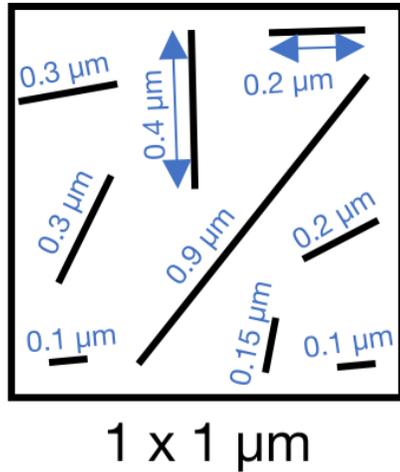


Figure 62 – Calculation of S_{full} and λ_C .

The calculation of fiber length density, or ρ_{FL} , is more straightforward. One simply measures the total length of fibers within a given frame, and divides by the area of that frame, as illustrated in **Figure 63**. It is similar to the area fraction of fibers, but is invariant to their width. One could also imagine calculating a decay function for ρ_{FL} as a function of frame size, similar to orientational order – this would quantify spatial inhomogeneities in

fiber packing. Since P3HT nanofibers tend to pack quite evenly, this was not considered in the current study.

Fiber Length Density, ρ_{FL}



$$\rho_{FL} = \frac{\text{Total length of fibers}}{\text{Area}}$$

$$= \frac{(0.3 + 0.3 + 0.1 + 0.4 + 0.9 + 0.2 + 0.2 + 0.15 + 0.1)}{(1 \times 1)}$$

$$\rho_{FL} = 2.65 \mu\text{m}^{-1}$$

Figure 63 – Calculation of fiber length density.

Finally, the calculation of fiber length and width distributions was added to GTFiber. Calculating the fiber length distribution is extremely straightforward: each fiber’s length equals the number of vectors in its contour times the vector step length. Any fiber containing a point that falls within 5% of the edge of the image is removed from the fiber length distribution to avoid edge effects. Fiber width is slightly less trivial, and is illustrated in **Figure 64**. For each fiber, 15 points are selected along its contour to sample its width. The black and white binarized image is used for this sampling. Line segments are projected perpendicular from the sampling points until they hit a black pixel, yielding a single fiber width sample. The width of that fiber is taken as the median of these samples. If a fiber contains less than 15 points, all of its points are used. Because GTFiber is optimized for

bulk fibrillar structures with large fiber populations, fiber widths are frequently on the order of <10 pixels. This means that fiber width calculation is extremely sensitive to the binarization process. For this reason, the use of GTFiber for quantitative fiber width measurement is discouraged unless fibers are more than 10 pixels wide.

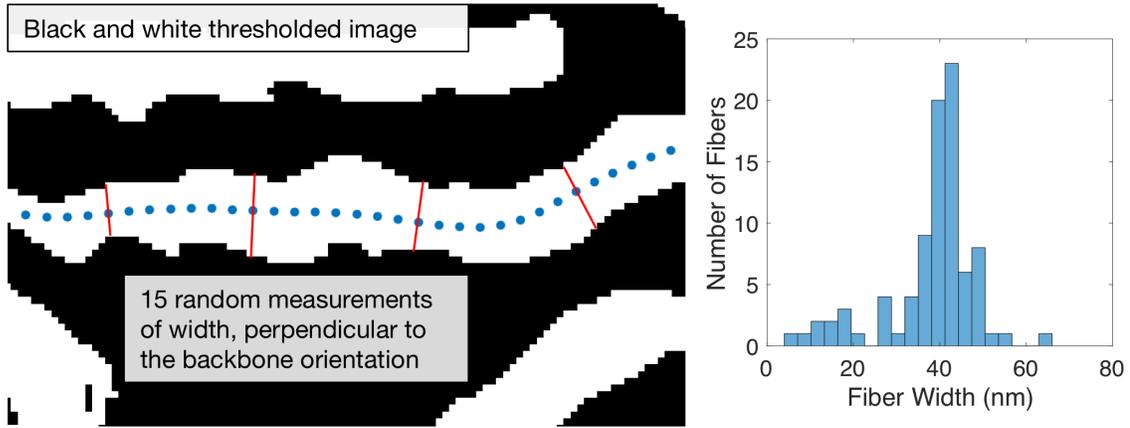


Figure 64 – Calculation of fiber width distributions.

One advantage of these order parameters is that they are defined entirely in physical units, and can thus be calculated on images of any size – but care should be taken when comparing them across different sized images. The dataset under consideration is composed of images ranging from 2 – 10 μm in width. Of these, 21 are 2 μm , 12 are 10 μm , while the remaining 104 are either 4, 5, or 7 μm . GTFiber automatically scales image processing parameters according to image size to ensure consistent segmentation across length scales, however, some measurements are inherently affected by image scale. For example, fiber length density is generally higher for smaller images because they pick up finer details that are lost in larger images. Also in smaller images, orientational order may not decay to an equilibrium value on the length scale of the image, resulting in an overestimated S_{full} . On the contrary, fiber widths are more accurately calculated in small

images where fibers are >10 pixels wide, rather than large images where they may be only a couple of pixels wide.

B.3. Accuracy and Sensitivity Analysis

The goal of GTFiber software is to provide accurate measurements of fibrillar structures. In quantifying accuracy, it is important to consider segmentation accuracy as well as measurement accuracy. In other words, did the program classify the correct pixels as fibers, and did it produce the same measurements of the structure that a human would?

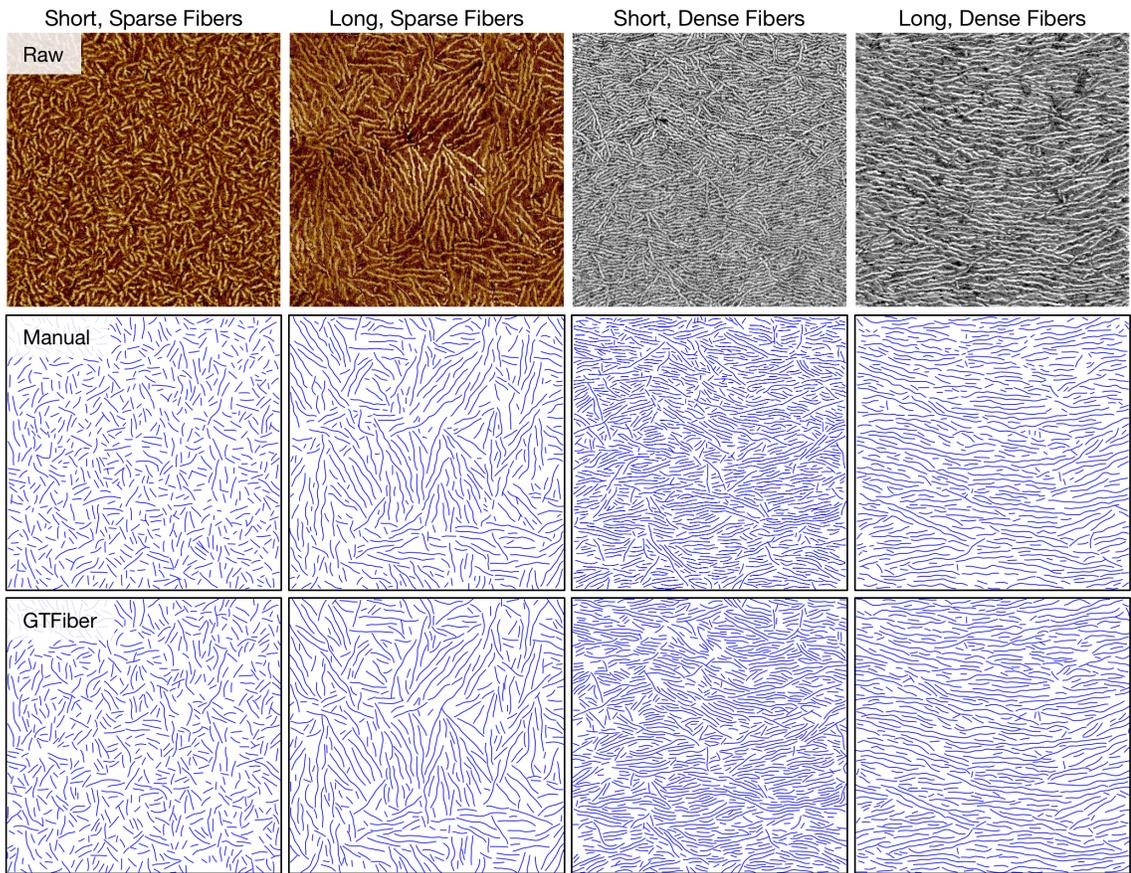
This leads naturally to a machine learning approach to parameter selection: provide a manually segmented training set (*i.e.* trace all fibers by hand), then train the image processing parameters so that they yield the same results as the manual segmentation. The problem with this approach for many researchers is that the generation of an adequate training set is a very tedious process, and by the end of it, they may already have all the analytical results they required in the first place. The philosophy behind GTFiber is that researchers should be able to quantify what they see in their images instantly without performing any manual segmentation. In general, the default parameters of GTFiber yield a reasonable segmentation result and only minor tweaking is required for further refinement. Data visualizations are provided at each step of processing so that the user can identify where errors occur and ultimately obtain a visually consistent segmentation that yields accurate measurements.

To quantify the accuracy and sensitivity of GTFiber, we performed a manual segmentation on example images and trained GTFiber's parameters to minimize the total percent error in the five main structural measurements. Optimized parameters were

obtained by gradient descent, minimizing the Euclidean distance between the structural metrics produced from a manual tracing in FiberApp and from automated extraction in GTFiber. The optimal parameters from training each individual image were averaged to obtain a universal parameter set, the results from which are presented in **Table 4**.

By visual inspection, the vectorized fiber plots look almost identical for all three cases, except possibly for the case of “short, dense fibers”. In the course of training, it was noted that GTFiber picks up on extremely small “fibers” that humans inconsistently identify. To mitigate the influence of these short fiber stubs, a “minimum fiber length” parameter was introduced – here, metrics are compared considering only the fibers exceeding 100 nm in length. Thus, while the visual segmentation results may not appear to be perfect, the structural metrics are largely accurate to within 10%, as tabulated in the continuation of **Table 4** on the page after the image results. Larger percent errors are only observed in cases where the measured value is low to begin with. The optimized parameters are also reported, which provide insight into heuristics to use for setting parameters. A guide to parameter selection is provided in **Table 6** – most parameters’ optimal settings are a simple function of the image size and the anticipated fiber width.

Table 4 – Accuracy of GTFiber in four different cases, using the same parameters, compared with manual tracing.



	Short, Sparse Fibers				Long, Sparse Fibers			
	Manual	GTFiber	Δ	%	Manual	GTFiber	Δ	%
Sfull	0.11	0.09	0.02	16.04	0.26	0.3	0.04	14.33
Decay Length (nm)	421	442	20.74	4.92	774	760	14.00	1.81
Mean Length (nm)	219	235	16.41	7.51	463	424	39.03	8.43
Median Length (nm)	203	210	7.40	3.65	382	360	21.60	5.66
Fiber Length Density (1/μm)	6.33	6.80	0.47	7.39	7.31	7.54	0.23	3.15

	Short, Dense Fibers				Long, Dense Fibers			
	Manual	GTFiber	Δ	%	Manual	GTFiber	Δ	%
Sfull	0.68	0.72	0.04	6.60	0.84	0.86	0.02	2.04
Decay Length (nm)	261	304	43.13	16.53	329	334	4.85	1.47
Mean Length (nm)	342	349	7.86	2.30	483	507	24.52	5.08
Median Length (nm)	270	270	0.00	0.00	348	330	18.00	5.17
Fiber Length Density (1/μm)	12.87	12.53	0.34	2.65	9.24	10.02	0.78	8.43

Parameters	Value	Metric	Average Error
Gaussian Smoothing (nm)	5	Sfull	0.03
Orientation Smoothing (nm)	15	Decay Length (nm)	20.68
Diffusion Time (s)	3	Mean Length (nm)	21.95
Top Hat Size (nm)	40	Median Length (nm)	11.75
Max Noise Area (nm ²)	1500	Fiber Length Density (1/ μm)	0.45
Skeletal Fringe Removal (nm)	60		
Stitch Gap Length (nm)	50		
Max Curvature (1/ μm)	7		
Vector Step Length (nm)	30		
Minimum Length (nm)	100		

Overall, however, GTFiber is relatively insensitive to parameter selection once a visually consistent segmentation has been achieved. A sensitivity analysis was performed to demonstrate this, presented in **Table 5**. Cells in this table quantify the partial derivative of each structural measurement with respect to each image processing parameter, calculated around the optimal parameter values from **Table 4**. A value of 1 indicates a 1% change in a structural metric per each 1% change in the respective parameters. A value of 0.01, then, indicates a 0.01% change in a structural metric per 1% change in the indicated parameter.

Vector step length appears to be the most influential parameter, especially in its influence on fiber length. This makes sense, as the step size also defines the minimum

measurable fiber length, which significantly affects the length distribution. Step length should be selected to be as low as possible, but should generally not be lower than 3 pixels. The mean fiber length is also affected strongly by the maximum stitching gap as well as the maximum curvature parameters, which makes sense, as increasing these parameters will favor stitching more segments together, creating a top-heavy length distribution. The remaining structural metrics have minimal sensitivity to changes in the image processing parameters.

Table 5 – Sensitivity of GTFiber parameters: percent change in each structural metric per percent change in each algorithm parameter.

	Sfull	Decay Length	Mean Length	Fiber Length Density
Gaussian Smoothing	0.077	0.057	0.010	0.073
Orientation Smoothing	0.005	0.038	0.049	0.058
Diffusion Time	0.032	0.035	0.009	0.014
Top Hat Size	0.020	0.013	0.049	0.014
Max Noise Area	0.006	0.028	0.058	0.016
Skeletal Fringe Removal	0.046	0.089	0.043	0.050
Stitch Gap Length	0.019	0.003	0.151	0.022
Max Curvature	0.054	0.004	0.393	0.031
Vector Step Length	0.054	0.208	0.907	0.048

Since human tracing is still prone to errors, another way to analyze extraction accuracy is with a known population of fibers. This could be accomplished with simulated images, or even more simply, with pictures of actual wires dispersed on a flat surface. A set of 50 images was taken of flexible white wires dispersed on black fabric, while attempting to create a variety of complex and/or ambiguous overlap situations. The images were then analyzed with GTFiber at the default parameter settings, yielding a fiber count and an average fiber length for each image. The true count was 26 fibers and the true mean length was 2.6 inches. The results are illustrated in **Figure 65**. Fiber counts and mean

lengths were generally overestimated but fell within 15% of their true value, consistent with the results of the previous optimization.

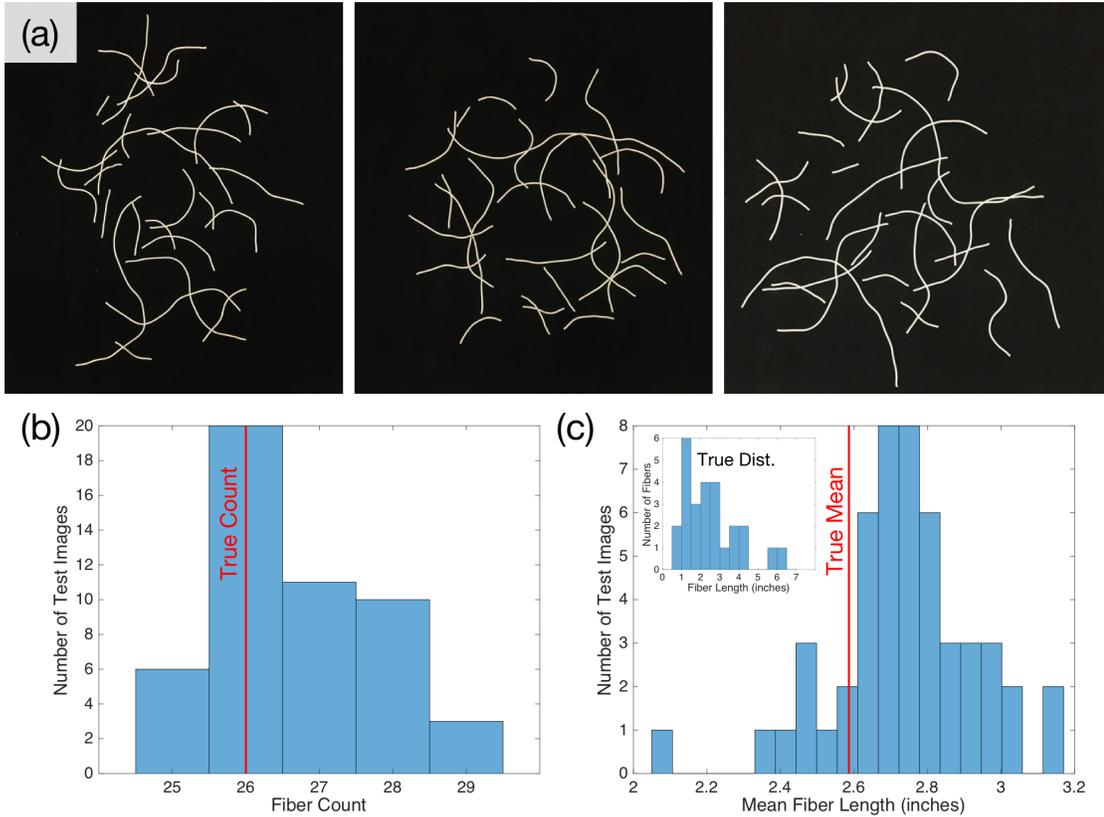


Figure 65 – (a) Representative pictures of flexible white wires on a black fabric background. (b) Histogram of fiber counts from all 50 pictures. (c) Histogram of mean fiber lengths from all 50 pictures; (inset) true fiber length distribution of the wire population.

Fiber extraction and measurement from images is a challenging problem in computer vision. Each new set of images poses a different set of challenges, and may not be amenable to the current approach. Images that create problems for GTFiber are highlighted in **Figure 66**.

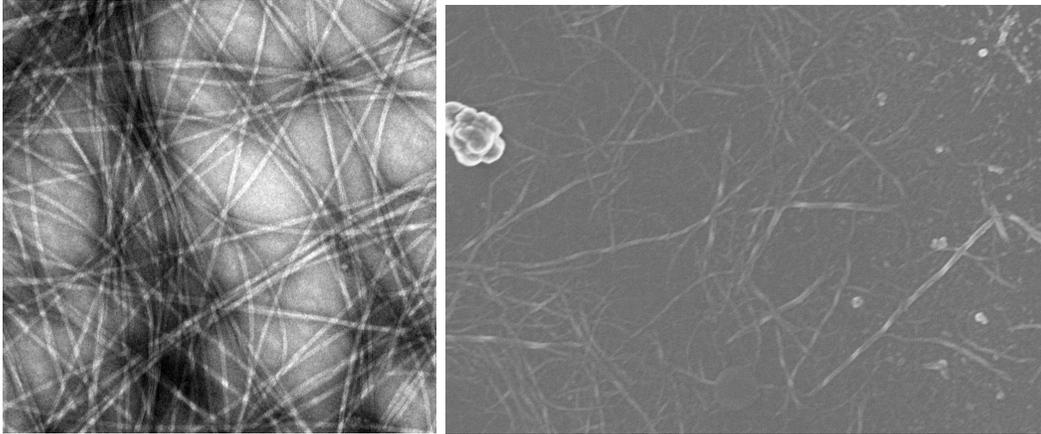


Figure 66 – Problematic cases for GTFiber.

These images exemplify some of the most difficult cases in fiber extraction. Indeed, even a human would be hard-pressed to definitively identify contiguous fibers in these images. In the image on the left, the bodies of the fibers are not significantly brighter than the background and are instead defined by the dark clouds around their edges. Thick fiber bundles create dark regions with high degrees of overlap that are nearly impossible to disentangle by eye. In the image on the right, a combination of large variations in fiber curvature and brightness, dense, overlapping clumps and extremely low contrast cause significant difficulty. To solve this problem in the general case, it is likely that a generative adversarial convolutional neural net approach would be required, combining both extensive simulations of fiber images as well as a significant number of accurately hand-labeled images capturing a wide variety of fiber populations, imaging techniques, and noise.^{153,190}

B.4. User Interface and Parameter Selection

Download software: [gtfiber.github.io](https://github.com/gtfiber/gtfiber.io)

The user interface is designed to produce measurements as quickly as possible, while providing visualizations that show what is happening at every step during processing.

To analyze an image with the default settings, follow these steps:

- 1) File -> Load Image
- 2) Enter image width in nanometers when prompted
- 3) Click “Run Filter” and wait for progress bar to complete
- 4) Click “Stitch Fibers” and wait for progress bar to complete
- 5) Click on any of the plotting and visualization options

If the skeletonized image that appears after “Run Filter” matches your original well, fiber stitching should be successful. If not, see the guide below for further parameter tweaking.

Bear in mind that not all images of fibers are amenable to this approach (see **Figure 66**).

Feel free to email the author for help with installation or usage.

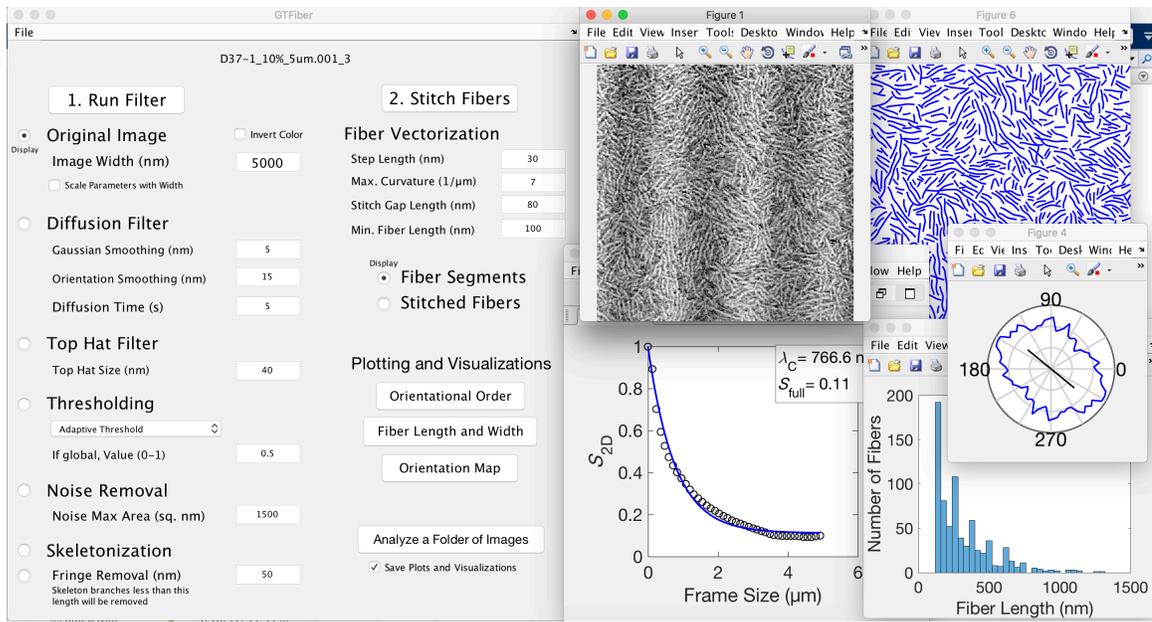


Figure 67 – GTFiber user interface and representative visualizations.

Table 6 – Guide to parameter selection and tuning.

Parameter	Recommended Value	How to Troubleshoot
Gaussian Smoothing (nm)	Half fiber width	If fibers are wide, turn up.
Orientation Smoothing (nm)	Fiber width	Keep as low as possible. Only increase if fiber edges are very jagged
Diffusion Time (s)	3–7 seconds	If thresholding results in holes inside fibers, turn up. If fibers are being run together, turn down.
Top Hat Size (nm)	2x fiber width	If wider fibers are getting lost, turn this up. Click between displaying Diffusion Filter result and Top Hat result to see the effect.
Max Noise Area (nm ²)	Area of shortest fiber (fiber width x min. length) (~1500 nm ² for a 5 x 5 μm image).	Click between displaying the Thresholding result and the Noise Removal result to see what this is filtering out. If adaptive thresholding is bringing out a lot of noise, try a Global Threshold with a value between 0.4 and 0.6.
Skeletal Fringe Removal (nm)	2x fiber width	Click between displaying Skeleton and Fringe Removal to see what this removes. If the skeleton still has lots of “spurs” or “fringes” after Fringe Removal, turn this up.
Stitch Gap Length (nm)	3x fiber width	After running “Stitch Fibers”, click between displaying Fiber Segments and Stitched Fibers. If segments are being combined that are too far apart, turn this down.
Max Curvature (μm ⁻¹)	10	If stitched fibers have unrealistic bends, turn down. If segments are not being stitched even though they should be, try turning up.
Vector Step Length (nm)	2x fiber width	For a 5μm, 512x512 image, 30 nm (3 pixels, or 0.6% of image width) was ideal. For higher resolution images, turn up.
Min. Fiber Length (nm)	No recommendation	Start at 0, turn up until small “nuisance” fibers are gone.

APPENDIX C. The OFET Database

C.1. Databases Drive Reproducibility

Academia is in the midst of a so-called reproducibility crisis.^{191,192} While this has been publicized largely in the life sciences, it is nonetheless a concern for research in nanomaterials.¹⁹³ Experimental systems have grown in complexity, data has grown in volume and velocity, and analysis is now performed by multiple layers of computing systems, resulting in published figures that are very far removed from the raw data used to generate them. Single words in a manuscript can now represent thousands of lines of computer code or a highly sensitive experimental method developed over multiple studies. Therein lies the heart of the reproducibility crisis: in the amount of information assumed to have been conveyed in the words of a text document.

Journals and research consortia have combatted this problem by setting up centralized data repositories for research methods, data, and analytical codes.¹⁹⁴ For example, in the field of Proteomics, the PRIDE Archive provides an explorable repository of mass spectrometry data from proteomics research, promoting transparency through the entire data analysis process.¹⁹⁵ Organic synthesis has benefitted from searchable databases like Scifinder[®] or Reaxys[®]. Material property simulations have benefitted more recently from the establishment of the Materials Genome Initiative and online collaborative organizations like nanoHUB.^{76,196} It is recognized that the volume and velocity of materials research data will soon reach a threshold across which large data repositories and tools from data science will be required for management and analysis.¹⁹⁷ The same will undoubtedly be true for nanomaterials research.

Exploratory nanomaterials research data presents a greater challenge due to the variability of types of information presented and the speed with which reporting practices change. This is not so different from an auction website like eBay, where sellers attempt to provide as detailed a description of their item as possible, but the relevant descriptive features of a given item are determined more or less by the marketplace, which is constantly changing. eBay can give sellers suggestions on what information to provide based on the listings of similar items. For example, if one were to sell a wristwatch, eBay suggests that the seller specify the band material and the movement mechanism based on the listings of similar items. In a similar vein, the Nature publishing group recently introduced an initiative to provide authors suggestions of information to include in methods sections.¹⁹³ Information like instrument settings and environmental conditions, referred to as metadata, can have a huge influence on measured material performance and is a vital part of any data repository. We aim to demonstrate the value of cataloging experimental metadata using the system of the solution-processed polymeric transistor as a case study.

Solution processable polymeric semiconductors are a promising class of electronic material for a variety of reasons: they can be coated on large areas, are mechanically flexible, and could enable high throughput additive manufacturing techniques.^{11,198–201} Commercial applications include solar energy, display, and sensing technologies.^{5,6,9,10,15} For almost two decades, the organic field effect transistor (OFET) has been used as a platform to study the process-structure-property relationships in polymeric semiconductors, because it is relevant to various applications, a facile platform for thin film deposition, and enables the calculation of charge carrier mobility (mobility), an important figure of merit for the performance of a given device.⁸⁴ Mobility is essentially a

model parameter defined as the velocity of a charge per applied electric field in a material sample; it has units of $cm^2 V^{-1} s^{-1}$ and is very sensitive to starting material, processing, device layout, and calculation method. Poly(3-hexylthiophene) (P3HT) is frequently used as a model material in studies of the effects of processing conditions on OFET mobility. We compile data from over 200 P3HT-based OFET devices published across the literature to envision what an experimental research database would look like for a nanomaterial system and how it could be used to generate physical insight as well as improve the reproducibility of experimental studies.

C.2. Database Curation

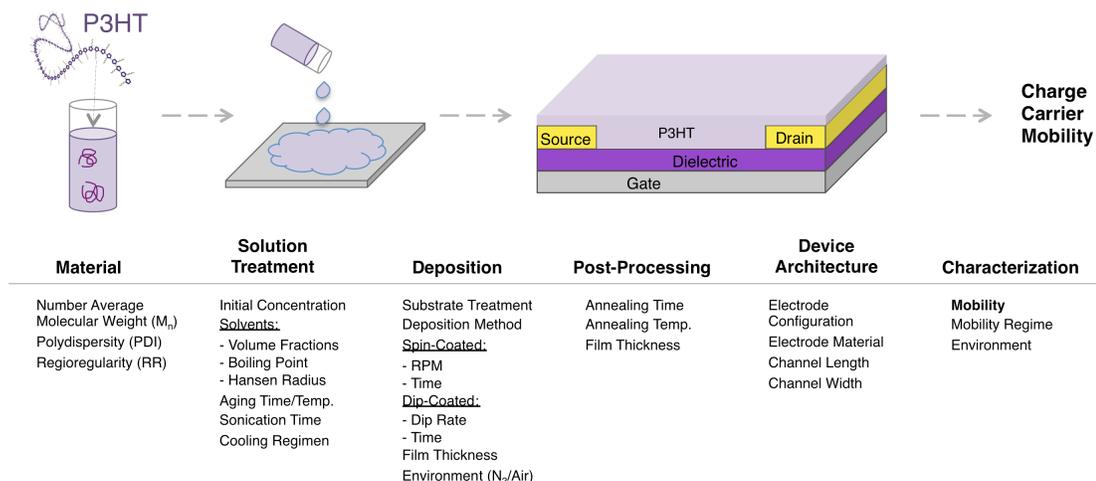


Figure 68 – Diagram of the process of fabricating and characterizing a P3HT-based OFET device. Below: list of all process variables included in our database.

Publications were selected for the database based on the following criteria: (1) the study must involve OFET devices whose active layer was comprised of neat P3HT, (2) mobility must be reported at room temperature, and (3) a majority of the relevant process and device parameters must be reported.^{26–28,35,40,49,80–82,85–87,89,90,202–206} The parameters that were recorded in the database are enumerated in Figure 68. They comprise everything from the molecular parameters of the starting material, to the machine settings of the various deposition techniques, to the analytical methods used to calculate mobility from the current-voltage measurements on the device.

Very few papers contained the complete set of desired processing information; the dataset is not sparse by any means, but missing information is a significant problem. Choi et al. discuss this further in an article on the best practices for reporting on OFET device

fabrication.²⁰⁷ The dataset is also wildly diverse, comprised of text-based designators (e.g. *bottom-gate*, *bottom-contact architecture* or *spin-coated...*), chemical compounds (*chloroform*, *chlorobenzene*, *acetone...*), numeric values (*5 mg/mL*, *40 kD...*), numeric ranges (*20 – 50 nm*), and even categories with mixed types (regioregularity is listed variously as a percentage or simply “*highly regioregular*”).

Data entry was performed in Excel, with each study receiving its own self-contained spreadsheet to handle all of this information. Mobility, typically presented in a graph, was extracted from figures using DataThief [www.datathief.org], a semi-automatic tool that can reliably extract quantitative data from scatter plots up to 3 significant figures. MATLAB scripts were written to compile the database into a structure array, filter and search the devices, and plot the results. For example, a one-line function call can find all devices made with chloroform as a solvent on a bottom-gate, bottom-contact architecture with gold electrodes, and plot the mobility of the filtered devices against their molecular weight. The entire database with accompanying graphical user interface is available for download from a GitHub repository at [<http://www.github.com/Imperssonator/OFET-Database>]. We now present select examples of the trends and insight that can be extracted from this database.

C.3. Results

C.3.1 Mobility versus Molecular Weight Distribution

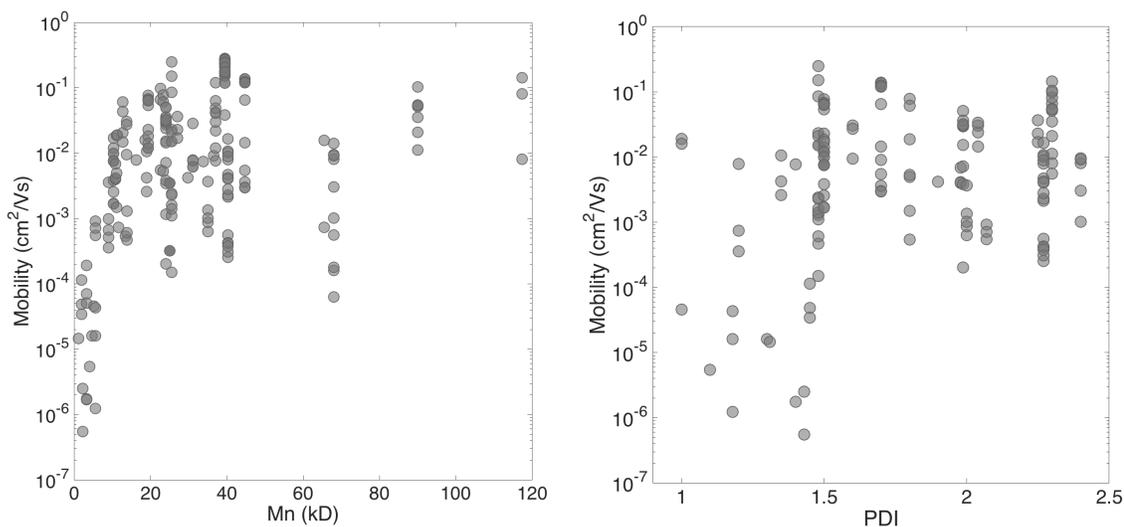


Figure 69 – (A) Hole mobility versus M_n (in kilodaltons) for 218 devices across the literature. Note the log scale on the y-axis. (B) Mobility versus polydispersity index for a subset (146) of the devices in (A). Markers are semi-transparent so that high data densities appear darker.

The first widely reported process-property trend in the literature is that of increasing mobility with molecular weight, plotted in **Figure 69A**. Kline et al., Zen et al., and Verilhac et al. conducted the seminal studies on this parameter, concluding variously that low molecular weights ($< 20,000 \text{ g mol}^{-1}$) introduce inherent crystalline disorder that limits their overall performance, while higher molecular weights may introduce chains that are long enough to extend across grain boundaries, providing high-mobility pathways for charge transport.^{26,35,36,82,204}

Polydispersity index (PDI) is more difficult to study as a process parameter, yet our dataset contains 146 devices for which PDI has been reported. It displays approximately

the same trend as molecular weight, as shown in **Figure 69B**, with devices with the lowest PDI showing a wider spread that encompasses very low performance devices ($< 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$). It is possible that this is due to a correlation between PDI and molecular weight, however this correlation is quite weak ($R^2 = 0.4$ from a linear fit), so there is some merit to the notion that higher Polydispersity index can benefit device performance. This could be explained on the same morphological grounds as the trend in molecular weight: highly polydisperse samples would contain many chains long enough to connect multiple crystalline grains. Given the difficulty of controlling molecular weight distribution accurately, it is all the more important that it be presented in as quantitative a manner as possible, because it is highly unlikely that authors from different groups will attain the same distribution.

C.3.2 Mobility versus Channel Length

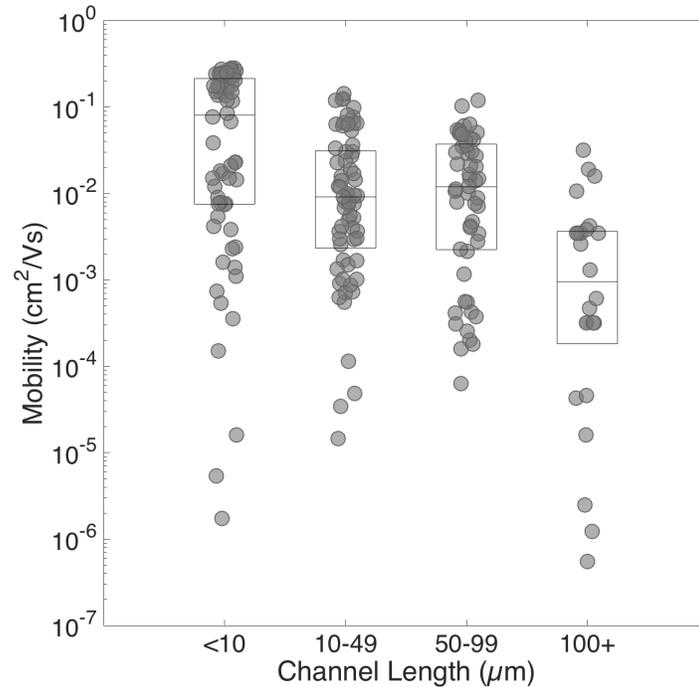


Figure 70 – Mobility versus groups of channel lengths as reported for 206 devices in the literature. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.

Channel length is a widely investigated device fabrication parameter due to its direct relationship with transistor packing density and power efficiency in electronic applications. Wang et al. and Chang et al. investigated the impact of channel length on mobility and discovered that shorter channels generally have higher mobility.^{25,202} Chang associated this trend with a decrease in the number of grain boundaries as the length scale of the channel approaches that of the polymeric crystalline domains. However, contributions from increased contact resistance and deviation from the gradual channel approximation were ignored. Chabynyc et al. demonstrated that field effect transistors

below 10 μm in length experience high contact resistance and deviate from the ideal current-voltage relationships shown above.²⁰⁸

Our database indicates that both of the above referenced analyses are correct. The box plot in **Figure 70** shows mobility versus four groups of channel lengths: median device performance for short channel devices is approximately one order of magnitude higher than devices with channels between 10 and 100 μm , and two orders of magnitude higher than devices with channels 100 μm and longer. The devices with the highest reported mobilities in the database have channels of length 5 μm and use an unconventional architecture with platinum electrodes.⁸⁶ Channel length will be revisited in a later section as we attempt to identify a standard device.

C.3.3 Mobility versus Deposition Method

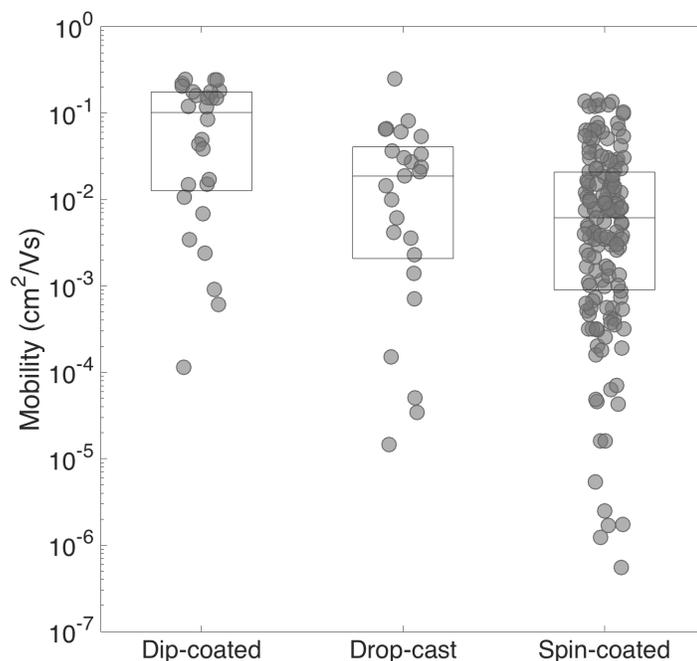


Figure 71 – Mobility versus solution deposition method for 204 devices in the literature: dip-coated, drop-cast, or spin-coated. Only devices with gold electrodes are considered in this search. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.

The three predominant methods of film deposition in studies of P3HT OFETs are dip-coating, drop-casting, or spin-coating, with spin-coating comprising the overwhelming majority of devices. This is probably due to its ability to create extraordinarily uniform thin films without concerns of dewetting or surface tension effects. However, as illustrated in **Figure 71**, the data indicate that spin-coating is not only the worst thin film deposition technique in terms of median, minimum and maximum mobility, but also the most variable. Surin, who conducted a study comprised of devices coated using all three of the above methods, postulated that for dip-coating and drop-casting, the evaporation of solvent is so

slow that films obtain a crystalline state closer to equilibrium than that of spin-coated films.²⁸ Thus, a broader range of kinetically trapped morphologies can be obtained by spin-coating. Aiyar, Chang and Choi each arrived at similar conclusions in studying solvent effects for spin-coated devices.^{80,81,89}

The spread of mobilities obtained from spin-coated devices occupies six decades of performance with a long tail of devices whose reported mobility descends as low as $10^{-6} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. The variability induced by the spin-coating process is also reflected in the rather large error bars regularly encountered with mobilities averaged over multiple devices spin-coated from the same solution. Variability in thin film deposition is troubling given that structural characterization, other than scanning probe microscopy (UV-Vis, X-Ray diffraction and scattering), is sometimes carried out on films deposited on separate substrates from the devices used for mobility measurement. It is a slight relief that spin-coating is unlikely to be used in an industrial process. As research moves to more relevant additive processing methods, device performance may also naturally increase.

C.3.4 Identifying a Standard Device Platform

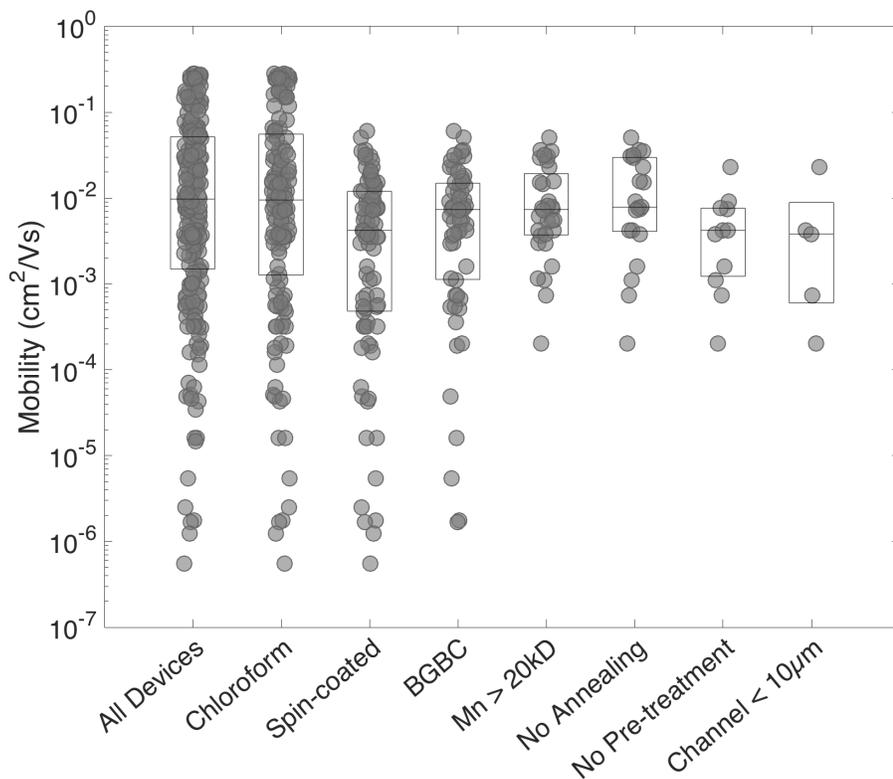


Figure 72 – Reported mobility of devices satisfying progressively tighter constraints. The constraint in each column stacks on all the previous constraints. Boxes indicate 25th percentile, median, and 75th percentile. Markers are semi-transparent so that high data densities appear darker.

Beyond what is presented above, global process-property trends are difficult to extract. A more fruitful endeavor may be to filter the database to isolate a standard device that is reported by many different authors, and examine what differences can be observed among that limited set. From an extensive perusal using our search functions, we have identified a set of processing conditions that appear relatively frequently: (1) Neat chloroform as a solvent, (2) spin-coated film, (3) bottom-gate, bottom-contact architecture (BGBC) with gold electrodes, (4) number average molecular weight > 20 kD, (5) no

annealing, and (6) no sonication, aging or other solution pre-treatment. Following the discussion in Section 3.2, we should also (7) ignore devices with channel lengths less than 10 μm . **Figure 72** illustrates how this progression of filters reduces our set down to five devices with a narrower range of mobilities, all from different authors.^{49,81,82,90,205} The full table of relevant processing conditions is reported for these devices in **Table 7**.

Table 7 – Processing conditions for 5 devices conforming to the standard defined above. ^a Regioregularity. ^b Initial concentration of P3HT in solution. ^c Hexamethyldisilazane.

Author/Year	Aiyar 2011	Bielecka 2011	Chang 2013	Park 2014	Verilhac 2006
M_n (kD)	24		40.3	24	27
M_w (kD)	47.7	65.5	91.481	47.7	60.75
PDI	1.9875		2.27	1.9875	2.25
R.R. ^a (%)	93	96.6	92	92	98
Solvent	CHCl_3	CHCl_3	CHCl_3	CHCl_3	CHCl_3
Init. Conc. ^b (mg/mL)	4	10	5	3	2
Substrate Treatment					HMDS ^c
Deposition Method	Spin-coated	Spin-coated	Spin-coated	Spin-coated	Spin-coated
Spin Rate (rpm)	1500	900	1500	2000	300
Spin Time (s)			60	60	30
Processing Environment	Air	Air	Air	N_2	Air
Mobility Environment	Air	Vacuum	N_2	Vacuum	N_2
Mobility Regime	Linear	Saturation	Saturation	Linear	Saturation
Electrode Configuration	BGBC	BGBC	BGBC	BGBC	BGBC
Electrode Material	Au	Au	Au	Au	Au
Channel Length (μm)	50	10	50	200	20
Channel Width (mm)	2	10	2	0.5	9
Mobility (cm^2/Vs)	0.000202	0.00073	0.00423	0.0038	0.0229

The devices in **Table 7** are presented in order of increasing mobility. Even for this tightly defined set of conditions, reported mobility spans two orders of magnitude. Subtle differences in the preparation and characterization of these devices can help elucidate the remaining sources of variance. The devices from Aiyar and Park are a particularly

instructive case: both are prepared from P3HT purchased from Sigma-Aldrich, with similar molecular weights.^{90,205} The only major difference is in their channel length (50 vs. 200 μm) and their mobility characterization environment (air vs. vacuum). Since the difference in channel length would suggest that Park's mobility be lower, it is the difference in characterization environment that apparently makes the largest difference, affecting an improvement of over an order of magnitude. The highest performing device of the set – that of Verilhac et al. – has two unique features: an HMDS-treated substrate and a remarkably low spin rate (300 *rpm*). Surface modification has been used in multiple studies to induce favorable edge-on crystalline orientation and reduce energetic traps at the dielectric interface.^{209,210} Following the discussion in Section 3.3, a low spin rate may also reduce the solvent evaporation rate, allowing more time for crystallization during deposition.

C.4. Commentary on Materials Processing Databases

Perhaps what is most remarkable about this dataset is that there are no two devices fabricated at the exact same conditions, yet the information needed to explain discrepancies between the closest devices is available and validates much of what has been learned up to this point. The ability to quickly explore and compare devices across this large set of data could yield useful information to many interested parties, which is why we have made it publicly accessible [<http://www.github.com/Imperssonator/OFET-Database>]. We have validated existing theories about the effects of molecular weight distribution, channel length, and deposition method, as well as identified a standard device that highlights the large discrepancies in performance measurement that still must be overcome. This was accomplished using only published process-property data – to say nothing of the even greater opportunity afforded by cataloging structure-property data.

While a database such as this is probably not a rigorous tool for knowledge extraction, it can be very helpful in hypothesis generation and experimental design. The main challenge with analysis using the database is that of imputation: the method used to fill in missing information. Machine learning and data mining methods surely exist for tackling this problem, either by automatically modeling missing information, or by automatically soliciting missing information from authors. Above all, a database-driven approach to experimental research must be flexible and able to rapidly adapt to the information that researchers provide and seek out. With the adoption of cutting-edge tools from the field of data science, experimental research, analysis and reproducibility will flourish as publishing enters a new age of accessible, interactive data.

Chapter 6. REFERENCES

1. Rughooputh, S. D. D. V.; Nowak, M.; Hotta, S.; Heeger, A. J.; Wudl, F. Soluble Conducting Polymers: The poly(3-Alkylthienylenes). *Synth. Met.* **1987**, *21*, 41–50.
2. Assadi, A.; Svensson, C.; Willander, M.; Inganäs, O. Field-effect Mobility of poly(3-hexylthiophene). *Appl. Phys. Lett.* **1988**, *53*, 195–197.
3. Krebs, F. C.; Gevorgyan, S. a.; Alstrup, J. A Roll-to-Roll Process to Flexible Polymer Solar Cells: Model Studies, Manufacture and Operational Stability Studies. *J. Mater. Chem.* **2009**, *19*, 5442.
4. Søndergaard, R.; Hösel, M.; Angmo, D.; Larsen-Olsen, T. T.; Krebs, F. C. Roll-to-Roll Fabrication of Polymer Solar Cells. *Mater. Today* **2012**, *15*, 36–49.
5. Günes, S.; Neugebauer, H.; Sariciftci, N. S. Conjugated Polymer-Based Organic Solar Cells. *Chem. Rev.* **2007**, *107*, 1324–1338.
6. Li, G.; Zhu, R.; Yang, Y. Polymer Solar Cells. *Nat. Photonics* **2012**, *6*, 153–161.
7. Burns, S. E.; Reynolds, K.; Reeves, W.; Banach, M.; Brown, T.; Chalmers, K.; Cousins, N.; Etchells, M.; Hayton, C.; Jacobs, K.; Menon, A.; Siddique, S.; Too, P.; Ramsdale, C.; Watts, J.; Cain, P.; von Werne, T.; Mills, J.; Curling, C.; Siringhaus, H.; Amundson, K.; McCreary, M. D. A Scalable Manufacturing Process for Flexible Active-Matrix E-Paper Displays. *J. Soc. Inf. Disp.* **2005**, *13*, 583.
8. Huitema, H. E. A.; Gelinck, G. H.; van der Putten, J. B. P. H.; Kuijk, K. E.; Hart, C. M.; Cantatore, E.; Herwig, P. T.; van Breemen, A. J. J. M.; de Leeuw, D. M. Plastic Transistors in Active-Matrix Displays. *Nature* **2001**, *414*, 599.
9. Gelinck, G. H.; Huitema, H. E. a; van Veenendaal, E.; Cantatore, E.; Schrijnemakers, L.; van der Putten, J. B. P. H.; Geuns, T. C. T.; Beenhakkers, M.; Giesbers, J. B.; Huisman, B.-H.; Meijer, E. J.; Benito, E. M.; Touwslager, F. J.; Marsman, A. W.; van Rens, B. J. E.; de Leeuw, D. M. Flexible Active-Matrix Displays and Shift Registers Based on Solution-Processed Organic Transistors. *Nat. Mater.* **2004**, *3*, 106–110.

10. Sekitani, T.; Nakajima, H.; Maeda, H.; Fukushima, T.; Aida, T.; Hata, K.; Someya, T. Stretchable Active-Matrix Organic Light-Emitting Diode Display Using Printable Elastic Conductors. *Nat. Mater.* **2009**, *8*, 494–499.
11. Zheng, H.; Zheng, Y.; Liu, N.; Ai, N.; Wang, Q.; Wu, S.; Zhou, J.; Hu, D.; Yu, S.; Han, S.; Xu, W.; Luo, C.; Meng, Y.; Jiang, Z.; Chen, Y.; Li, D.; Huang, F.; Wang, J.; Peng, J.; Cao, Y. All-Solution Processed Polymer Light-Emitting Diode Displays. *Nat. Commun.* **2013**, *4*, 1971.
12. Majewski, L. a.; Balocco, C.; King, R.; Whitelegg, S.; Song, a. M. Fast Polymer Nanorectifiers for Inductively Coupled RFID Tags. *Mater. Sci. Eng. B Solid-State Mater. Adv. Technol.* **2008**, *147*, 289–292.
13. Arias, A. C.; MacKenzie, J. D.; McCulloch, I.; Rivnay, J.; Salleo, A. Materials and Applications for Large Area Electronics: Solution-Based Approaches. *Chem. Rev.* **2010**, *110*, 3–24.
14. Knopfmacher, O.; Hammock, M. L.; Appleton, A. L.; Schwartz, G.; Mei, J.; Lei, T.; Pei, J.; Bao, Z. Highly Stable Organic Polymer Field-Effect Transistor Sensor for Selective Detection in the Marine Environment. *Nat. Commun.* **2014**, *5*, 2954.
15. Schwartz, G.; Tee, B. C.-K.; Mei, J.; Appleton, A. L.; Kim, D. H.; Wang, H.; Bao, Z. Flexible Polymer Transistors with High Pressure Sensitivity for Application in Electronic Skin and Health Monitoring. *Nat. Commun.* **2013**, *4*, 1859.
16. Lu, N.; Kim, D.-H. Flexible and Stretchable Electronics Paving the Way for Soft Robotics. *Soft Robot.* **2014**, *1*, 53–62.
17. Bauer, S.; Bauer-Gogonea, S.; Graz, I.; Kaltenbrunner, M.; Keplinger, C.; Schw??diauer, R. 25th Anniversary Article: A Soft Future: From Robots and Sensor Skin to Energy Harvesters. *Adv. Mater.* **2014**, *26*, 149–162.
18. Sirringhaus, H. 25th Anniversary Article: Organic Field-Effect Transistors: The Path Beyond Amorphous Silicon. *Adv. Mater.* **2014**, *26*, 1319–1335.
19. Tiwari, S.; Greenham, N. C. Charge Mobility Measurement Techniques in Organic Semiconductors. *Opt. Quantum Electron.* **2009**, *41*, 69–89.
20. Salleo, A. Charge Transport in Polymeric Transistors. *Mater. Today* **2007**, *10*, 38–

45.

21. Mandoc, M. M.; Koster, L. J. a; Blom, P. W. M. Optimum Charge Carrier Mobility in Organic Solar Cells. *Appl. Phys. Lett.* **2007**, *90*, 23–25.
22. Kwiatkowski, J. J. From Molecules to Mobilities: Modelling Charge Transport in Organic Semiconductors, 2008, Vol. PhD Thesis.
23. Mei, J.; Diao, Y.; Appleton, A. L.; Fang, L.; Bao, Z. Integrated Materials Design of Organic Semiconductors for Field-Effect Transistors. *J. Am. Chem. Soc.* **2013**, *135*, 6724–6746.
24. Usta, H.; Facchetti, A. Polymeric and Small-Molecule Semiconductors for Organic Field-Effect Transistors. In *Large Area and Flexible Electronics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2015; pp. 1–100.
25. Wang, G.; Swensen, J.; Moses, D.; Heeger, A. J. Increased Mobility from Regioregular poly(3-Hexylthiophene) Field-Effect Transistors. *J. Appl. Phys.* **2003**, *93*, 6137.
26. Kline, R. J.; McGehee, M. D.; Kadnikova, E. N.; Liu, J.; Fréchet, J. M. J. Controlling the Field-Effect Mobility of Regioregular Polythiophene by Changing the Molecular Weight. *Adv. Mater.* **2003**, *15*, 1519–1522.
27. Chang, J.-F.; Sun, B.; Breiby, D. W.; Nielsen, M. M.; Sölling, T. I.; Giles, M.; McCulloch, I.; Sirringhaus, H. Enhanced Mobility of Poly(3-Hexylthiophene) Transistors by Spin-Coating from High-Boiling-Point Solvents. *Chem. Mater.* **2004**, *16*, 4772–4776.
28. Surin, M.; Leclère, P.; Lazzaroni, R.; Yuen, J. D.; Wang, G.; Moses, D.; Heeger, A. J.; Cho, S.; Lee, K. Relationship between the Microscopic Morphology and the Charge Transport Properties in poly(3-Hexylthiophene) Field-Effect Transistors. *J. Appl. Phys.* **2006**, *100*, 33712.
29. 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.
30. Yan, H.; Chen, Z.; Zheng, Y.; Newman, C.; Quinn, J. R.; Dötz, F.; Kastler, M.;

- Facchetti, A. A High-Mobility Electron-Transporting Polymer for Printed Transistors. *Nature* **2009**, *457*, 679–686.
31. Tessler, N.; Preezant, Y.; Rappaport, N.; Roichman, Y. Charge Transport in Disordered Organic Materials and Its Relevance to Thin-Film Devices: A Tutorial Review. *Adv. Mater.* **2009**, *21*, 2741–2761.
 32. Malik, S.; Nandi, A. K. Crystallization Mechanism of Regioregular poly(3-Alkyl Thiophene)s. *J. Polym. Sci. Part B Polym. Phys.* **2002**, *40*, 2073–2085.
 33. Lim, J. A.; Liu, F.; Ferdous, S.; Muthukumar, M.; Briseno, A. L. Polymer Semiconductor Crystals. *Mater. Today* **2010**, *13*, 14–24.
 34. Lan, Y.-K.; Huang, C.-I. Charge Mobility and Transport Behavior in the Ordered and Disordered States of the Regioregular poly(3-Hexylthiophene). *J. Phys. Chem. B* **2009**, *113*, 14555–14564.
 35. Kline, R. J.; McGehee, M. D.; Kadnikova, E. N.; Liu, J.; Fréchet, J. M. J.; Toney, M. F. Dependence of Regioregular Poly(3-Hexylthiophene) Film Morphology and Field-Effect Mobility on Molecular Weight. *Macromolecules* **2005**, *38*, 3312–3319.
 36. Noriega, R.; Rivnay, J.; Vandewal, K.; Koch, F. P. V; 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.
 37. Chen, T.-A.; Wu, X.; Rieke, R. D. Regiocontrolled Synthesis of Poly (3-Alkylthiophenes) Mediated by Rieke Zinc: Their Characterization and Solid-State Properties. *J. Am. Chem. Soc.* **1995**, *117*, 233–244.
 38. Oh, J. Y.; Shin, M.; Lee, T. II; Jang, W. S.; Min, Y.; Myoung, J.; Baik, H. K.; Jeong, U. Self-Seeded Growth of Poly(3-Hexylthiophene) (P3HT) Nanofibrils by a Cycle of Cooling and Heating in Solutions. *Macromolecules* **2012**, *45*, 7504–7513.
 39. Fu, Y.; Lin, C.; Tsai, F. Y. High Field-Effect Mobility from poly(3-Hexylthiophene) Thin-Film Transistors by Solvent-Vapor-Induced Reflow. *Org. Electron. physics, Mater. Appl.* **2009**, *10*, 883–888.
 40. Newbloom, G. M.; Kim, F. S.; Jenekhe, S. A.; Pozzo, D. C. Mesoscale Morphology and Charge Transport in Colloidal Networks of Poly(3-Hexylthiophene).

Macromolecules **2011**, *44*, 3801–3809.

41. Sun, S.; Salim, T.; Wong, L. H.; Foo, Y. L.; Boey, F.; Lam, Y. M. A New Insight into Controlling poly(3-Hexylthiophene) Nanofiber Growth through a Mixed-Solvent Approach for Organic Photovoltaics Applications. *J. Mater. Chem.* **2011**, *21*, 377–386.
42. Lobov, G. S.; Zhao, Y.; Marinins, A.; Yan, M.; Li, J.; Sugunan, A.; Thylén, L.; Wosinski, L.; Östling, M.; Toprak, M. S.; Popov, S. Dynamic Manipulation of Optical Anisotropy of Suspended Poly-3-Hexylthiophene Nanofibers. *Adv. Opt. Mater.* **2016**, *4*, 1651–1656.
43. Jeong, Y. J.; Lee, H.; Lee, B. S.; Park, S.; Yudistira, H. T.; Choong, C. L.; Park, J. J.; Park, C. E.; Byun, D. Directly Drawn poly(3-Hexylthiophene) Field-Effect Transistors by Electrohydrodynamic Jet Printing: Improving Performance with Surface Modification. *ACS Appl. Mater. Interfaces* **2014**, *6*, 10736–10743.
44. Lim, H.; Ho, C.-C.; Wu, S.-J.; Tsai, H.-C.; Su, W.-F.; Chao, C.-Y. A poly(3-Hexylthiophene) Block Copolymer with Macroscopically Aligned Hierarchical Nanostructure Induced by Mechanical Rubbing. *Chem. Commun.* **2013**, *49*, 9146.
45. Na, J. Y.; Kang, B.; Sin, D. H.; Cho, K.; Park, Y. D. Understanding Solidification of Polythiophene Thin Films during Spin-Coating: Effects of Spin-Coating Time and Processing Additives. *Sci. Rep.* **2015**, *5*, 13288.
46. Singh, K. a.; Young, T.; McCullough, R. D.; Kowalewski, T.; Porter, L. M. Planarization of Polymeric Field-Effect Transistors: Improvement of Nanomorphology and Enhancement of Electrical Performance. *Adv. Funct. Mater.* **2010**, *20*, 2216–2221.
47. Wang, G.; Hirasa, T.; Moses, D.; Heeger, A. J. Fabrication of Regioregular poly(3-Hexylthiophene) Field-Effect Transistors by Dip-Coating. *Synth. Met.* **2004**, *146*, 127–132.
48. Chu, P.-H.; Kleinhenz, N.; Persson, N.; McBride, M.; Hernandez, J. L.; Fu, B.; Zhang, G.; Reichmanis, E. Toward Precision Control of Nanofiber Orientation in Conjugated Polymer Thin Films: Impact on Charge Transport. *Chem. Mater.* **2016**, *28*, 9099–9109.
49. Bielecka, U.; Lutsyk, P.; Janus, K.; Sworakowski, J.; Bartkowiak, W. Effect of

- Solution Aging on Morphology and Electrical Characteristics of Regioregular P3HT FETs Fabricated by Spin Coating and Spray Coating. *Org. Electron.* **2011**, *12*, 1768–1776.
50. Brinkmann, M. Structure and Morphology Control in Thin Films of Regioregular poly(3-Hexylthiophene). *J. Polym. Sci. Part B Polym. Phys.* **2011**, *49*, 1218–1233.
 51. Kim, Y.; Cook, S.; Tuladhar, S. M.; Choulis, S. a.; Nelson, J.; Durrant, J. R.; Bradley, D. D. C.; Giles, M.; McCulloch, I.; Ha, C.-S.; Ree, M. A Strong Regioregularity Effect in Self-Organizing Conjugated Polymer Films and High-Efficiency Polythiophene:fullerene Solar Cells. *Nat. Mater.* **2006**, *5*, 197–203.
 52. Persson, N.; McBride, M.; Grover, M.; Reichmanis, E. Silicon Valley Meets the Ivory Tower: Searchable Data Repositories for Experimental Nanomaterials Research. *Curr. Opin. Solid State Mater. Sci.* **2016**, *20*, 338–343.
 53. Niles, E. T.; Roehling, J. D.; Yamagata, H.; Wise, A. J.; Spano, F. C.; Moulé, A. J.; Grey, J. K. J-Aggregate Behavior in Poly-3-Hexylthiophene Nanofibers. *J. Phys. Chem. Lett.* **2012**, *3*, 259–263.
 54. Clark, J.; Chang, J. F.; Spano, F. C.; Friend, R. H.; Silva, C. Determining Exciton Bandwidth and Film Microstructure in Polythiophene Films Using Linear Absorption Spectroscopy. *Appl. Phys. Lett.* **2009**, *94*, 2007–2010.
 55. Kleinhenz, N.; Persson, N.; Xue, Z.; Chu, P. H.; Wang, G.; Yuan, Z.; McBride, M. A.; Choi, D.; Grover, M. A.; Reichmanis, E. Ordering of Poly(3-Hexylthiophene) in Solutions and Films: Effects of Fiber Length and Grain Boundaries on Anisotropy and Mobility. *Chem. Mater.* **2016**, *28*, 3905–3913.
 56. Rivnay, J.; Mannsfeld, S. C. B.; Miller, C. E.; Salleo, A.; Toney, M. F. Quantitative Determination of Organic Semiconductor Microstructure from the Molecular to Device Scale. *Chem. Rev.* **2012**, *112*, 5488–5519.
 57. Zhang, R.; Li, B.; Iovu, M. C.; Jeffries-EL, M.; Sauvé, G.; Cooper, J.; Jia, S.; Tristram-Nagle, S.; Smilgies, D. M.; Lambeth, D. N.; McCullough, R. D.; Kowalewski, T. Nanostructure Dependence of Field-Effect Mobility in Regioregular Poly(3-Hexylthiophene) Thin Film Field Effect Transistors. *J. Am. Chem. Soc.* **2006**, *128*, 3480–3481.
 58. Jimison, L. H.; Toney, M. F.; McCulloch, I.; Heeney, M.; Salleo, A. Charge-

Transport Anisotropy Due to Grain Boundaries in Directionally Crystallized Thin Films of Regioregular Poly(3-Hexylthiophene). *Adv. Mater.* **2009**, *21*, 1568–1572.

59. Kwiatkowski, J. J.; Jimison, L. H.; Salleo, A.; Spakowitz, A. J. A Boltzmann-Weighted Hopping Model of Charge Transport in Organic Semicrystalline Films. *J. Appl. Phys.* **2011**, *109*, 1–6.
60. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, 1–98.
61. Yang, H.; Lindquist, B. W. Three-Dimensional Image Analysis of Fibrous Materials. **2000**, *4115*, 275–282.
62. Shen, H.; Nutt, S.; Hull, D. Direct Observation and Measurement of Fiber Architecture in Short Fiber-Polymer Composite Foam through Micro-CT Imaging. *Compos. Sci. Technol.* **2004**, *64*, 2113–2120.
63. Pins, G. D.; Christiansen, D. L.; Patel, R.; Silver, F. H. Self-Assembly of Collagen Fibers. Influence of Fibrillar Alignment and Decorin on Mechanical Properties. *Biophys. J.* **1997**, *73*, 2164–2172.
64. Tang, X.; Rupp, B.; Yang, Y.; Edwards, T. D.; Grover, M. A.; Bevan, M. A. Optimal Feedback Controlled Assembly of Perfect Crystals. *ACS Nano* **2016**, *10*, 6791–6798.
65. Fullwood, D. T.; Niezgodna, S. R.; Adams, B. L.; Kalidindi, S. R. Microstructure Sensitive Design for Performance Optimization. *Prog. Mater. Sci.* **2010**, *55*, 477–562.
66. Bredfeldt, J. S.; Liu, Y.; Pehlke, C. A.; Conklin, M. W.; Szulczewski, J. M.; Inman, D. R.; Keely, P. J.; Nowak, R. D.; Mackie, T. R.; Eliceiri, K. W. Computational Segmentation of Collagen Fibers from Second-Harmonic Generation Images of Breast Cancer. *J. Biomed. Opt.* **2014**, *19*, 16007.
67. Lupulescu, A. I.; Rimer, J. D. In Situ Imaging of Silicalite-1 Surface Growth Reveals the Mechanism of Crystallization. *Science (80-.)*. **2014**, *344*, 729–732.
68. Jalili, R.; Morshed, M.; Ravandi, S. A. H. Fundamental Parameters Affecting Electrospinning of PAN Nanofibers as Uniaxially Aligned Fibers. *J. Appl. Polym.*

Sci. **2006**, *101*, 4350–4357.

69. Vader, D.; Kabla, A.; Weitz, D.; Mahadevan, L. Strain-Induced Alignment in Collagen Gels. *PLoS One* **2009**, *4*, e5902.
70. Natarajan, B.; Lachman, N.; Lam, T.; Jacobs, D.; Long, C.; Zhao, M.; Wardle, B. L.; Sharma, R.; Liddle, J. A. The Evolution of Carbon Nanotube Network Structure in Unidirectional Nanocomposites Resolved by Quantitative Electron Tomography. *ACS Nano* **2015**, *9*, 6050–6058.
71. Wählby, C.; Kametsky, L.; Liu, Z. H.; Riklin-Raviv, T.; Conery, A. L.; O'Rourke, E. J.; Sokolnicki, K. L.; Visvikis, O.; Ljosa, V.; Irazoqui, J. E.; Golland, P.; Ruvkun, G.; Ausubel, F. M.; Carpenter, A. E. An Image Analysis Toolbox for High-Throughput *C. Elegans* Assays. *Nat. Methods* **2012**, *9*, 714–716.
72. Stein, A. M.; Vader, D. A.; Jawerth, L. M.; Weitz, D. A.; Sander, L. M. An Algorithm for Extracting the Network Geometry of Three-Dimensional Collagen Gels. *J. Microsc.* **2008**, *232*, 463–475.
73. Murphy, J. N.; Harris, K. D.; Buriak, J. M. Automated Defect and Correlation Length Analysis of Block Copolymer Thin Film Nanopatterns. *PLoS One* **2015**, *10*, e0133088.
74. Usov, I.; Mezzenga, R. FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects. *Macromolecules* **2015**, *48*, 1269–1280.
75. Jordens, S.; Isa, L.; Usov, I.; Mezzenga, R. Non-Equilibrium Nature of Two-Dimensional Isotropic and Nematic Coexistence in Amyloid Fibrils at Liquid Interfaces. *Nat. Commun.* **2013**, *4*, 1917.
76. Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 11002.
77. Kalidindi, S. R.; Brough, D. B.; Li, S.; Cecen, A.; Blekh, A. L.; Congo, F. Y. P.; Campbell, C. Role of Materials Data Science and Informatics in Accelerated Materials Innovation. *MRS Bull.* **2016**, *41*, 596–602.

78. Krause, M.; Hausherr, J. M.; Burgeth, B.; Herrmann, C.; Krenkel, W. Determination of the Fibre Orientation in Composites Using the Structure Tensor and Local X-Ray Transform. *J. Mater. Sci.* **2010**, *45*, 888–896.
79. Weickert, J. Coherence-Enhancing Diffusion Filtering. *Int. J. Comput. Vis.* **1999**, *31*, 111–127.
80. Choi, D.; Chang, M.; Reichmanis, E. Controlled Assembly of Poly(3-Hexylthiophene): Managing the Disorder to Order Transition on the Nano- through Meso-Scales. *Adv. Funct. Mater.* **2015**, *25*, 920–927.
81. Chang, M.; Choi, D.; Fu, B.; Reichmanis, E. Solvent Based Hydrogen Bonding : Impact on Poly (3-Hexylthiophene) Nanoscale Morphology and Charge. *ACS Nano* **2013**, *7*, 5402–5413.
82. Verilhac, J.-M.; LeBlevenec, G.; Djurado, D.; Rieutord, F.; Chouiki, M.; Travers, J.-P.; Pron, A. Effect of Macromolecular Parameters and Processing Conditions on Supramolecular Organisation, Morphology and Electrical Transport Properties in Thin Layers of Regioregular poly(3-Hexylthiophene). *Synth. Met.* **2006**, *156*, 815–823.
83. Wang, G.; Persson, N.; Chu, P.; Kleinhenz, N.; Fu, B.; Chang, M.; Deb, N.; Mao, Y.; Wang, H.; Grover, M. A.; Reichmanis, E. Microfluidic Crystal Engineering of π -Conjugated Polymers. *ACS Nano* **2015**, *9*, 8220–8230.
84. Sirringhaus, H.; Brown, P. J.; Friend, R. H.; Nielsen, M. M.; Bechgaard, K.; Langeveld-Voss, B. M. W.; Spiering, A. J. H.; Janssen, R. A. J.; Meijer, E. W.; Herwig, P.; de Leeuw, D. M. Two-Dimensional Charge Transport in Self-Organized, High-Mobility Conjugated Polymers. *Nature* **1999**, *401*, 685–688.
85. Zhao, K.; Khan, H. U.; Li, R.; Su, Y.; Amassian, A. Entanglement of Conjugated Polymer Chains Influences Molecular Self-Assembly and Carrier Transport. *Adv. Funct. Mater.* **2013**, *23*, 6024–6035.
86. Cho, S.; Lee, K.; Yuen, J.; Wang, G.; Moses, D.; Heeger, A. J.; Surin, M.; Lazzaroni, R. Thermal Annealing-Induced Enhancement of the Field-Effect Mobility of Regioregular poly(3-Hexylthiophene) Films. *J. Appl. Phys.* **2006**, *100*, 114503.
87. Park, Y. D. The Molecular Structures of Poly(3-Hexylthiophene) Films Determine the Contact Properties at the Electrode/Semiconductor Interface. *Bull. Korean*

Chem. Soc. **2014**, *35*, 2277–2280.

88. Chang, M.; Lee, J.; Chu, P.-H.; Choi, D.; Park, B.; Reichmanis, E. Anisotropic Assembly of Conjugated Polymer Nanocrystallites for Enhanced Charge Transport. *ACS Appl. Mater. Interfaces* **2014**, *6*, 21541–21549.
89. Aiyar, A. R.; Hong, J.-I.; Izumi, J.; Choi, D.; Kleinhenz, N.; Reichmanis, E. Ultrasound-Induced Ordering in Poly(3-Hexylthiophene): Role of Molecular and Process Parameters on Morphology and Charge Transport. *ACS Appl. Mater. Interfaces* **2013**, *5*, 2368–2377.
90. Aiyar, A. R.; Hong, J.-I.; Nambiar, R.; Collard, D. M.; Reichmanis, E. Tunable Crystallinity in Regioregular Poly(3-Hexylthiophene) Thin Films and Its Impact on Field Effect Mobility. *Adv. Funct. Mater.* **2011**, *21*, 2652–2659.
91. Park, M. S.; Aiyar, A.; Park, J. O.; Reichmanis, E.; Srinivasarao, M. Solvent Evaporation Induced Liquid Crystalline Phase in Poly(3-Hexylthiophene). *J. Am. Chem. Soc.* **2011**, *133*, 7244–7247.
92. Kleinhenz, N.; Rosu, C.; Chatterjee, S.; Chang, M.; Nayani, K.; Xue, Z.; Kim, E.; Middlebrooks, J.; Russo, P. S.; Park, J. O.; Srinivasarao, M.; Reichmanis, E. Liquid Crystalline Poly(3-Hexylthiophene) Solutions Revisited: Role of Time-Dependent Self-Assembly. *Chem. Mater.* **2015**, *27*, 2687–2694.
93. Ayres, C.; Bowlin, G. L.; Henderson, S. C.; Taylor, L.; Shultz, J.; Alexander, J.; Telemeco, T. A.; Simpson, D. G. Modulation of Anisotropy in Electrospun Tissue-Engineering Scaffolds: Analysis of Fiber Alignment by the Fast Fourier Transform. *Biomaterials* **2006**, *27*, 5524–5534.
94. Lee, J. Y.; Bashur, C. A.; Goldstein, A. S.; Schmidt, C. E. Polypyrrole-Coated Electrospun PLGA Nanofibers for Neural Tissue Applications. *Biomaterials* **2009**, *30*, 4325–4335.
95. Niezgodna, S. R.; Kanjarla, A. K.; Kalidindi, S. R. Novel Microstructure Quantification Framework for Databasing, Visualization, and Analysis of Microstructure Data. *Integr. Mater. Manuf. Innov.* **2013**, *2*, 3.
96. Hu, L.; Hecht, D. S.; Grüner, G. Percolation in Transparent and Conducting Carbon Nanotube Networks. *Nano Lett.* **2004**, *4*, 2513–2517.

97. Perona, P.; Malik, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639.
98. Bai, X.; Zhou, F.; Xue, B. Image Enhancement Using Multi Scale Image Features Extracted by Top-Hat Transform. *Opt. Laser Technol.* **2012**, *44*, 328–336.
99. Fujiyoshi, H.; Lipton, A. J. Real-Time Human Motion Analysis by Image Skeletonization. In *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201)*; IEEE Comput. Soc, 1998; pp. 15–21.
100. Naf, M.; Kubler, O.; Kikinis, R.; Shenton, M. E.; Szekely, G. Characterization and Recognition of 3D Organ Shape in Medical Image Analysis Using Skeletonization. In *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*; IEEE, 1996; pp. 139–150.
101. Saeed, K.; Tabędzki, M.; Rybnik, M.; Adamski, M. K3M: A Universal Algorithm for Image Skeletonization and a Review of Thinning Techniques. *Int. J. Appl. Math. Comput. Sci.* **2010**, *20*, 317–335.
102. Brinkmann, M.; Chandezon, F.; Pansu, R. B.; Julien-Rabant, C. Epitaxial Growth of Highly Oriented Fibers of Semiconducting Polymers with a Shish-Kebab-like Superstructure. *Adv. Funct. Mater.* **2009**, *19*, 2759–2766.
103. Hashimoto, T.; Murase, H.; Ohta, Y. A New Scenario of Flow-Induced Shish-Kebab Formation in Entangled Polymer Solutions. *Macromolecules* **2010**, *43*, 6542–6548.
104. Mykhaylyk, O. O.; Chambon, P.; Impradice, C.; Fairclough, J. P. A.; Terrill, N. J.; Ryan, A. J. Control of Structural Morphology in Shear-Induced Crystallization of Polymers. *Macromolecules* **2010**, *43*, 2389–2405.
105. Mukhija, D.; Solomon, M. J. Nematic Order in Suspensions of Colloidal Rods by Application of a Centrifugal Field. *Soft Matter* **2011**, *7*, 540–545.
106. Diao, Y.; Tee, B. C.-K.; Giri, G.; Xu, J.; Kim, D. H.; Becerril, H. A.; Stoltenberg, R. M.; Lee, T. H.; Xue, G.; Mannsfeld, S. C. B.; Bao, Z. Solution Coating of Large-Area Organic Semiconductor Thin Films with Aligned Single-Crystalline Domains. *Nat. Mater.* **2013**, *12*, 665–671.

107. Wang, G.; Chu, P.-H.; Fu, B.; He, Z.; Kleinhenz, N.; Yuan, Z.; Mao, Y.; Wang, H.; Reichmanis, E. Conjugated Polymer Alignment: Synergisms Derived from Microfluidic Shear Design and UV Irradiation. *ACS Appl. Mater. Interfaces* **2016**, *8*, 24761–24772.
108. Qu, G.; Kwok, J. J.; Diao, Y. Flow-Directed Crystallization for Printed Electronics. *Acc. Chem. Res.* **2016**, *49*, 2756–2764.
109. Chang, M.; Choi, D.; Egap, E. Macroscopic Alignment of One-Dimensional Conjugated Polymer Nanocrystallites for High-Mobility Organic Field-Effect Transistors. *ACS Appl. Mater. Interfaces* **2016**, *8*, 13484–13491.
110. Le Berre, M.; Chen, Y.; Baigl, D. From Convective Assembly to Landau–Levich Deposition of Multilayered Phospholipid Films of Controlled Thickness. *Langmuir* **2009**, *25*, 2554–2557.
111. Katz, E.; Yarin, A. L.; Salalha, W.; Zussman, E. Alignment and Self-Assembly of Elongated Micronsize Rods in Several Flow Fields. *J. Appl. Phys.* **2006**, *100*, 34313.
112. Himmelberger, S.; Vandewal, K.; Fei, Z.; Heeney, M.; Salleo, A. Role of Molecular Weight Distribution on Charge Transport in Semiconducting Polymers. *Macromolecules* **2014**, *47*, 7151–7157.
113. 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, Z.; Luscombe, C.; McLachlan, M. A.; Heeney, M.; Rumbles, G.; Silva, C.; Salleo, A.; Nelson, J.; Smith, P.; Stingelin, N. 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.
114. Persson, N. E.; McBride, M. A.; Grover, M. A.; Reichmanis, E. Automated Analysis of Orientational Order in Images of Fibrillar Materials. *Chem. Mater.* **2017**, *29*, 3–14.
115. Bronstein, H.; Hurhangee, M.; Fregoso, E. C.; Beatrup, D.; Soon, Y. W.; Huang, Z.; Hadipour, A.; Tuladhar, P. S.; Rossbauer, S.; Sohn, E.-H.; Shoaee, S.; Dimitrov, S. D.; Frost, J. M.; Ashraf, R. S.; Kirchartz, T.; Watkins, S. E.; Song, K.; Anthopoulos, T.; Nelson, J.; Rand, B. P.; Durrant, J. R.; McCulloch, I. Isostructural, Deeper Highest Occupied Molecular Orbital Analogues of Poly(3-Hexylthiophene) for High-Open Circuit Voltage Organic Solar Cells. *Chem. Mater.* **2013**, *25*, 4239–4249.

116. Wade, J.; Wood, S.; Beatrup, D.; Hurhangee, M.; Bronstein, H.; McCulloch, I.; Durrant, J. R.; Kim, J. Operational Electrochemical Stability of Thiophene-Thiazole Copolymers Probed by Resonant Raman Spectroscopy. *J. Chem. Phys.* **2015**, *142*, 244904.
117. Heeney, M.; Zhang, W.; Crouch, D. J.; Chabynyc, M. L.; Gordeyev, S.; Hamilton, R.; Higgins, S. J.; McCulloch, I.; Skabara, P. J.; Sparrowe, D.; Tierney, S. Regioregular poly(3-Hexyl)selenophene: A Low Band Gap Organic Hole Transporting Polymer. *Chem. Commun.* **2007**, 5061–5063.
118. Shahid, M.; McCarthy-Ward, T.; Labram, J.; Rossbauer, S.; Domingo, E. B.; Watkins, S. E.; Stingelin, N.; Anthopoulos, T. D.; Heeney, M. Low Band Gap Selenophene–diketopyrrolopyrrolepolymers Exhibiting High and Balanced Ambipolar Performance in Bottom-Gate Transistors. *Chem. Sci.* **2012**, *3*, 181–185.
119. Burkhart, B.; Khlyabich, P. P.; Thompson, B. C. Influence of the Ethylhexyl Side-Chain Content on the Open-Circuit Voltage in Rr-Poly(3-Hexylthiophene- Co -3-(2-Ethylhexyl)thiophene) Copolymers. *Macromolecules* **2012**, *45*, 3740–3748.
120. Howard, J. B.; Noh, S.; Beier, A. E.; Thompson, B. C. Fine Tuning Surface Energy of Poly(3-Hexylthiophene) by Heteroatom Modification of the Alkyl Side Chains. *ACS Macro Lett.* **2015**, *4*, 725–730.
121. Chen, Z.; Lemke, H.; Albert-Seifried, S.; Caironi, M.; Nielsen, M. M.; Heeney, M.; Zhang, W.; McCulloch, I.; Sirringhaus, H. High Mobility Ambipolar Charge Transport in Polyselenophene Conjugated Polymers. *Adv. Mater.* **2010**, *22*, 2371–2375.
122. Ong, B. S.; Wu, Y.; Liu, P.; Gardner, S. Structurally Ordered Polythiophene Nanoparticles for High-Performance Organic Thin-Film Transistors. *Adv. Mater.* **2005**, *17*, 1141–1144.
123. Rivnay, J. Organic Electronics: Efficiency through Dilution. *Nat. Mater.* **2016**, *15*, 594–595.
124. Chang, M.; Su, Z.; Egap, E. Alignment and Charge Transport of One-Dimensional Conjugated Polymer Nanowires in Insulating Polymer Blends. *Macromolecules* **2016**, *49*, 9449–9456.
125. Chang, M.; Choi, D.; Wang, G.; Kleinhenz, N.; Persson, N.; Park, B.; Reichmanis,

- E. Photoinduced Anisotropic Assembly of Conjugated Polymers in Insulating Polymer Blends. *ACS Appl. Mater. Interfaces* **2015**, *7*, 14095–14103.
126. Choi, D.; Kim, H.; Persson, N.; Chu, P.-H.; Chang, M.; Kang, J.-H.; Graham, S.; Reichmanis, E. Elastomer-Polymer Semiconductor Blends for High Performance Stretchable Charge Transport Network. *Chem. Mater.* **2016**.
127. Chen, Z.; Lee, M. J.; Shahid Ashraf, R.; Gu, Y.; Albert-Seifried, S.; Meedom Nielsen, M.; Schroeder, B.; Anthopoulos, T. D.; Heeney, M.; McCulloch, I.; Siringhaus, H. High-Performance Ambipolar Diketopyrrolopyrrole-thieno[3,2-B]thiophene Copolymer Field-Effect Transistors with Balanced Hole and Electron Mobilities. *Adv. Mater.* **2012**, *24*, 647–652.
128. Zhang, X.; Bronstein, H.; Kronemeijer, A. J.; Smith, J.; Kim, Y.; Kline, R. J.; Richter, L. J.; Anthopoulos, T. D.; Siringhaus, H.; Song, K.; Heeney, M.; Zhang, W.; McCulloch, I.; DeLongchamp, D. M. Molecular Origin of High Field-Effect Mobility in an Indacenodithiophene–benzothiadiazole Copolymer. *Nat. Commun.* **2013**, *4*, 2238.
129. Caironi, M.; Bird, M.; Fazzi, D.; Chen, Z.; Di Pietro, R.; Newman, C.; Facchetti, A.; Siringhaus, H. Very Low Degree of Energetic Disorder as the Origin of High Mobility in an N-Channel Polymer Semiconductor. *Adv. Funct. Mater.* **2011**, *21*, 3371–3381.
130. Zhang, X.; Hudson, S. D.; DeLongchamp, D. M.; Gundlach, D. J.; Heeney, M.; McCulloch, I. In-Plane Liquid Crystalline Texture of High-Performance Thienothiophene Copolymer Thin Films. *Adv. Funct. Mater.* **2010**, *20*, 4098–4106.
131. Takacs, C. J.; Treat, N. D.; Krämer, S.; Chen, Z.; Facchetti, A.; Chabinyc, M. L.; Heeger, A. J. Remarkable Order of a High-Performance Polymer. *Nano Lett.* **2013**, *13*, 2522–2527.
132. Brinkmann, M.; Hartmann, L.; Biniek, L.; Tremel, K.; Kayunkid, N. Orienting Semi-Conducting π -Conjugated Polymers. *Macromol. Rapid Commun.* **2014**, *35*, 9–26.
133. Chang, M.; Lee, J.; Kleinhenz, N.; Fu, B.; Reichmanis, E. Photoinduced Anisotropic Supramolecular Assembly and Enhanced Charge Transport of Poly(3-Hexylthiophene) Thin Films. *Adv. Funct. Mater.* **2014**, *24*, 4457–4465.

134. Sharma, S.; Nair, S. S.; Zhang, Z.; Ragauskas, A. J.; Deng, Y. Characterization of Micro Fibrillation Process of Cellulose and Mercerized Cellulose Pulp. *RSC Adv.* **2015**, *5*, 63111–63122.
135. Arosio, P.; Beeg, M.; Nicoud, L.; Morbidelli, M. Time Evolution of Amyloid Fibril Length Distribution Described by a Population Balance Model. *Chem. Eng. Sci.* **2012**, *78*, 21–32.
136. Wie, J. J.; Nguyen, N. A.; Cwalina, C. D.; Liu, J.; Martin, D. C.; Mackay, M. E. Shear-Induced Solution Crystallization of Poly(3-Hexylthiophene) (P3HT). *Macromolecules* **2014**, *47*, 3343–3349.
137. Newbloom, G. M.; Weigandt, K. M.; Pozzo, D. C. Electrical, Mechanical, and Structural Characterization of Self-Assembly in Poly(3-Hexylthiophene) Organogel Networks. *Macromolecules* **2012**, *45*, 3452–3462.
138. Persson, N. E.; Chu, P.-H.; McBride, M.; Grover, M.; Reichmanis, E. Nucleation, Growth, and Alignment of Poly(3-Hexylthiophene) Nanofibers for High-Performance OFETs. *Acc. Chem. Res.* **2017**, *50*, 932–942.
139. Li, J.-H.; Xi, Y.; Pozzo, L. D.; Xu, J.-T.; Luscombe, C. K. Macroscopically Aligned Nanowire Arrays of π -Conjugated Polymers via Shear-Enhanced Crystallization. *J. Mater. Chem. C* **2017**, *5*, 5128–5134.
140. Kimata, S.; Sakurai, T.; Nozue, Y.; Kasahara, T.; Yamaguchi, N.; Karino, T.; Shibayama, M.; Kornfield, J. A. Molecular Basis of the Shish-Kebab Morphology in Polymer Crystallization. *Science* **2007**, *316*, 1014–1017.
141. Baghgar, M.; Labastide, J. A.; Bokel, F.; Hayward, R. C.; Barnes, M. D. Effect of Polymer Chain Folding on the Transition from H- to J-Aggregate Behavior in P3HT Nanofibers. *J. Phys. Chem. C* **2014**, *118*, 2229–2235.
142. Skrypnichuk, V.; Wetzelaer, G.-J. A. H.; Gordiichuk, P. I.; Mannsfeld, S. C. B.; Herrmann, A.; Toney, M. F.; Barbero, D. R. Ultrahigh Mobility in an Organic Semiconductor by Vertical Chain Alignment. *Adv. Mater.* **2016**, *28*, 2359–2366.
143. Shaw, L.; Hayoz, P.; Diao, Y.; Reinspach, J. A.; To, J. W. F.; Toney, M. F.; Weitz, R. T.; Bao, Z. Direct Uniaxial Alignment of a Donor–Acceptor Semiconducting Polymer Using Single-Step Solution Shearing. *ACS Appl. Mater. Interfaces* **2016**, *8*, 9285–9296.

144. Papathanasiou, T. D.; Guell, D. C. *Flow-Induced Alignment in Composite Materials*; Elsevier, 1997.
145. Lee, P.; Lin, R.; Moon, J.; Lee, L. P. Microfluidic Alignment of Collagen Fibers for in Vitro Cell Culture. *Biomed. Microdevices* **2006**, *8*, 35–41.
146. Kao, K. Y.; Lo, S. C.; Chen, H. L.; Chen, J. H.; Chen, S. A. Gelation of a Solution of poly(3-Hexylthiophene) Greatly Retards Its Crystallization Rate in the Subsequently Cast Film. *J. Phys. Chem. B* **2014**, *118*, 14510–14518.
147. Griffin, D. J.; Grover, M. A.; Kawajiri, Y.; Rousseau, R. W. Data-Driven Modeling and Dynamic Programming Applied to Batch Cooling Crystallization. *Ind. Eng. Chem. Res.* **2016**, *55*, 1361–1372.
148. Korevaar, P. a.; De Greef, T. F. a; Meijer, E. W. Pathway Complexity in π -Conjugated Materials. *Chem. Mater.* **2014**, *26*, 576–586.
149. Brinkmann, M.; Rannou, P. Molecular Weight Dependence of Chain Packing and Semicrystalline Structure in Oriented Films of Regioregular Poly(3-Hexylthiophene) Revealed by High-Resolution Transmission Electron Microscopy. *Macromolecules* **2009**, *42*, 1125–1130.
150. Zhao, Y.; Zhao, X.; Zang, Y.; Di, C.; Diao, Y.; Mei, J. Conjugation-Break Spacers in Semiconducting Polymers: Impact on Polymer Processability and Charge Transport Properties. *Macromolecules* **2015**, *48*, 2048–2053.
151. Zhao, B.; Zikry, M. A. The Effects of Structural Disorders and Microstructural Mechanisms on Semi-Crystalline P3HT Behavior. *Polymer (Guildf)*. **2015**, *57*, 1–11.
152. Gemünden, P.; Poelking, C.; Kremer, K.; Andrienko, D.; Daoulas, K. C. Nematic Ordering, Conjugation, and Density of States of Soluble Polymeric Semiconductors. *Macromolecules* **2013**, *46*, 5762–5774.
153. DeCost, B. L.; Holm, E. A. A Computer Vision Approach for Automated Analysis and Classification of Microstructural Image Data. *Comput. Mater. Sci.* **2015**, *110*, 126–133.
154. Dondossola, E.; Holzapfel, B. M.; Alexander, S.; Filippini, S.; Hutmacher, D. W.;

- Friedl, P. Examination of the Foreign Body Response to Biomaterials by Nonlinear Intravital Microscopy. *Nat. Biomed. Eng.* **2016**, *1*, 7.
155. Wirix, M. J. M.; Bomans, P. H. H.; Friedrich, H.; Sommerdijk, N. A. J. M.; de With, G. Three-Dimensional Structure of P3HT Assemblies in Organic Solvents Revealed by Cryo-TEM. *Nano Lett.* **2014**, *14*, 2033–2038.
156. Schuettfort, T.; Watts, B.; Thomsen, L.; Lee, M.; Siringhaus, H.; McNeill, C. R. Microstructure of Polycrystalline PBTTT Films: Domain Mapping and Structure Formation. *ACS Nano* **2012**, *6*, 1849–1864.
157. DeLongchamp, D. M.; Kline, R. J.; Jung, Y.; Germack, D. S.; Lin, E. K.; Moad, A. J.; Richter, L. J.; Toney, M. F.; Heeney, M.; McCulloch, I. Controlling the Orientation of Terraced Nanoscale “Ribbons” of a Poly(thiophene) Semiconductor. *ACS Nano* **2009**, *3*, 780–787.
158. Schmidt-hansberg, B.; Sanyal, M.; Klein, M. F. G.; Pfaff, M.; Schnabel, N.; Jaiser, S.; Vorobiev, A.; Mu, E.; Colsmann, A.; Scharfer, P.; Gerthsen, D. Moving through the Phase Diagram : Morphology Formation in Solution Cast Polymer À Fullerene Blend Films for Organic Solar Cells. *ACS Nano* **2011**, *5*, 8579–8590.
159. Nikiforov, M. P.; Darling, S. B. Concurrent Quantitative Conductivity and Mechanical Properties Measurements of Organic Photovoltaic Materials Using AFM. *J. Vis. Exp.* **2013**, 1–8.
160. Mulfari, D.; Celesti, A.; Fazio, M.; Villari, M.; Puliafito, A. Using Google Cloud Vision in Assistive Technology Scenarios. In *2016 IEEE Symposium on Computers and Communication (ISCC)*; IEEE, 2016; pp. 214–219.
161. Pinheiro, P. O.; Collobert, R.; Dollar, P. Learning to Segment Object Candidates. In *Advances in Neural Information Processing Systems 28*; Cortes, C.; Lawrence, N. D.; Lee, D. D.; Sugiyama, M.; Garnett, R., Eds.; Curran Associates, Inc., 2015; pp. 1990–1998.
162. Hsiao, L. C.; Schultz, B. A.; Glaser, J.; Engel, M.; Szakasits, M. E.; Glotzer, S. C.; Solomon, M. J. Metastable Orientational Order of Colloidal Discoids. *Nat. Commun.* **2015**, *6*, 8507.
163. Goldberg, I. G.; Allan, C.; Burel, J.-M.; Creager, D.; Falconi, A.; Hochheiser, H.; Johnston, J.; Mellen, J.; Sorger, P. K.; Swedlow, J. R. The Open Microscopy

Environment (OME) Data Model and XML File: Open Tools for Informatics and Quantitative Analysis in Biological Imaging. *Genome Biol.* **2005**, *6*, R47.

164. DeCost, B. L.; Francis, T.; Holm, E. A. Exploring the Microstructure Manifold: Image Texture Representations Applied to Ultrahigh Carbon Steel Microstructures. *Acta Mater.* **2017**, *133*, 30–40.
165. Chowdhury, A.; Kautz, E.; Yener, B.; Lewis, D. Image Driven Machine Learning Methods for Microstructure Recognition. *Comput. Mater. Sci.* **2016**, *123*, 176–187.
166. Lubbers, N.; Lookman, T.; Barros, K. Inferring Low-Dimensional Microstructure Representations Using Convolutional Neural Networks. **2016**, 1–25.
167. Lapshin, R. V. Automatic Drift Elimination in Probe Microscope Images Based on Techniques of Counter-Scanning and Topography Feature Recognition. *Meas. Sci. Technol.* **2007**, *18*, 907–927.
168. M. El Rifai, O.; Youcef-Toumi, K. On Automating Atomic Force Microscopes: An Adaptive Control Approach. *Control Eng. Pract.* **2007**, *15*, 349–361.
169. Fairbairn, M. W.; Moheimani, S. O. R. Control Techniques for Increasing the Scan Speed and Minimizing Image Artifacts in Tapping-Mode Atomic Force Microscopy: Toward Video-Rate Nanoscale Imaging. *IEEE Control Syst.* **2013**, *33*, 46–67.
170. Kalidindi, S. R. *Hierarchical Materials Informatics: Novel Analytics for Materials Data*; Elsevier, 2015.
171. Averbuch, A.; Coifman, R. R.; Donoho, D. L.; Elad, M.; Israeli, M. Fast and Accurate Polar Fourier Transform. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 145–167.
172. Averbuch, A.; Shabat, G.; Shkolnisky, Y. Direct Inversion of the 3D Pseudo-Polar Fourier Transform. **2015**.
173. Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

174. Arandjelovic, R.; Zisserman, A. All About VLAD. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE, 2013; pp. 1578–1585.
175. Jégou, H.; Douze, M.; Schmid, C.; Pérez, P. Aggregating Local Descriptors into a Compact Representation. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2010**, 3304–3311.
176. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE, 2014; pp. 1717–1724.
177. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **2014**.
178. Van Der Maaten, L. Accelerating T-SNE Using Tree-Based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221–3245.
179. Casciato, M. J.; Vastola, J. T.; Lu, J. C.; Hess, D. W.; Grover, M. A. Initial Experimental Design Methodology Incorporating Expert Conjecture, Prior Data, and Engineering Models for Deposition of Iridium Nanoparticles in Supercritical Carbon Dioxide. *Ind. Eng. Chem. Res.* **2013**, *52*, 9645–9653.
180. Kim, S.; Kim, H.; Lu, R. W.; Lu, J.-C.; Casciato, M. J.; Grover, M. A. Adaptive Combined Space-Filling and D -Optimal Designs. *Int. J. Prod. Res.* **2015**, *53*, 5354–5368.
181. Chen, Y.; Elenee Argentinis, J.; Weber, G. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. *Clin. Ther.* **2016**, *38*, 688–701.
182. Savva, M.; Kong, N.; Chhajta, A.; Fei-Fei, L.; Agrawala, M.; Heer, J. ReVision: Automated Classification, Analysis and Redesign of Chart Images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*; ACM Press: New York, New York, USA, 2011; p. 393.
183. Ray Choudhury, S.; Mitra, P.; Giles, C. L. Automatic Extraction of Figures from Scholarly Documents. In *Proceedings of the 2015 ACM Symposium on Document Engineering - DocEng '15*; ACM Press: New York, New York, USA, 2015; pp. 47–50.

184. Al-Zaidy, R. A.; Giles, C. L. Automatic Extraction of Data from Bar Charts. *Proc. Knowl. Capture Conf. ZZZ - K-CAP 2015* **2015**, 1–4.
185. Mosser, L.; Dubrule, O.; Blunt, M. J. Reconstruction of Three-Dimensional Porous Media Using Generative Adversarial Neural Networks. **2017**.
186. Weickert, J.; Schar, H. A Scheme for Coherence-Enhancing Diffusion Filtering with Optimized Rotation Invariance. *J. Vis. Commun. Image Represent.* **2002**, *13*, 103–118.
187. Yanowitz, S. D.; Bruckstein, a. M. A New Method for Image Segmentation. *Comput. Vision, Graph. Image Process.* **1989**, *46*, 82–95.
188. Smith, M. B.; Li, H.; Shen, T.; Huang, X.; Yusuf, E.; Vavylonis, D. Segmentation and Tracking of Cytoskeletal Filaments Using Open Active Contours. *Cytoskeleton* **2010**, *67*, 693–705.
189. Casey, J. *Exploring Curvature*; Springer Science & Business Media, 2012.
190. Bales, B.; Pollock, T.; Petzold, L. Segmentation-Free Image Processing and Analysis of Precipitate Shapes in 2D and 3D. *Model. Simul. Mater. Sci. Eng.* **2017**, *25*, 45009.
191. Editorial. Journals Unite for Reproducibility. *Nature* **2014**, *515*, 7–7.
192. Editor. Announcement: Reducing Our Irreproducibility. *Nature* **2013**, *496*, 398–398.
193. Editorial. Joining the Reproducibility Initiative. *Nat. Nanotechnol.* **2014**, *9*, 949–949.
194. Editorial. Code Share. *Nature* **2014**, *514*, 536–536.
195. PRIDE Archive <http://www.ebi.ac.uk/pride/archive/> (accessed Jan 12, 2016).
196. nanoHUB <https://nanohub.org/> (accessed Jan 13, 2016).

197. Kalidindi, S. R.; De Graef, M. Materials Data Science: Current Status and Future Outlook. *Annu. Rev. Mater. Res.* **2015**, *45*, 171–193.
198. Sirringhaus, H.; Ando, M. Materials Challenges and Applications of Solution Processed Organic Field-Effect Transistors. *MRS Bull.* **2008**, *33*.
199. Loo, Y.-L.; McCulloch, I. Progress and Challenges in Commercialization of Organic Electronics. *MRS Bull.* **2008**, *33*, 653–662.
200. Sheats, J. R. Manufacturing and Commercialization Issues in Organic Electronics. *J. Mater. Res.* **2011**, *19*, 1974–1989.
201. Krebs, F. C. All Solution Roll-to-Roll Processed Polymer Solar Cells Free from Indium-Tin-Oxide and Vacuum Coating Steps. *Org. Electron.* **2009**, *10*, 761–768.
202. Chang, J.-F.; Clark, J.; Zhao, N.; Sirringhaus, H.; Breiby, D. W.; Andreasen, J. W.; Nielsen, M. M.; Giles, M.; Heeney, M.; McCulloch, I. Molecular-Weight Dependence of Interchain Polaron Delocalization and Exciton Bandwidth in High-Mobility Conjugated Polymers. *Phys. Rev. B* **2006**, *74*, 115318.
203. Jiang, C.; Cheng, X.; Wu, X.; Yang, X.; Yin, B.; Hua, Y.; Wei, J.; Yin, S. Effects of P3HT Concentration on the Performance of Organic Field Effect Transistors. *Optoelectron. Lett.* **2011**, *7*, 30–32.
204. Zen, A.; Pflaum, J.; Hirschmann, S.; Zhuang, W.; Jaiser, F.; Asawapirom, U.; Rabe, J. P.; Scherf, U.; Neher, D. Effect of Molecular Weight and Annealing of poly(3-Hexylthiophene)s on the Performance of Organic Field-Effect Transistors. *Adv. Funct. Mater.* **2004**, *14*, 757–764.
205. Park, B.; Ko, D. Charge Transport in Ordered and Disordered Regions in Pristine and Sonicated-Poly(3-Hexylthiophene) Films. *J. Phys. Chem. C* **2014**, *118*, 1746–1752.
206. Scharsich, C.; Lohwasser, R. H.; Sommer, M.; Asawapirom, U.; Scherf, U.; Thelakkat, M.; Neher, D.; Köhler, A. Control of Aggregate Formation in poly(3-Hexylthiophene) by Solvent, Molecular Weight, and Synthetic Method. *J. Polym. Sci. Part B Polym. Phys.* **2012**, *50*, 442–453.
207. Choi, D.; Chu, P.-H.; McBride, M.; Reichmanis, E. Best Practices for Reporting

Organic Field Effect Transistor Device Performance. *Chem. Mater.* **2015**, *27*, 4167–4168.

208. Chabinyo, M. L.; Lu, J.-P.; Street, R. A.; Wu, Y.; Liu, P.; Ong, B. S. Short Channel Effects in Regioregular Poly(thiophene) Thin Film Transistors. *J. Appl. Phys.* **2004**, *96*, 2063–2070.
209. Kim, D. H.; Park, Y. D.; Jang, Y.; Yang, H.; Kim, Y. H.; Han, J. I.; Moon, D. G.; Park, S.; Chang, T.; Chang, C.; Joo, M.; Ryu, C. Y.; Cho, K. Enhancement of Field-Effect Mobility Due to Surface-Mediated Molecular Ordering in Regioregular Polythiophene Thin Film Transistors. *Adv. Funct. Mater.* **2005**, *15*, 77–82.
210. Kline, R. J.; DeLongchamp, D. M.; Fischer, D. a.; Lin, E. K.; Heeney, M.; McCulloch, I.; Toney, M. F. Significant Dependence of Morphology and Charge Carrier Mobility on Substrate Surface Chemistry in High Performance Polythiophene Semiconductor Films. *Appl. Phys. Lett.* **2007**, *90*, 62117.