SEM IMAGE REDUCTION AND ANALYSIS FOR MACHINE LEARNING OPTIMIZATION OF METAL HALIDE PEROVSKITES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Shanza Riaz

August 2023

ABSTRACT

Solar energy resources are poised as the biggest renewable competitor of fossil fuels. Solar perovskite thin films are an emerging technology with efficiencies close to established photovoltaic technologies. However, structural and thermal instability and an exponentially large number of possible compositions and processing parameters have led to reproducibility and scalability issues even at the lab scale. A DOE-funded collaboration across universities and national labs is creating a multi-scale statistical machine learning (ML) model to optimize the entire process from perovskite synthesis to device performance. The collaboration is called SPIRALs, which stands for Science and Processing Informed by Rational Algorithmic Learning (SPIRALs). A key challenge is integrating data from diverse data sources, such as experimental data and SEM images, as inputs to the ML algorithms. This work reports the Python code developed to reduce SEM images using spatial functions such as power spectral densities and autocorrelation functions. Quantitative analysis of these functions is done to derive spatial parameters such as the correlation length, grain size, fractal dimension, and Hurst exponent that describe the surface morphology of thin films. These parameters can tie synthesis methods to device performance and act as a check for ML framework predictions. The developed coding methodology and subsequent analysis can be applied to any thin film image and extended to different instruments to facilitate the SPIRALs project.

BIOGRAPHICAL SKETCH

Shanza was born in Rawalpindi, Pakistan in 1997. Growing up as an army brat, she moved to a new city every two years and has lived in all major provinces of the country. With change being a constant in her life, Shanza resolved to pursue her higher education abroad as early as middle school. She was an inquisitive child with a penchant for diverse subjects but found herself most captivated by the sciences. However, it was in 2016 that she discovered her true calling when she witnessed firsthand the profound impact of climate change. Each winter, the haunting sight of photochemical smog enveloping her cherished city, Lahore, underscored the urgency of tackling this environmental challenge. Recognizing the dependency of developing nations on fossil fuels and their environmental repercussions, she became impassioned to seek solutions through the development of renewable energy sources.

Driven by her aspirations, Shanza ventured to the United States to pursue her Bachelor's degree in Chemical Engineering at the Illinois Institute of Technology in Chicago, Illinois. She worked in the energy industry during the pandemic and the technical challenges in the energy sector prompted her interest in research in this space. She pursued a Master of Science degree which allowed her to collaborate with talented researchers in the emerging solar energy field. Having witnessed the potential impact of such research, Shanza envisions a future of continued dedication to exploring advancements in renewable energy within an industry-facing role. She ardently seeks to contribute meaningfully to the ongoing battle against climate change.

To Mama, Baba, Amna, and Hanan for your prayers, love, and support.

I love you all.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to the esteemed members of my committee, Dr. Tobias Hanrath, Dr. Lara Estroff, and Dr. John Marohn, for their invaluable support and guidance throughout this academic journey. I am particularly appreciative of Dr. Lara Estroff and Dr. John Marohn for taking me on as a research student and graciously serving as my joint advisors outside my field. Your guidance and mentorship have been pivotal during these last two years.

I want to convey my appreciation to the entire SPIRALs team. Engaging with such talented scientists from various fields and universities has been an immensely rewarding experience.

I also want to thank all members of the Marohn group, for their support, friendship, and inspired group meeting snacks! I want to give a special shout out to the EFM subgroup especially Rachael, Virginia, and Chris for showing me the ropes in the lab and for their help with my research and thesis. I wish you all the best for the future and I look forward to seeing your continued success.

In particular, I wish to extend my heartfelt thanks to Dr. John Marohn. Thank you, John, for being an exceptional mentor, and for your continuous support over the last two years. I am truly grateful for your instrumental role in reshaping my thesis and for guiding me patiently, through many coding challenges. Your passion for science and research is a big motivation for me going forward.

And last but not least, I want to express my heartfelt gratitude to my friends and family. Mama and Baba, you've been my pillars of support, always believing

in me and letting me chase my dreams abroad. To my younger siblings: you have been my biggest cheerleaders throughout, and I hope I've done you proud. To all my friends and family in Lahore, Ithaca, and everywhere else: thank you for always being there for me. I'm so lucky to have you all in my life!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and Background

Photovoltaic (PV) technologies are poised as the biggest competitor to fossil fuels for meeting global energy demand since they rely on an abundant renewable resource, solar energy. In comparison to current PV technology, metal halide perovskite (MHP) solar cells promise to be a cost-effective and highly efficient energy resource. They can be utilized at a TW-scale if they overcome the limitations of efficiency, scalability, sustainability, and stability [1].

1.1 Perovskite Crystal Structure and Material Properties

Perovskites, named after Russian mineralogist Lev Perovski, are part of a family of ABX_3 materials that share the same crystal structure as $CaTiO_3$. They have shared corner octahedra that extend in 3 dimensions as shown in Fig. 1.1. An MHP's typically contain an A-site cation like cesium, methylammonium, or formamidinium; a B-site cation like lead; and an X-site halide anion (I^-, Br^-, or Cl^-) [2]. Fig. 1.1 illustrates the general cubic perovskite crystal structure configuration.

In the early 1920s, Goldschmidt proposed a "tolerance factor" T to study the stability of oxide perovskites. Geometrically, for an ideal ABO_3 perovskite, the ratio of the bond length of the $A-O$ bond to the bond length of $B-O$ bond is $2:1$.

Assuming that the bond length is roughly the sum of two ionic radii, the tolerance factor T value of an ideal perovskite should be equal to 1.0. However, Goldschmidt found that, as an experimental fact, T values of most cubic perovskites are in the range of 0.8–0.9, and distorted perovskites occur outside this range [3]. This observation led to the following definition of the Goldschmidt tolerance factor for an ABX_3 perovskite [4].

$$T = \frac{R_a + R_b}{\sqrt{2}(R_b + R_x)} \tag{1.1}$$

In equation 1.1 the radius of the A cation is R_a, the radius of the B cation is R_b, and the radius of the X cation is R_x. Depending on T, perovskites form and phase changes occur and the ABX_3 structure can either be orthorhombic or cubic [5]. If T is less than 0.70, perovskites do not form, and if T is greater than 1, large cubic phase distortions are observed. If T lies in the narrower range 0.89–1.0, the cubic structure shown in Fig. 1.1 is likely, with lower T values giving less symmetric tetragonal or orthorhombic structures [6].

In the 1950s Muller and Roy developed plots of the ionic radii R_a versus R_b, termed 'structural maps' to study ternary oxide stability [7]. Li *et al.* built on this understanding and developed the octahedral factor, μ, that measures the ratio of the ionic radii of the cation and anion [5].

$$\mu = \frac{R_b}{R_x} \tag{1.2}$$

Equations 1.1 and 1.2 form the criterion for perovskite formability. For halide perovskites (X=F, Cl, Br, I), $0.81 < T < 1.11$ and $0.44 < \mu < 0.90$. Fig. 1.1 shows the structural map for 12 stable methylammonium (MA) and ethylammonium (EA;

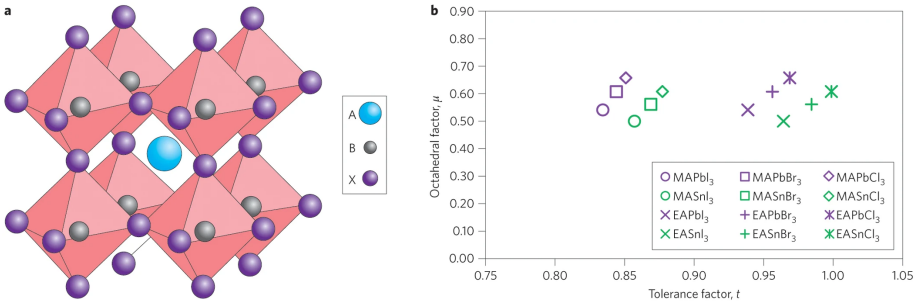

Figure 1.1: Cubic metal halide perovskite structure and structural map for perovskite formation. (a) 3-D cubic metal halide perovskite structure with A and B site cations and X site anions. (b) Calculated T and μ factors for 12 halide perovskites. Fig. 1.1 adapted from [9].

$CH_3CH_2NH_3$) halide perovskites developed from these equations. The corresponding formamidinium ($NH_2CH{=}NH_2$) based halides are expected to have intermediate values between those shown in Fig. 1.1. This formability space provides the option for many perovskites. More than 500,000 options for MHPs are possible, even if the M-site cation is restricted to Pb^{2+} alone [8]. In some cases, an intermediate tetragonal phase may also be present. T and u factor rules do not predict the temperature at which the cubic phase forms, and so heating is often required to reach the cubic phase with good photo absorption [9].

1.2 Advantages of Metal Halide Perovskites

Perovskite solar cells have high efficiencies that have been attributed to exceptional material properties such as remarkably high absorption over the visible spectrum, low exciton binding energy, charge carrier diffusion lengths in the

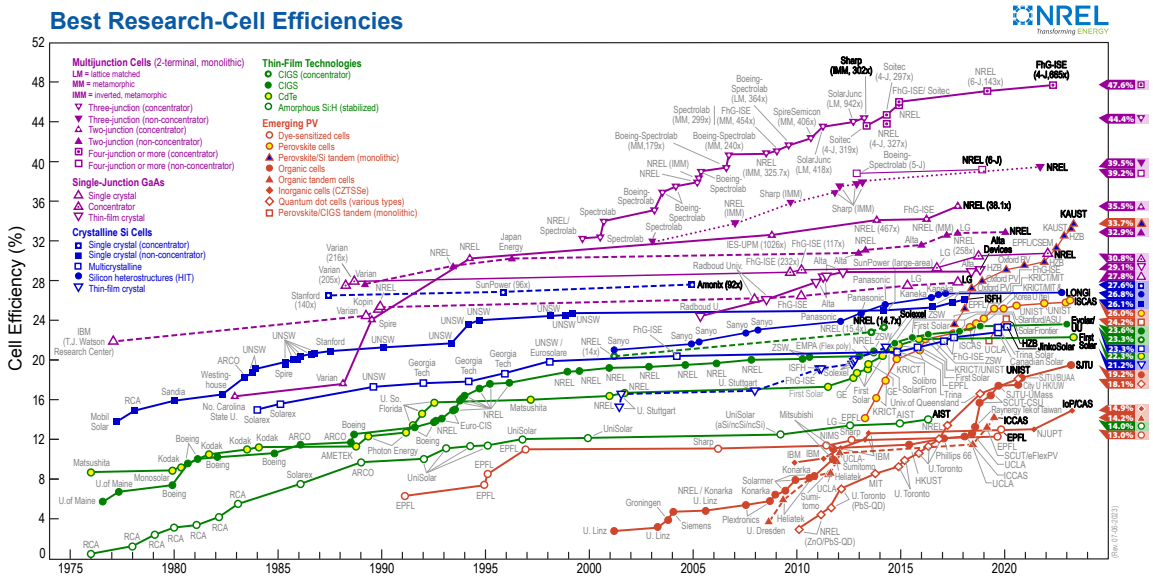

Figure 1.2: Efficiencies of research solar cells, comparison of single junction Si photovoltaic cells (at 26.7%) with perovskites (at 26.0%). [1]

millimeter range, a sharp optical band edge, and a tunable band gap from 1.1 to 2.3 eV by interchanging the above cations, metals and/or halides [10]. Fig. 1.2 compares the growth of perovskite solar cells (PSCs) with other photovoltaic cell types, such as those made from Si, copper, indium gallium selenide/sulfide, etc. Single-junction perovskite systems achieved 26.0% efficiency in 2023, while perovskite-silicon tandem systems achieved 33.7% efficiency. In comparison, Si hetero-structure cells have currently demonstrated 26.7% power conversion efficiency. These power conversion efficiencies have yet to plateau and are increasingly close to the Shockley-Queisser limit, which provides a theoretical maximum efficiency of 33 % for a single-junction solar cell [11].

[1] This plot is courtesy of the National Renewable Energy Laboratory, Golden, CO.

Moreover, perovskites can be easily processed using techniques such as spin coating, blade coating slot-dye coating, and atomic layer deposition, at room temperature making them one of the most versatile PV technologies [10].

1.3 Limitations of Metal Halide Perovskites

1.3.1 Thermal and Structural Instability

The biggest impediments to perovskite success are thermal and structural instabilities in pure perovskite compounds. External factors such as light, moisture, and oxygen cause chemical- and photo-oxidation degradation of perovskites [12]. Additionally, they undergo phase transitions, crystallizing to their non-cubic, non-conductive δ-phases at room temperature while the photoactive α-phase is stable at high temperatures. For example, $CsPbI_3$ has a band gap of 1.73 eV and good emissivity, but its photo-active α-phase is only stable at temperatures above 300°C and it crystallizes to an orthorhombic δ-phase structure at room temperature [13].

To address this issue, compositional engineering techniques have become a key design principle for perovskites [14]. This is because a reduction of band gap leads to an increase in electronic conductivity [13]. Moreover, modulating cation and halide compositions reduce the lattice strain making it harder to form vacancies, which in turn offers greater moisture stability. Furthermore, various film synthetic methods contribute to the development of stable inorganic PSCs.

For example, mixing additives with perovskite precursors minimizes defect traps by introducing an intermediate phase, and reduces lattice strain [12]. These approaches, while improving stability, increase the synthetic and processing complexity of perovskites.

1.3.2 Microscopic Grain Properties

Halide perovskites are a type of semiconductor that have unique optoelectronic properties. They are more tolerant to defects than conventional semiconductors, which has led to their use in the development of efficient photovoltaic devices[15, 16]. Early development of PSCs attributed the progress in efficiency to these unique defect structures in the bulk and benign grain boundaries [17]. However, recent research has shown conflicting reports on the role of grain boundaries. Studies have revealed that increasing grain size from nanometers to micrometers leads to improved device performance and elongated charge-carrier lifetimes, indicating a possible non-benign role of grain boundaries [18, 19, 20, 21, 22]. Other research suggests that grain boundaries contribute to non-radiative recombination and reduce device performance [15, 23]. Additionally, a recent theoretical study pointed out that GBs may even be the major recombination sites in the standard iodide-based perovskites, which seems to be consistent with the recent experimental efforts [24].

Even though it has been established that grain boundaries play a key role in the

performance of thin-film optoelectronic devices, their effect on halide perovskite materials is still not understood. The biggest factor limiting progress is the inability to identify grain boundaries. Microscopic image analysis techniques and softwares can misidentify grain boundaries, leading to conflicting literature reports about their influence. Electron backscatter diffraction (EBSD) is the most widely used method to identify grain boundaries but it destroys halide perovskite thin films [15, 25]. The spatially-resolved autocorrelation functions and radial power spectral densities developed in my research project are used to identify grain sizes and correlation lengths in perovskite thin films as explained in Sections 2 and 3. This analysis can be extended further to isolate grain boundaries.

Additionally, defects are primarily found either on the film surface, the bulk of the grain, or the boundary between neighboring grains. Increasing the grain size can also affect bulk and surface properties such as increasing crystallinity, reducing defect density at the surface and in the bulk, and reducing structural defects associated with pinhole formation in addition to possible changes in grain boundary properties. So, It is crucial to study the impact of these different microscopic factors on the electro-optical properties of polycrystalline perovskite thin films.

Moreover, another challenge in terms of material stability is understanding where degradation starts. Grain boundaries or the surface of perovskite films or both are considered to be the weakest points where the degradation starts[26, 27]. However, the degradation process is not yet understood fully. Additionally, ion migration in perovskite layers can also impact device stability, and controlling

grain size has been explored to mitigate this effect. Larger grain sizes have been shown to increase the activation energy for ion migration, potentially reducing degradation [28]. Thus, various microscopic factors affect device performance, posing a monumental experimental and computational challenge.

1.3.3 Processing Conditions

Metal halide perovskite device performance is also affected by factors such as the solvent, precursor chemicals, annealing temperatures, and different annealing profiles. These factors alter the crystallization pathway of the perovskite, which can lead to phase transformations. These properties impact the device's grain boundaries, impurity phases, and electrical conductivity which all affect performance [29].

This vast parameter and compositional space surrounding MHP device design have created scalability and reproducibility issues within and amongst labs. To tackle these challenges both on the microscopic and macroscopic scale, there is a need for statistical machine learning (ML) guidance.

1.3.4 Machine Learning Use in Literature

Machine learning techniques include computer algorithms that automatically evolve and adapt to the environment via experience, to make decisions and extract

knowledge [30]. Combining density functional theory data and statistical analysis, Park *et al.* suggested that the effective atomic radii and number of the lone pairs in the A-site are adequate to predict the perovskite stability [31]. Within the compositional engineering space, Yang *et al.* employed a machine learning process that identified the relationships between the octahedral structures and the anharmonicity of the halogen species [32]. Jacobs *et al.* computationally identified 15 out of 1845 halide perovskites that are comprised of nontoxic elements, stable in a humid operating environment, and have an optimal bandgap for single-junction cell applications [33]. Saidi *et al.* employed a hierarchical convolutional neural network to investigate the electronic properties of the halide perovskite materials, using neutral network element to calculate the proper structural and electronic features, obtained root-mean-square errors for the lattice constants, octahedral angle, and bandgap for the MHPs [34]. Bayesian optimization is employed by Clancy *et al.* to design the halide perovskite materials with the optimal combination of cations, halides, and solvents [35].

Based on published papers, Odabasi *et al.* employed machine learning to analyze the experimental details and device performance of perovskite solar cells that were reported in the literature from 2013 to 2018 and confirmed that the machine learning techniques could not only help develop new halide perovskite materials but could also facilitate the design of compatible neighboring materials and fabrication routes. This finding facilitates the selection of suitable hole transport layer (HTL), electron transport layer (ETL), solvents, anti-solvents, and spinning times [36].

Suitable processing conditions of the perovskite solar cells can also be optimized via the machine learning model. Ren *et al.* linked material parameters at certain conditions to illumination-dependent current-voltage (JV) measurements to get final device efficiency. Their model provided physical insights about the location of the best performing and robust processing conditions in solar cell devices [37].

Machine learning has also been combined with various experimental techniques to evaluate halide perovskite functionality. Ali *et al.* employed machine learningto evaluate X-ray diffraction, and scanning electron microscopy data to investigate the double-cation perovskites recovery in the cubic phase when <10 mol % cesium is added [38]. Theoretical simulations were also run to understand process kinetics such as anion exchange in halide perovskite nanoplates [39].

Machine learning methods have employed image recognition to analyze the size distributions of perovskite crystals and find boundaries of the perovskite crystal grains in SEM images [40]. Using a blind-source separation technique on SEM energy dispersive x-ray (EDX) spectral images, Jany *et al.* successfully extracted the nanostructures' pure X-ray signal to determine the chemical composition of the perovskite structures [41]. Madsen *et al.* developed a deep learning-based algorithm for the recognition of local atomic structures in TEM images, which is stable to microscope parameters and noise [42]. Liu *et al.* have developed a resolution-enhancement method for post-processing images from atomic force microscopy (AFM) using a deep convolution neural network to derive high-

resolution topography information from low-resolution topography images [43]. There is still an opportunity to retrieve a large number of data from images of experimental work since there are extensive text and figure-based content in the halide perovskite and materials science literature [30].

These papers showcase the application of machine learning techniques in predicting various perovskite parameters, such as solar cell performance, band gap, material properties, and thermo-electric behavior. However, there is still a vacuum for a combined work of high-throughput calculation, machine learning, data mining, and experiments that encompasses the comprehensive structural, mechanical, electronic, and optical properties of the halide perovskite materials [30].

1.3.5 SPIRALs Collaboration

The DOE-funded Science and Processing Informed by Rational Algorithmic Learning (SPIRAL) collaboration between groups at Cornell, Johns Hopkins, the University of Virginia (UVA), the Pacific Northwest National Laboratory (PNNL), and the National Renewable Energy Laboratory (NREL) is tackling this challenge by applying new statistical machine learning (ML) tools for the predictive formulation engineering of solution-processed, solid-state perovskites all the way from individual reagents in solution to functional devices. This is done by developing a convergent machine-learning framework that takes data input derived from first-principles simulations and detailed experimental characterization [8].

One of the main hurdles faced by the collaboration lies in handling the multi-dimensional, multi-scale combinatorial aspects of the perovskite formulation systems, starting with data sources. To address this, the BayesOpt algorithm developed by Clancy *et al.* is used to run through multi-scale data inputs from various experimental sources [35]. However, integrating data from disparate sources, such as experimental data and images, presents challenges. In particular, SEM images contain large volumes of metadata that add noise to the algorithm and hinder data optimization. Another challenge is the lack of information on the role of defects in perovskite degradation [8].

In this work, I have developed a Python code to reduce SEM images into their power spectra and autocorrelation functions. These functions are analyzed to identify surface morphology parameters specific to each image. The grain size and correlation length of the thin film are also identified which can facilitate the defect analysis work being done by the Clancy group. Instead of using equipment-specific image analysis tools such as ImageJ, the developed Python code can easily be integrated into the broader machine-learning workflow. When combined with information on processing conditions, the code will serve as a check on film parameters extracted during the experiment and through computation.

Chapter 2

SEM Image Reduction Using Spatial Functions

2.1 SEM Image Analysis Application for Thin Films

Scanning electron microscopy (SEM) imaging is a technique used to examine the surface morphology and microstructure of perovskite thin films at high resolution. It involves scanning a focused electron beam over a sample's surface and detecting the backscattered electrons to create an image. SEM imaging provides valuable information about the topography and spatial features of a surface [44].

Images analyzed in this thesis were taken on the Zeiss Gemini 500 scanning electron microscope at Cornell.[1] The code to derive the image pixel size from an image has been optimized to extract the metadata from images taken with this particular equipment. The code needs to be modified slightly based on the model of SEM used since the pixel size reference in the metadata varies. This modification was successfully made to analyze SEM images taken with the FEI Quanta 650 Scanning Electron Microscope sent over by the Choi group at the University of Virginia.

In the context of thin film imaging, SEM is particularly useful for characteriz-

[1]The author acknowledges the use of facilities and instrumentation supported by NSF through the Cornell University Materials Research Science and Engineering Center DMR-1719875.

ing the morphology and quality of thin film structures, such as perovskite films. It has been used to identify voids in the films [45] as well as identify defects and changes in surface morphology such as grain size which are shown to affect device performance and reproducibility [46].

Overall, SEM image analysis plays a vital role in the characterization and understanding of perovskite materials. SEM images provide valuable insights into the microstructure, morphology, and composition of perovskite-based devices such as solar cells, LEDs, and photodetectors [44].

2.1.1 Challenges in SEM Image Analysis

A cursory qualitative analysis of SEM images can be done visually; however, height information is not available just by looking at the image; getting additional surface morphology factors such as grain size requires further analysis. This is usually done with image analysis software such as ImageJ which is time-consuming and performed on individual images. Additionally, to employ machine learning to make predictions with image analysis, a large body of images need to be compared. This brings in the issues of instrument resolution, noise, and other instrument settings which hinder image comparison [47].

Conventionally, the deviation of a surface from its mean plane is assumed to be a random process for which statistical parameters such as the variances of the height, slope, and curvature are used for characterization [48]. The most com-

mon descriptor for surface morphology is the root mean square (RMS) roughness that reduces the whole information content of the image to one single RMS value. Surfaces with very different morphology may have the same RMS value of surface roughness and for many surfaces, the RMS roughness depends on the length scale used for the measurement. The precise description of surface morphology, therefore, calls for more sophisticated tools [49].

To address these challenges, spatial functions such as the power spectral density function (PSDF) and auto-covariance (or autocorrelation) function (ACF) are used to represent properties of all wavelengths or spatial sizes within the image [47, 50]. Spatial functions, in contrast to conventional statistical analysis such as RMS values, break down the surface image into its frequency domain and acquire detailed roughness data to describe surface features[51]. Additionally, obtaining roughness values by integrating the area under the different PSD curves in the same range removes the problem of each instrument having different upper and lower band limits on the spatial function curves [52].

Furthermore, these spatial functions are analyzed using methods of fractal geometry (explained in Chapter 3) to obtain topographical parameters (fractal dimension, correlation length, and Hurst exponent), which are relatively independent of the resolution of the instrument and scale invariant [52]. Using these parameters, we can easily compare SEM images taken on different instruments and feed this data as inputs into the machine learning algorithm to get informed outputs. This fractal analysis is explained with examples in the next chapter.

2.2 Mathematical Equations for Spatial Functions

In literature, power spectral density functions used for profiling data differ due to different normalization procedures [53]. This creates reproducibility issues and hinders the comparison of results. Hence, I derived the normalization factors for spatial functions prior to coding.

2.2.1 Fourier Transform and PSD Equations for SEM Images

For the general case, suppose the topography of a surface within a Cartesian space $S(x, y)$ is described by a continuous function $S(k_x, k_y)$, where k_x and k_y are wave vectors (spatial frequencies) and L_x and L_y are the lengths in x and y directions respectively. The integration is computed over the full area $L_x L_y$ of the entire periodic surface from negative to positive infinity. This gives the continuous Fourier transform $S(k_x, k_y)$ equation 2.1.

$$\hat{S}(k_x, k_y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \;\; S(x, y) \;\; e^{2\pi i(xk_x + yk_y)} \quad [\text{a.u.nm}^2] \tag{2.1}$$

The units for equation 2.1 depend on the units of the scan, which usually has units of length.

In the case of an SEM image, spatial functions exist as discrete points on a grid with N_x and N_y representing the total number of pixels in the x and y direction of the global coordinate system. N_x and N_y are set to be equal so that the computation

of the radial average of the power spectrum can be simplified.

$S_{k,\ell}$ represents the discrete data obtained by sampling at discrete spatial grid points $(k\Delta x, \ell\Delta y)$ where Δx and Δy represent the image pixels size with units of length, the total spatial length is given by $L_x = \Delta x\, N$ and $L_y = \Delta y\, N$, and the frequency step is given by $\Delta k_x = 1/L_x$ and $\Delta k_y = 1/L_y$ in the x and y directions respectively. The discrete Fourier transformation $\hat{S}_{m,n}$ with indices $m = 0, 1, 2, \ldots, N-1$ and $n = 0, 1, 2, \ldots, N-1$ in the frequency domain is

$$\hat{S}_{m,n} = \Delta x \Delta y \sum_{k=0}^{N_x-1} \sum_{\ell=0}^{N_y-1} S_{k,\ell}\, e^{2\pi i\left(\frac{k\,m}{N_x} + \frac{\ell\,n}{N_y}\right)} \quad [\text{a.u.nm}^2]\cdot \tag{2.2}$$

We find the magnitude squared of the Fourier transform to get the 2-D power spectral density $\hat{P}_{m,n}$ in the frequency domain as shown.

$$\hat{P}_{m,n} = \frac{1}{L_x L_y} |\hat{S}_{m,n}|^2 \quad [\text{a.u.}^2\text{nm}^2] \tag{2.3}$$

or

$$\hat{P}_{m,n} = \Delta k_x \Delta k_y\, |\hat{S}_{m,n}|^2 \quad [\text{a.u.}^2\text{nm}^2] \tag{2.4}$$

The units are determined by the normalization factors found in the above equations which account for the finite number of measured values. The analysis of the 2-D and 1-D radial PSD is discussed in Chapter 3.

2.2.2 Autocorrelation Function Equations

The autocorrelation function represents the correlation between different points in the measured values [54, 55]. For the discrete function, $S_{k,\ell}$, data is displaced by indices k' and ℓ'. The autocorrelation function $C_{k',\ell'}$ is defined as [48].

$$C_{k',\ell'} = \frac{1}{N_x N_y} \sum_{k=0}^{N_x-1} \sum_{\ell=0}^{N_y-1} S^*_{k,\ell} \cdot S_{k+k',l+\ell'} \quad [\text{a.u.}^2] \tag{2.5}$$

where $S^*_{k,\ell}$ is the complex conjugate of $S_{k,\ell}$.

The autocorrelation function is also found using the Fast Fourier transform (FFT) technique. Since the height information is not shown on the SEM Image, I use the FFT technique to find the autocorrelation function. The Fourier transform of the autocorrelation function is related to the PSD function through the Wiener-Khinchin theorem such that the inverse FFT of the PSD function is taken to get the ACF [54].

$$C = \mathcal{F}^{-1}(\hat{P}) \leftrightarrow \mathcal{F}\,(C) = \hat{P} \tag{2.6}$$

where $\mathcal{F}$ denotes the Fourier transform operator. Using the PSD function equations 2.3, the discrete ACF for an SEM image is given by

$$C_{k',\ell'} = \frac{1}{L_x L_y} \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} \hat{P}_{m,n}\, e^{-2\pi i\left(\frac{k'm}{N_x}+\frac{\ell'n}{N_y}\right)} \quad [\text{a.u.}^2]. \tag{2.7}$$

2.2.3 Parseval's Theorem

Parseval's theorem relates the "energy" of a function in the time domain to the "energy" of its Fourier transform in the frequency domain. It states that the total "energy" of a function is equal to the sum of the squared magnitudes of its Fourier coefficients. For the case where the displacement is 0 (at the origin), and infinity, such that $k' = \ell' = 0$, the autocorrelation function simplifies to the summation of the squared function $S_{k,\ell}$ as shown in equation 2.8.

$$C_{0,0} = \frac{1}{N_x N_y} \sum_{k=0}^{N_x-1} \sum_{\ell=0}^{N_y-1} |S_{k,\ell}|^2 \quad [\text{a.u.}^2] \tag{2.8}$$

If the reference is taken as the mean plane of the surface, the displacement values will correspond to the mean of the function and the autocorrelation function corresponds to the mean squared value of the function [48, 56]. In the case of SEM images, this corresponds to the mean squared value of the original surface.

Similarly, for $C_{(0,0)}$, k' and ℓ' in equation 2.7 become 0 and the ACF simplifies to the summation of the PSD function which corresponds to the area under the PSD function curve as shown.

$$C_{(0,0)} = \frac{1}{L_x L_y} \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} \hat{P}_{m,n} \quad [\text{a.u.}^2] \tag{2.9}$$

Equation 2.8 and 2.9 are equal since the magnitude squared of the FFT is equal to the PSD as shown above in equation 2.3, and 2.4.

Hence, we have shown that the mean squared value of the image is equal to

the area under the PSD function, which verifies Parseval's theorem. Using this proof, I have validated the normalization factors derived from these equations in my code.

2.3 Image Reduction Code Methodology

The developed code is broken down into two steps. The first step consists of computing the 2-D power spectral density, 2-D autocorrelation function, 1-D radial power spectral density, and 1-D radial autocorrelation plots. In the second step, these plots are fit and analyzed with fitting methods used in the literature. This step-by-step coding methodology for image reduction is explained in this section and in the subsequent image analysis is discussed in Chapter 3 of this thesis.

2.3.1 Image Metadata Extraction

To start, a library of SEM images was created that was used to extract images in Python. By creating a path to this image library, the images can be accessed using the Python Imaging Library (PIL). Then, image metadata was retrieved and stored as a string. The basic metadata tags don't store the image pixel size in terms of length explicitly, so regular expressions `re.search()` were used to search for a pattern in the `metadata` string. For the Zeiss Gemini 500 scanning electron microscope, this pattern is identified and matched as the phrase `Image Pixel`

`Size =`, and the units are matched as either `nm` or `microns`. The numeric value after the `Image Pixel Size =` was extracted and stored in the d_x and d_y variables and used throughout the code. The function to generalize this process is shown.

```
def get_image_pixel_size(image_path):
    image = Image.open(image_path)
    metadata = str(image.tag_v2)
    match = re.search(r'Image Pixel Size =
        ([0-9]{1,4}(?:\.[0-9]*)) (nm|microns)', metadata)
    if match:
        image_pixel_size = float(match.group(1))
        unit = match.group(2)
        print(f"Image Pixel Size: {image_pixel_size} {unit}")
        dx, dy = image_pixel_size
        width, height = image.size
        total_image_size = image_pixel_size * width,
            image_pixel_size* height
        return total_image_size, dx, dy, unit
```

The code can be applied to images from any SEM, but the regular expression will need to be edited based on how the image pixel size is stored in the metadata while the rest of the code can be used as is.

2.3.2 2-D FFT and PSD Plots

The next step was to crop the image, such that the total pixel size N_x and N_y is the same in the x and y direction. The variable $npts$, set by the user via trail-and-error, is used to remove the scale-bar from the image. The new value of $N = N_x = N_y$ was calculated such that a new value N is the nearest power of 2 that is less than the total number of pixels that are set as the function argument. This allowed the efficient use of the discrete FFT algorithm, cut down on processing time for big SEM images, and obtain a symmetric image for further reduction. For consistency and symmetry, the image was cropped from the center outwards. This symmetrization procedure was followed by subtracting the mean value of the image for baseline correction, to reduce any systematic biases or variations in the image data, ensuring accurate analysis later. The new values of the total pixel size N_x, N_y and total length L_x, L_y after cropping were stored and used in the next steps as shown in the function below.

```
def crop_image(img_plotted, img, npts):
    Nx, Ny = img[0:npts, :].shape
    N = 2 ** (math.ceil(math.log2(npts)) - 1)
    img_cropped = img[0:npts, :][
    int(Nx / 2 - N / 2):int(Nx / 2 + N / 2),
    int(Ny / 2 - N / 2):int(Ny / 2 + N / 2)]
    img_cropped = img_cropped - img_cropped.mean()
    Nx, Ny = N
```

```
    Lx= Nx*dx
    Ly = Nx*dy
```

Next, a 2-D Fourier transform was applied to the image, converting the image to the frequency domain. To keep the code simple and reproducible, the NumPy library was used instead of the Scipy library to run the FFT algorithm. The resulting Fourier transform is cyclicly shifted so that the zero-frequency point is located in the center of the Fourier-transformed data. The real and complex conjugate parts of the Fourier transform were multiplied together and divided by $L_x L_y$ to get the 2-D power spectral density (PSD). The resulting 2-D FFT and PSD plots were used to examine the frequency components present in the image and gain insights into its spatial characteristics. The normalization factors for the FFT and PSD steps were derived from the mathematical calculations done in section 2.2.1 and used in the code. The logarithmic values of the FFT and PSD data were plotted so that intensity variations are more visible. The relevant code from the function to compute the FFT and PSD is shown below.

```
def compute_psd(img_cropped):
    FT = dx * dy * np.fft.fftshift(np.fft.fft2(img_cropped))
    img_psd = np.real(FT * np.conj(FT)) / (Lx*Ly)
    #code to plot the FFT and PSD
    return img_psd
```

2.3.3 Parsevals Proof

The code was validated by applying the Parseval theorem outlined in section 2.2.3. Confirming that the mean square value of the image equals the sum of the power spectral density (equivalently, area under the power spectral density curve). Testing this equality confirms that the normalization factors and code for the FFT and PSD plots are correctly derived and coded.

```
def rms_squared(img_cropped):
        img_rms_squared = np.var(img_cropped)
        rounded_rms_squared = round(img_rms_squared, 2)
        print(f"RMS squared value of the image:{
           rounded_rms_squared}(a.u.^2)")
        return img_rms_squared
def psd_sum(img_psd):
        dkx = 1/Lx
        dky = 1/Ly
        psd_sum = dkx*dky*np.sum(img_psd)
        rounded_psd_sum = round(psd_sum, 2)
        print(f"Total Sum of the area under the PSD: {
           rounded_psd_sum}(a.u.^2)")
        return psd_sum
```

2.3.4 1-D Radial PSD Plot

Since the 2-D FFT and PSD plots are not easy to analyze, the 1-D radial power spectral density were plotted to get further insights into the frequency distribution of the image. First, 1-D arrays for the total length and the corresponding spatial frequency in the x and y direction were created. Then a 2-D grid was created from the spatial frequencies and the radial frequency $K[\text{nm}^{-1}]$ values at each point of the grid were calculated using the Euclidean distance formula. The radial frequency was indexed by sorting its elements in ascending order and this was used to reorder the flattened K and 2-D PSD arrays. Then, a moving filter was applied to the sorted PSD array using the `np.convolve()` function with the mode set to `same` which convolves the array with a window of the same size as the total pixel size N. This windowing corrupts data points at the corners so a small number of points away from each end of the arrays are sliced away as shown in the code.

```
def compute_K_freq(img_psd_freq):
    kx = dkx*np.arange(-Lx / 2 , Lx / 2, Lx/512)
    ky = dky*np.arange(-Ly / 2, Ly / 2, Ly/512)
    kx_freq = np.fft.fftshift(np.fft.fftfreq(kx.size))
    ky_freq = np.fft.fftshift(np.fft.fftfreq(ky.size))
    kX, kY = np.meshgrid(kx_freq, ky_freq)
    K = np.sqrt(kX ** 2 + kY ** 2)
    return K
```

```
K = compute_K_freq(img_psd)
def compute_psd_radial(K, npts):
    K_index = K.flatten().argsort()
    K_sorted = K.flatten()[K_index]
    img_sorted = img_psd.flatten()[K_index]
    img_sorted_avg = np.convolve(
        img_sorted,
        np.ones(npts) / npts,
        mode='same')
    kr = K_sorted[npts:-npts]
    psdr = img_sorted_avg[npts:-npts]
    return kr, psdr
```

Finally, the 1-D radial power spectral density was plotted versus the radial frequency as shown in Fig. 2.1 (e). The log-log radial PSD is plotted to cover a larger spatial range [54].

2.3.5 2-D and 1-D Autocorrelation Plots

As explained mathematically in Section 2.2.2, the inverse Fourier transform of the power spectral density is the autocorrelation function. Consequently, the 2-D autocorrelation function was computed from the 2-D PSD plots. The `np.fft.ifftshift()` function was used to shift the zero frequency component

to its original position to undo the shift from the PSD computation. The autocorrelation function was normalized and then cyclicly shifted again to center it at zero.

Next, 1-D arrays of the spatial distance in the x and y direction were created and their 2-D grid was created. The radial distance R[nm] was calculated from the origin for each point in the 2-D grid using the Euclidean distance, equation 2.10.

$$R = \sqrt{x^2 + y^2}. \tag{2.10}$$

A 1-D array is created to store the R values corresponding to the maximum value of the flattened 2-D autocorrelation array for a given bin size using a `for` loop. The 1-D radial autocorrelation is created and plotted using a binning method. This procedure sorts the data by radial distance and made the plot easier to visualize and analyze to get the grain size.

The average grain size is determined as the radius at which the first maximum in the radial autocorrelation function occurs. This average grain size value serves as an indicator of the characteristic size of structures present in the image.

```
def autcorr_2dpsd(img_psd_freq, Lx, Ly):
    autocorr = np.fft.ifft2(np.fft.ifftshift(img_psd_freq))
    autocorr = np.real(autocorr/autocorr[0,0])
    autocorr = np.fft.ifftshift(autocorr)
    x=np.arange(start=-Lx/2, stop=Lx/2, step=Ly/512)
    y=np.arange(start=-Ly/2, stop=Ly/2, step=Ly/512)
```

```
    xo, yo = np.meshgrid(x,y)
    R = np.sqrt(xo ** 2 + yo ** 2).flatten()
    R_bins = np.linspace(0, np.max(R), 200)
    autocorr_radial = np.zeros(len(R_bins) - 1)
    for i in range(len(R_bins) - 1):
        indices = np.where((R >= R_bins[i]) & (R < R_bins[i+1]))
        autocorr_radial[i] = np.max(autocorr.flatten()[indices])
```

The code described above was used to reduce and transform an SEM image of a $PbCsBr_3$ thin film as shown in Fig. 2.1. This code was run on a simple image with one circle, an image containing concentric circles, and a simple sinusoidal grating image. These served as test cases to verify the code and extract the grain size from the autocorrelation function. These results can be compared to the actual SEM image PSD and autocorrelation plots for further analysis as discussed in Section 3.2.4.

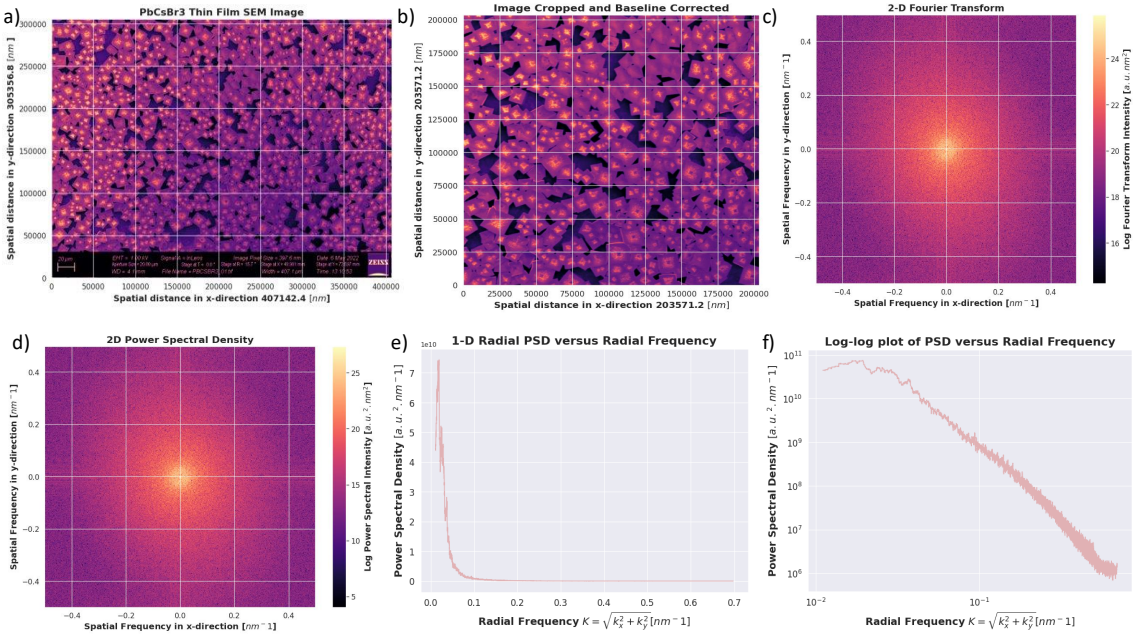

Figure 2.1: $PbCsBr_3$ thin film SEM image reduction into 2-D FFT, 2-D PSD, and 1-D RPSD functions. (a) Original SEM image of $PbCsBr_3$ thin film; (b) cropped and baseline corrected image; (c) 2-D FFT of image; (d) 2-D PSDF of image; (e) 1-D radial PSD plot; and (f) log-log radial PSD plot.

Chapter 3

Qualitative and Quantitative Image Analysis

3.1 Qualitative Analysis

Having set up the code to extract PSD and ACF plots, we can move on to the qualitative and quantitative analysis of these spatial functions.

The 2-D FFT and PSD plots (Fig. 2.1 (c) and (d)) of the $CsPbBr_3$ sample provides some basic qualitative information regarding the surface. The periodic features in the real image produce peaks in the FFT image, and the frequency of the peaks are the wavelengths of those periodic features. The size of the bright central feature represents the general size of the lateral surface structures. Lines or streaks indicate a preferred orientation of the surface microstructure [53]. The 2-D PSD and FFT plots are useful when comparing images with very different surface morphologies since they offer a qualitative view of the image. Further processing is required to develop a more quantitative understanding of the surface features.

This is where the 1-D radial PSD plots come in. 1-D radial PSD averages out any orientation-dependent structural information. Nevertheless, the radial PSD plot allows for a facile analysis of surface morphology — the peaks in the plot indicate dominant frequencies or features in the image. The amplitude of the peaks is proportional to the strength or intensity of those features. Analyzing

the peak positions, widths, and amplitudes give insights into the characteristic length scales and periodicity present in the surface. Generally, periodic surfaces have peaks, whereas homogeneous thin films typically exhibit a relatively flat or uniform radial PSD with low power at high frequencies and a dominant peak or plateau at a specific frequency related to the film's characteristic features as seen in Fig. 2.1(e) and (f). The shape and position of this peak or plateau provide insights into the film's microstructure, surface roughness, or periodicity [57].

3.2 Quantitative Analysis

As Mandelbrot explained in the early 1980s, rough surfaces formed by natural phenomena are self-similar fractals (isotropic magnification) or self-affine fractals (anisotropic magnification). When a rough surface is repeatedly magnified, increasing details of roughness are observed right down to the nanoscale which look quite similar at a range of magnifications. Such a behavior can be characterized by fractal geometry using scale-independent parameters that provide information on the roughness structure at all length scales [58].

This concept of fractal geometry has also been extended to describe the surface morphology and complexity of thin films, which can often be considered to be statistically self-affine [48]. Additional studies have established that AFM images can exhibit an isotropic growing surface similar to natural fractal surfaces [51, 59, 60, 61]. Thus, SEM images can also be considered isotropic and fractal analysis can

be applied. The log-log RPSD plots of these isotropic, self-affine surfaces follow a power law as established by Mandelbrot [62].

3.2.1 Inverse Power Law Method Derivation

Theory of Fractal Geometry

Euclidean geometry describes ordered objects such that for any given integer N, a segment of the length L is the sum of N straight segments of length $r = L/N$. Using these variables, the dimension of the object is found such that

$$D = \frac{\log N}{\log (1/r)} \tag{3.1}$$

Thus, the dimensions for points, curves, surfaces, and cubes are 0, 1, 2, and 3, respectively. Associated with each dimension is a measure of the object such as the length of a line, the area of a surface, and the volume of a cube. These measures are invariant with respect to the unit of measurement, implying that the dimension of a line remains independent of whether a centimeter or a micrometer scale is used.

However, most objects found in nature [58], rough surfaces for example, appear disordered and irregular for which the measures of length, area, and volume are scale-dependent. This suggests that the dimensions of such objects cannot be integers. Such objects are said to be fractal.

Mandelbrot determined this fractal dimension for the British coastline by

showing that for a decreasing unit of measurement, the length of a natural coastline does not converge but, instead, increases monotonically. On plotting the length L as a function of the unit of measurement ϵ on a log-log plot, he found a simple relation of the form:

$$L = \epsilon^{(1-D)} \tag{3.2}$$

where the real number D associated with every coastline is the dimension of the coastline. This study marked the origin of fractal geometry [58].

Self-Affine Fractals

There are two main categories of fractals: self-similar and self-affine. Fractal structures are said to be self-similar if a small piece is cut out of a fractal and magnified isotropically, both the original and the magnification look the same. Natural fault surfaces have been shown to possess self-similar features, where equal magnification occurs in all directions [59, 62].

However, not all fractal objects are self-similar. Many objects in nature exhibit unequal scaling magnification in different directions, known as anisotropic magnification, leading to the concept of self-affine geometry as defined by the following relationship [59, 63]:

$$x \to \lambda(x) \tag{3.3}$$

$$Z(x) \mapsto \sigma Z(x) \tag{3.4}$$

where σ is a homogeneous function of λ. The homogeneity index H, known as the roughness Hurst exponent or Holder condition is given as:

$$\sigma = \lambda^H \tag{3.5}$$

The Hurst exponent ranges between 0 and 1. In the case of self-similar geometry, H is equal to 1, indicating equal scaling factors in all directions.

An example of a self-affine fractal is fractional Brownian motion (fBm), which expands upon the concept of Brownian motion. Brownian motion describes the erratic movement of small particles suspended in a liquid. The function $Z(t)$ represents the x-location of a particle undergoing Brownian motion as a function of time t. While the curve of $Z(t)$ exhibits similarities to fractals, it is not self-similar but self-affine. This is demonstrated by the statistical properties of the Brownian function, where the scaling factor differs between time and position, indicating self-affinity.

$$Z(\gamma t) = \gamma^{l/2}\left(\overline{B}(t)\right) \tag{3.6}$$

where $\overline{B}(t)$ represents the probability distribution of the variable B at time t.

An fBm is a generalization of Brownian motion, and its statistical properties are described by:

$$[Z(t_2) - Z(t_1)] \propto |t_2 - t_1|^{2H} \tag{3.7}$$

This equation demonstrates the scaling relationship between increments of the fBm function, where H represents the Hurst exponent. When H= 0.5, the fBm reduces to standard Brownian motion.

The statistical self-affinity of fBm functions makes them fractal functions. The fractal dimension D of an fBm is related to the Hurst exponent H and the dimension E of the Euclidean space as $D = E + 1 - H$ [59, 64, 63].

Inverse Power Law Method Derivation

In 1985 Mandelbrot showed that the lengths of self-affine fractal curves do not follow the relation $L = \epsilon^{(1-D)}$ for all values of ϵ and therefore the dimension of self-affine curves cannot be obtained by measuring their lengths, but instead is obtained from their power spectra [59, 62].

For these self-affine fractal surfaces, the scaling of any measure of this irregularity is the power law, the exponent of which defines or is related to the fundamental measure, the fractal dimension D. The general power law equation for a power spectrum $S(v)$ that follows a power law is given in equation 3.8 [64].

$$S(v) = \frac{1}{v^b} = \frac{1}{v^{2H+E}}, \quad b = 2H + E \tag{3.8}$$

This relationship highlights the self-affinity of the power spectrum and is used in the fractal analysis of the radial power spectral density plots of perovskite thin films as explained below.

3.2.2 Inverse Power Law Method Analysis

For SEM images, the constant-slope region of the log RPSD plot characterizes the self-similar regime of the thin film. This region was analyzed using the inverse power law or fractal model [51, 59, 65, 66, 67, 68, 64, 69, 70].

The inverse power law $S(x)$ is

$$S(x) = \frac{a}{x^b} \tag{3.9}$$

The fractal dimension D, is related to the gradient of the log-log RPSD graph b, using [64, 69, 63]:

$$D_s = \frac{1}{2}(8 - b) = \frac{1}{2}\left(8 - (E + 2H)\right), \quad E = 2 \tag{3.10}$$

where b is the exponent of the slope and is defined as $b = E + 2H$, with E being the Euclidean dimension of the Fourier transform used $E = 2$, and H being the roughness scaling factor or Hurst exponent. The Hurst exponent H was derived for self-affine surfaces in Appendix 3.2.1.

From equation 3.10, the fractal dimension D can be used to derive the Hurst exponent H through the following relation:

$$D_{\mathrm{s}} = 3 - H \tag{3.11}$$

where $D = D_{\mathrm{s}} - 1$ for an isotropic and homogeneous surface [59, 63].

The fractal dimension D describes the irregularity of the surface roughness, *i.e*. higher fractal dimension corresponds to a rougher fracture surface [71]. The

value of the fractal dimension D is between 1 and 2 and surface fractal dimension D_s values range from 2 to 3 [56]. Therefore, only power-law regimes with slopes between -2 and -4 satisfy these relationships.

The Hurst exponent describes the process used in forming the measured surface and it usually varies between 0 and 1. The exponent indicates the spatial distribution of the roughness of a surface; if H varies between 0 and 0.5, it means that the spatial frequency of the surfaces is continuously alternating between high and low values, which indicates a uniform roughness distribution. However, when H is greater than 0.5, it indicates spatial frequencies with positive autocorrelation, meaning there are sudden changes in surface features and nonuniform distribution of roughness [53].

The fractal dimension D and Hurst exponent H were used to quantitatively analyze the surface roughness and morphology to compare different surfaces. Functions were set up in Python to fit the log-log RPSD plots to the inverse power law and to derive the fractal dimension and Hurst exponent using Equation 3.10. Fig. 3.1 shows the fitted plots for two $CsPbBr_3$ perovskite thin film images of the same magnification.

Table 3.1: Inverse power law fit parameters

Image Name	***b*** (Slope Exponent)	***D*** (Fractal Dimension)	***H*** (Hurst Exponent)
Image A	2.73	2.63	0.37
Image B	2.06	3.97	0.03

Based on Fig. 3.1, Table 3.1 gives the values of the D and H. The power law

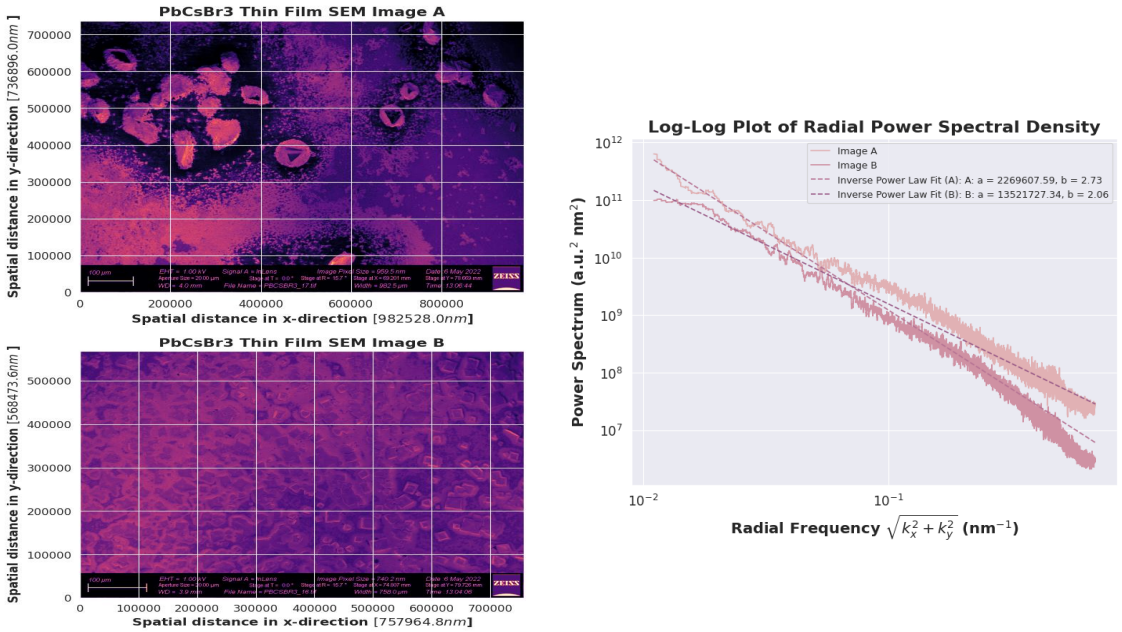

Figure 3.1: Inverse power law fitted log-log RPSD plots of Image A and Image B of different $CsPbBr_3$ thin films.

model was fit to the entire frequency range. From these roughness parameters, we can intuit that the thin film represented by image A has greater non-uniformly rough features compared to the thin film in image B.

From Fig. 3.1, it can be seen that the log-log RPSD plots aren't fitted especially well by the inverse power law, especially at high frequencies. Most thin films have both periodic and self-affine characteristics, which are reflected in the RPSD plots. For more homogeneous thin films, the higher frequency regions are flatter with small peaks representing the periodicity of the grains as is shown in Fig. 2.1(f). To account for this periodicity, the inverse power law is fitted only to the low-frequency, constant slope regions of the log-log RPSD plot going forward.

3.2.3 *K*-correlation Model Fit and Analysis

The high-frequency regions of the log-log RPSD plot were fit with the *k*-correlation model [51, 53, 65, 72, 73]. Taking the inverse Fourier transform of the height-height correlation function gives the power spectral density plot as explained in Section 2.2.2 [74]. The log-log PSD plots of self-affine thin films are empirically fitted using the *k*-correlation model [75, 76]. The deviation from the constant-slope power law at high frequencies is represented by A, B, and C, the fitting constants of the *k*-correlation function. The *k*-correlation model is

$$S(x) = \frac{A}{(B^2 \cdot x^2)^{(C+1)/2}} \tag{3.12}$$

The peak or knee at the high-frequency region of the log-log PSD of a self-affine thin film is quantitatively described by parameter B. The value of B indicates the location of the knee and it is related to radially-averaged correlation length and grain size. It describes the transition from a low to a high spatial frequency region. At higher spatial frequencies, the slope is described by C, and in that region, the surface is said to be fractal. Parameter A, also known as the shoulder parameter, is the value of the spectrum at low frequency where the power spectrum is nearly flat [53, 72]. These fitting parameters describe the topography of the surface and are specifically used to find the correlation length τ which describes the size distribution of the surface grains:

$$\tau^2 = \frac{(C-1)^2 B^2}{2\pi^2 C} \tag{3.13}$$

The *k*-correlation model equation was used to fit the high-frequency region of the log-log RPSD plot of the same $CsPbBr_3$ thin film analyzed in previous sections as shown in Fig. 3.2.

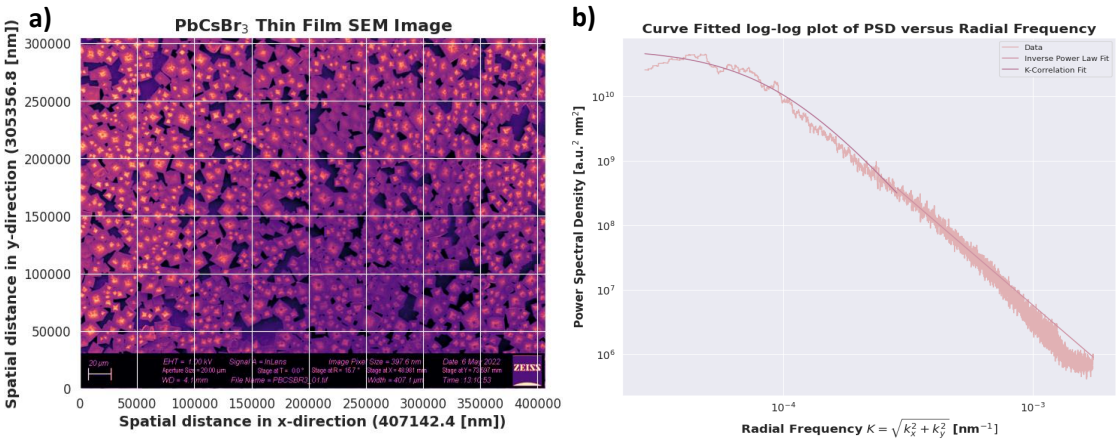

Figure 3.2: K-correlation model fitted log-log RPSD plot of a $CsPbBr_3$ thin film

The k-correlation model was fit to the image in Fig. 3.2 and the values of the parameters A, B, and C were used to calculate the correlation length. These values are given in table 3.2.

Table 3.2: K-correlation model fit parameters.

A [a.u^{2}nm^2] (Shoulder parameter)	B [nm] (Knee location)	C (Slope)	τ [nm] (Surface Correlation Length)
5.46e+10	10435	3.5	3140

Additionally, the low and high-frequency regions need to be adjusted when fitting the log-log RPSD plots depending on the shape of the plot. This created a scaling issue and the correlation length values calculated using the k-correlation parameters were affected based on how good the curve fitting is. To avoid this human error, the radial autocorrelation function was analyzed to get spatial distance values, such as the correlation length and grain size of the thin film. The analysis of these images and their associated parameters is discussed in Section 4.

3.2.4 Autocorrelation Function Analysis

The 1-D radial autocorrelation function is a mathematical tool used to analyze the spatial arrangement or texture of an image. The autocorrelation function measures the similarity between a signal and a delayed version of itself at different distances or lags [48, 54, 77, 78]. For a homogeneous thin film, the autocorrelation function typically shows a decay from its maximum value at zero distance to a relatively constant or small value as the distance increases. The rate of decay and the range of distances over which the correlation persists depend on the film's micro-structure and the characteristic length scales of its surface features. In the case of an image, it quantifies how similar the intensity values of pixels are as a function of their separation distance [79].

Mathematically, the discrete autocorrelation function $C_{k',\ell'}$ equation is derived in Section 2.2.2, equation 2.5. For rough, self-affine surfaces, such as those of perovskite thin films, the normalized $C_{k',\ell'}$ is maximum when the displacement is zero i.e. $k', \ell' = 0$. It has an exponential decay represented by equation 3.14

$$C_{k',\ell',\beta} = \exp - \left(\frac{(k', \ell')}{\beta} \right) \tag{3.14}$$

Additionally, the decay in the autocorrelation function was also used to quantify the surface morphology by the correlation length γ. The correlation length γ is the length at which the value of the autocorrelation function drops to $1/e$ as established in literature [74, 80]. Using equation 3.14, the correlation length corre-

sponds to:

$$\gamma = \beta \tag{3.15}$$

This is based on the assumption that the correlation length is taken when two points on a function have just reached the condition where they can be regarded as being independent. This follows from the fact that when $C_{k',\ell'}$ is close to unity, two points on the function at a distance $C_{k',\ell'}$ apart are strongly interdependent. However, when $C_{k',\ell'}$ attains values close to zero, two points on the function at a distance $C_{k',\ell'}$ apart are weakly correlated. A shorter correlation length γ indicates smaller features, while a longer correlation length γ suggests the presence of larger-scale features. [54].

Since the radial autocorrelation function is a measure of the displacement in the x and y direction, equation 3.14 and the derived correlation length γ can be applied to the plot. Using this mathematical basis, the normalized radial autocorrelation plots generated from the code outlined in Section 2.3.5 were analyzed.

Fig.3.3 (b) shows the normalized radial autocorrelation plot of an artificial image of concentric circles with a radius of 50 nm. From the plot, it can be seen that the distance at which the first peak occurs is related to the characteristic average grain size. This analysis was applied to the radial autocorrelation function of the $CsPbBr_3$ thin film in Fig. 3.3 (d) to get the average grain size as well.

The average grain size and correlation length can be used to quantify the size and spacing of grains in the thin film. Generally, the larger the grain size, the larger the separation distance at which the first peak appears as is seen in the case

of the radial autocorrelation function of the $CsPbBr_3$ thin film Image A shown in Fig. 3.4 (b). This is also shown by the larger correlation length of Image A versus Image B.

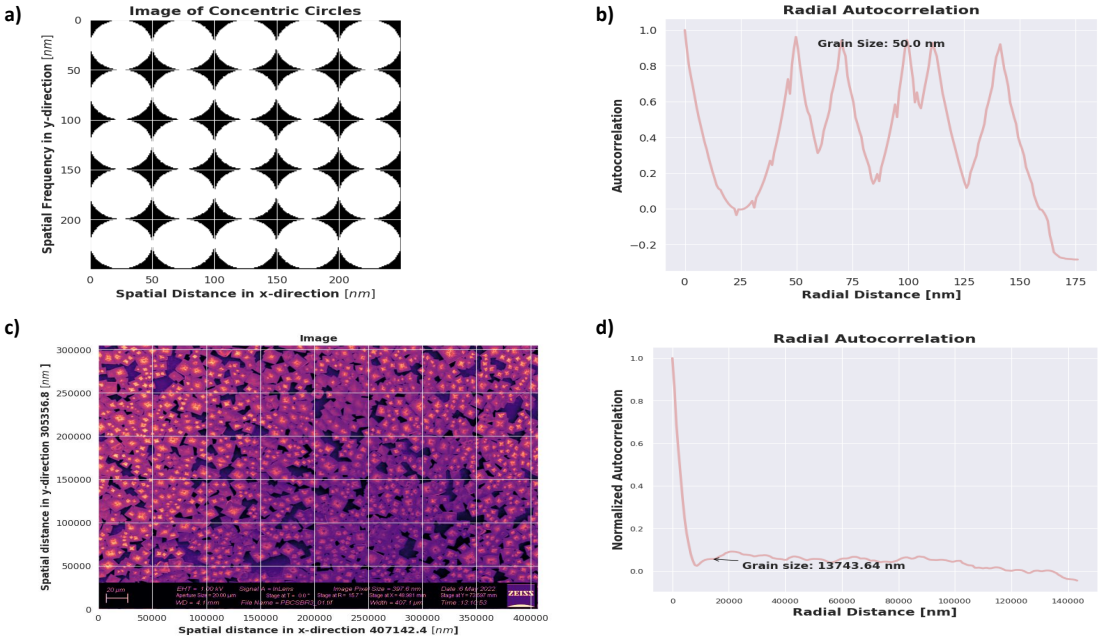

Figure 3.3: Normalized radial autocorrelation function plots for a test case circles image and a thin film image of a lead-halide perovskite film. (a) Images of concentric circles with a radius of 50 nm. (b) Radial autocorrelation function of (a) with its average grain size indicated. (c) Image of $CsPbBr_3$ thin film (d) Radial autocorrelation function of $CsPbBr_3$ thin film image with its average grain size indicated.

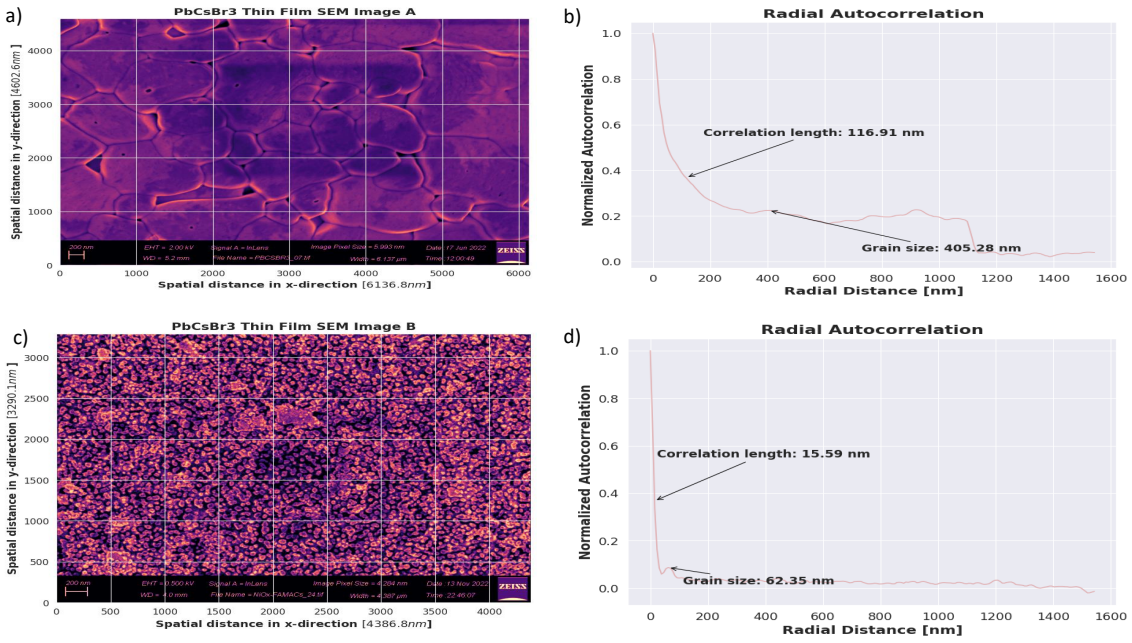

Figure 3.4: Comparison of correlation length γ between two SEM images of similar magnification. (a) $CsPbBr_3$ thin film Image A with larger grains. (b) Radial autocorrelation function of Image A with its average grain size and correlation length. (c) $CsPbBr_3$ thin film Image B with smaller grains. (d) Radial autocorrelation function of Image A with its average grain size and correlation length indicated.

Chapter 4

Results and Discussion

4.1 Results and Discussion

A total of 39 SEM images of $CsPbBr_3$ thin films, taken on the Zeiss Gemini 500 SEM, were analyzed. These films were synthesized using different solvents and processing conditions. The resulting thin films had varied surface morphology and features that were reflected in the varied radial PSD profiles.

The *k*-correlation model was fit to the high-frequency region of the log-log RPSD plot. The *k*-correlation fitting parameters, such as the shoulder parameter A, the knee location parameter B, and the slope C were used to find the correlation length τ. The power law model uses the slope exponent b to calculate the surface fractal dimension D and the Hurst exponent H. Additionally, the radial autocorrelation function was also used to find the correlation length γ and the average grain size. The surface morphology parameters for each image are presented in Table 4.1.

Generally, the high-frequency regime varies more resulting in greater fluctuations in the *k*-correlation parameters compared to the power law exponent. Hence, we find that the Hurst exponent and fractal dimension are more reliable measures of surface roughness and offer a better comparison between images. The thin

films were sorted based on their Hurst exponent values from low to high as the non-uniform surface roughness increases in Table 4.1.

Table 4.1: Surface roughness parameters of 39 $CsPbBr_3$ thin film SEM images

Image	A [a.u.$^2\mu$m^2]	B [μm]	C	τ [μm]	b	D	H	γ [μm]	Grain Size [μm]
1	2.87e+09	76.79	2.27	14.60	1.70	3.15	-0.15	6.33	35.70
2	1.14e+08	44.15	1.10	0.96	1.98	3.01	-0.01	3.27	17.00
3	1.29e+11	10467.5	0.83	431.00	2.29	2.86	0.14	2.15	18.50
4	4.34e+06	11.59	0.57	1.47	2.50	2.75	0.25	0.81	8.48
5	1.09e+04	0.79	2.73	0.19	2.52	2.74	0.26	0.25	0.76
6	9.34e+06	6.89	2.51	1.48	2.56	2.72	0.28	1.78	10.60
7	1.84e+05	2.66	1.72	0.33	2.63	2.68	0.32	0.47	1.77
8	5.70e+06	9.19	1.99	1.45	2.69	2.66	0.34	2.13	10.60
9	5.30e+05	4.71	0.97	0.04	2.72	2.64	0.36	0.48	2.02
10	6.79e+05	9.08	0.25	3.07	2.74	2.63	0.37	0.53	4.62
11	4.89e+14	10509.6	1.74	133.00	2.77	2.62	0.38	12.90	240.00
12	1.29e+11	3139.11	0.83	13.20	2.78	2.61	0.39	1.22	6.09
13	1.81e+04	0.23	30.00	0.27	2.83	2.59	0.41	0.36	1.78
14	1.65e+14	4662.64	1.84	65.10	2.95	2.53	0.47	2.73	8.20
15	1.82e+03	0.75	0.49	0.12	2.98	2.51	0.49	0.26	0.52
16	7.82e+05	1.13	1.46	0.10	3.09	2.46	0.54	0.23	1.97
17	2.28e+05	1.19	2.28	0.23	3.10	2.45	0.55	0.25	1.46

Table 4.1: (continued)

Image	A [a.u.$^2\mu$m^2]	B [μm]	C	τ [μm]	b	D	H	γ [μm]	Grain Size [μm]
18	7.48e+06	3.64	0.77	0.21	3.12	2.44	0.56	0.41	2.07
19	1.69e+06	1.02	5.00	0.41	3.22	2.39	0.61	0.53	3.02
20	2.00e+08	44.31	2.55	9.67	3.23	2.38	0.62	10.50	40.70
21	4.60e+07	56.77	1.49	5.12	3.29	2.36	0.64	13.50	43.10
22	2.36e+05	2.62	0.90	0.07	3.29	2.35	0.65	0.36	2.31
23	2.01e+05	1.45	0.63	0.15	3.39	2.30	0.70	0.08	0.59
24	5.45e+06	4.68	3.05	1.24	3.40	2.30	0.70	1.66	6.20
25	1.28e+13	3660.55	1.33	236.00	3.50	2.25	0.75	0.89	6.04
26	3.83e+11	2959.49	1.13	82.80	3.51	2.24	0.76	0.89	6.04
27	5.46e+07	10.44	3.50	3.14	3.56	2.22	0.78	3.62	13.70
28	4.29e+05	3.08	1.79	0.41	3.58	2.21	0.79	0.48	2.83
29	1.89e+05	0.43	10.00	0.28	3.64	2.18	0.82	0.36	1.24
30	3.14e+08	168.69	0.96	1.49	3.67	2.17	0.83	0.16	0.57
31	3.66e+05	1.97	2.58	0.44	3.70	2.15	0.85	0.53	1.78
32	2.41e+05	1.09	3.61	0.34	3.70	2.15	0.85	0.46	1.81
33	1.75e+04	0.66	4.45	0.24	3.71	2.14	0.86	0.32	1.55
34	1.34e+06	3.17	1.17	0.11	3.81	2.10	0.90	0.09	0.37
35	2.85e+05	1.80	2.71	0.42	3.86	2.07	0.93	0.53	4.44
36	1.11e+05	0.60	2.19	0.11	3.91	2.05	0.95	0.10	0.48

Table 4.1: (continued)

Image	A [a.u.$^2\mu$m^2]	B [μm]	C	τ [μm]	b	D	H	γ [μm]	Grain Size [μm]
37	1.81e+06	1.97	2.64	0.45	3.96	2.02	0.98	0.71	2.77
38	8.87e+03	0.34	1.71	0.04	3.99	2.01	0.995	0.28	0.46
39	2.80e+03	0.04	5.50	0.02	4.00	2.00	1.00	0.02	0.06

Fig. 4.1 displays the images with the lowest and the highest Hurst exponent values. The images are numbered corresponding to their numbering in table 4.1.

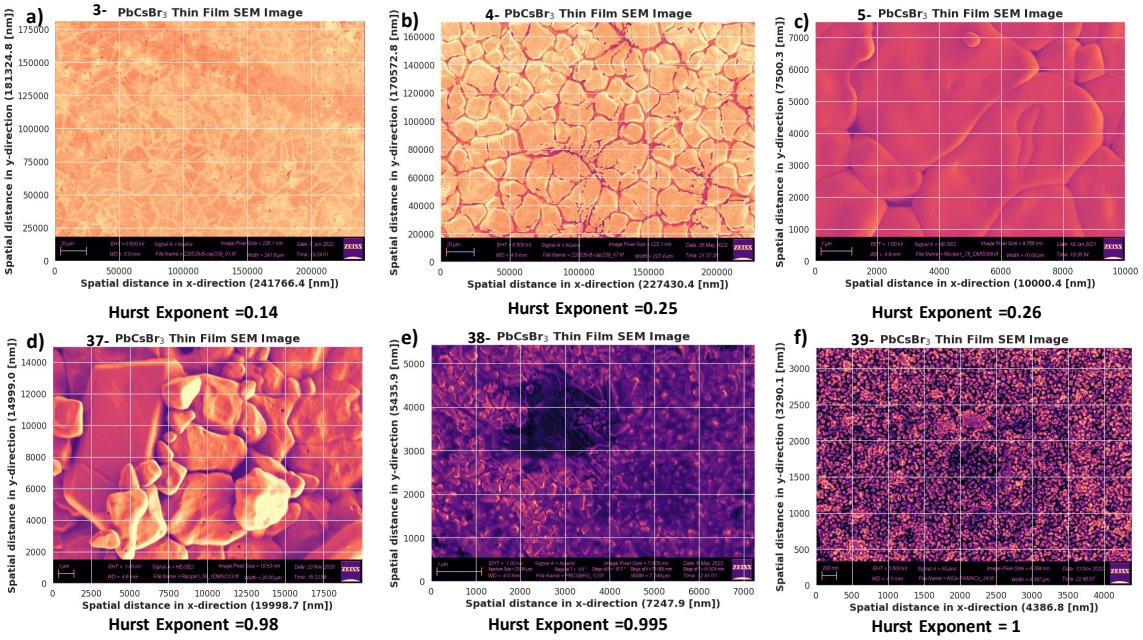

Figure 4.1: $CsPbBr_3$ thin film images with the lowest and highest Hurst exponent values based on Table 4.1. (a) Image 3, Hurst exponent = 0.14. (b) Image 3, Hurst exponent = 0.25. (c) Image 4, Hurst exponent = 0.26. (d) Image 37, Hurst exponent = 0.98. (e) Image 38, Hurst exponent = 0.995. (f) Image 39, Hurst exponent = 1.

Based on the analysis of Table 4.1 and Fig. 4.1, films with prominent surface features and more defined grains have a knee region with a pronounced peak and the correlation length increases as the size distribution of the grains increases. The knee region flattens as the height variation decreases in the samples as well. For more homogeneous, uniform films, this knee region gets smaller and the power spectra tend towards a straighter slope, more closely following the power law model. The slope exponent b was used to calculate the Hurst exponent, which measures the extent of uniform roughness in the film. We find that homogeneous films with similar surface features, and fewer pinholes, have the lowest Hurst exponent. The exponent increases with an increase in film roughness, and non-uniformity, and is generally large for images with varied grain sizes.

Outliers such as very non-uniformly rough surfaces, especially those with a 100-micron length scale have log-log RPSD plots that aren't fit well by using either the power law and k-correlation model. In these cases, only the power law fit is used. An example of this is Fig. 3.1. In this case, the correlation length was calculated using the radial autocorrelation function. The Hurst exponent values don't seem to be comparable to the rest of the images and don't follow the general trend. Some outliers have negative Hurst exponent values, as seen for Images 1 and 2 in Table 4.1. In these cases, the log-log RPSD plots follow the power law at the high-frequency region and the k-correlation model for the low-frequency region so the curve fits were bad.

Moreover, since the log-log RPSD plot shape varies, the low and high-

frequency regions are defined arbitrarily to allow for the best curve fit [53]. The correlation length τ calculated using the k-correlation model parameters fluctuates as the curve fit changes based on the range of frequencies the model is fit for. This frequency range is set to allow for the best curve fit, where k-correlation model parameter values aren't unphysical. For example, the values of A and B and C cannot be negative and since C is an exponent, its value can't be too large. The frequency range is also chosen such that the Hurst exponent H is between 0 to 1.

Additionally, k-correlation is an empirical model that is not derived mathematically. However, it does account for the knee in the high-frequency region of the PSD that measures the surface spatial feature size. On the other hand, the correlation length τ found mathematically using the radial autocorrelation function is more rigorously derived and is not dependent on the manual curve fitting. There is a need for further analysis of the correlation length to understand its physical implications in this application and to determine the best method to derive it.

Furthermore, from this analysis, it was observed that the comparison of surface roughness parameters is most accurate when done for images of similar magnification. As the magnification decreases (image pixel size increases), the resolution of surface features drops and the Hurst exponent doesn't change as drastically. So, for higher magnification images, the Hurst exponent (H) has comparable values to rougher low magnification images.

The initial image analysis reported here needs to be extended to get better com-

parisons between different images, study the effects of contrast, and magnification on the image power spectra, and compare to existing literature. Our observations and surface roughness parameters need to be validated using device-quality thin films. The device performance data for these films will allow us to understand the significance of these parameters. Using machine learning algorithms.

4.2 Conclusions

In conclusion, a versatile code for image reduction has been developed, enabling the extraction of various important image features such as the 2-D FFT, 2-D PSD, 2-D autocorrelation function, 1-D radial PSD, and 1-D radial autocorrelation functions. This code is designed with efficiency in mind, allowing it to seamlessly loop through a database of image data. As a result, the code is scalable and adaptable to handle different sets of images.

In addition, curve fitting of the log-log radial power spectral density using the inverse power law and k-correlation model has been implemented for 39 images of $CsPbBr_3$ or $MAPbI_3$ thin films acquired from different scanning electron microscope (SEM) models, The Zeiss Gemini 500 and the FEI Quanta 650 Scanning Electron Microscope. By conducting the curve fitting on each individual image's log-log radial power spectral density plot, we were able to extract and store parameters such as the fractal dimension D, Hurst exponent H and the k-correlation model-based correlation length τ. Based on the values of these parameters, we

were able to compare the surface roughness of different SEM images.

The average grain size and correlation length γ were computed from the normalized radial autocorrelation function for each image. By analyzing and comparing these parameters, we were able to gain a better understanding of the perovskite thin film surface morphology and its spatial features such as grain size.

4.3 Further Work

The developed code has been extended to accommodate images acquired from various other scanning electron microscopes (SEMs). Minor adjustments can be made to the code to facilitate the extraction of metadata from atomic force microscopy (AFM) images. Radial PSDF and radial autocorrelation plots can be compared to the existing SEM images to validate surface morphology parameters. Further analysis of spatial functions to get the size distribution of the grains can be done as well.

To expand the scope of the research, we recommended exploring and identifying the parameters for a larger data set of images. This expanded analysis will contribute to the establishment of a comprehensive image database, encompassing spatial functions and corresponding roughness parameters. Finally, this data set can be integrated into the multi-scale machine learning algorithm to allow for better correlation of experimental data sources to facilitate ML predictions for per-

ovskite optimization.

BIBLIOGRAPHY

[1] J. J. Berry, J. Van De Lagemaat, M. M. Al-Jassim, S. Kurtz, Y. Yan, and K. Zhu, "Perovskite Photovoltaics: The Path to a Printable Terawatt-Scale Technology," *ACS Energy Letters*, vol. 2, pp. 2540–2544, Nov. 2017.

[2] M. A. Green, A. Ho-Baillie, and H. J. Snaith, "The emergence of perovskite solar cells," *Nature Photon*, vol. 8, pp. 506–514, July 2014.

[3] V. M. Goldschmidt, T. Barth, G. Lunde, and W. Zachariasen, "Die gesetze der krystallochemie. naturwissensenschaffen," *Norwegian Science and Technology Oslo, Mat.-Natural Sciences*, vol. 8, pp. 477–485, 1926.

[4] C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, and M. Scheffler, "New tolerance factor to predict the stability of perovskite oxides and halides," *Sci. Adv.*, vol. 5, p. eaav0693, Feb. 2019.

[5] C. Li, K. C. K. Soh, and P. Wu, "Formability of ABO3 perovskites," *Journal of Alloys and Compounds*, vol. 372, pp. 40–48, June 2004.

[6] Z. Li, M. Yang, J.-S. Park, S.-H. Wei, J. J. Berry, and K. Zhu, "Stabilizing Perovskite Structures by Tuning Tolerance Factor: Formation of Formamidinium and Cesium Lead Iodide Solid-State Alloys," *Chemistry of Materials*, vol. 1, pp. 284–292, Dec. 2015.

[7] O. Muller and R. R, "The Major Ternary Structural Families," *Berlin, Heidelberg, Springer,New York*, June 1974.

[8] L. A. Estroff, J. A. Marohn, R. Hernandez, P. Clancy, J. Choi, J. Tao, R. Lindsey, and D. Moore, "Personal communication. "formulation engineering of energy materials via multiscale learning spirals," *U.S. Department of Energy grant DE-SC0022305*, 2021.

[9] Y. G. Kim, T.-Y. Kim, J. H. Oh, K. S. Choi, Y.-J. Kim, and S. Y. Kim, "Cesium lead iodide solar cells controlled by annealing temperature," *The Royal Society of Chemistry*, vol. 19, no. 8, pp. 6257–6263, 2017.

[10] M. Saliba, T. Matsui, J.-Y. Seo, K. Domanski, J.-P. Correa-Baena, M. K. Nazeeruddin, S. M. Zakeeruddin, W. Tress, A. Abate, A. Hagfeldt, and M. Grätzel, "Cesium-containing triple cation perovskite solar cells: Improved stability, reproducibility and high efficiency," *Energy and Environmental Science*, vol. 9, no. 6, pp. 1989–1997, 2016.

[11] W. Shockley and H. J. Queisser, "Detailed Balance Limit of Efficiency of p-n Junction Solar Cells," *Journal of Applied Physics*, vol. 32, pp. 510–519, Mar. 1961.

[12] W. Xiang, S. F. Liu, and W. Tress, "A review on the stability of inorganic metal halide perovskites: Challenges and opportunities for stable solar cells," *Energy and Environmental Science*, vol. 14, pp. 2090–2113, Apr. 2021.

[13] C. C. Stoumpos, C. D. Malliakas, and M. G. Kanatzidis, "Semiconducting Tin and Lead Iodide Perovskites with Organic Cations: Phase Transitions, High Mobilities, and Near-Infrared Photoluminescent Properties," *Inorganic Chemistry*, vol. 52, pp. 9019–9038, Aug. 2013.

[14] H. N. Yang, G. C. Wang, and T. M. Lu, "Instability in Low-Temperature Molecular-Beam Epitaxy Growth of Si/Si(111)," *Physical Review Letters*, vol. 73, pp. 2348–2351, Oct. 1994.

[15] G. W. P. Adhyaksa, S. Brittman, H. Āboliņš, A. Lof, X. Li, J. D. Keelor, Y. Luo, T. Duevski, R. M. A. Heeren, S. R. Ellis, D. P. Fenning, and E. C. Garnett, "Understanding Detrimental and Beneficial Grain Boundary Effects in Halide Perovskites," *Advanced Materials*, vol. 30, p. 1804792, Dec. 2018.

[16] A. Walsh, D. O. Scanlon, S. Chen, X. G. Gong, and S.-H. Wei, "Self-Regulation Mechanism for Charged Point Defects in Hybrid Halide Perovskites," *Angewandte Chemie International Edition*, vol. 54, no. 6, pp. 1791–1794, 2015.

[17] A. Kojima, K. Teshima, Y. Shirai, and T. Miyasaka, "Organometal Halide Perovskites as Visible-Light Sensitizers for Photovoltaic Cells," *American Chemical Society*, vol. 131, pp. 6050–6051, Apr. 2009.

[18] Z. Chu, M. Yang, P. Schulz, D. Wu, X. Ma, E. Seifert, L. Sun, X. Li, K. Zhu,

and K. Lai, "Impact of grain boundaries on efficiency and stability of organic-inorganic trihalide perovskites," *Nature Communications*, vol. 8, pp. 1–8, Dec. 2017.

[19] N. D. Marco, H. Zhou, Q. Chen, P. Sun, Z. Liu, L. Meng, E.-P. Yao, Y. Liu, A. Schiffer, and Y. Yang, "Guanidinium: A Route to Enhanced Carrier Lifetime and Open-Circuit Voltage in Hybrid Perovskite Solar Cells," *Nano Letters*, vol. 16, pp. 1009–1016, Jan. 2016.

[20] B. Yang, O. Dyck, J. Poplawsky, J. Keum, A. Puretzky, S. Das, I. Ivanov, C. Rouleau, G. Duscher, D. Geohegan, and K. Xiao, "Perovskite Solar Cells with Near 100% Internal Quantum Efficiency Based on Large Single Crystalline Grains and Vertical Bulk Heterojunctions," *Journal of American Chemical Society*, vol. 137, pp. 9210–9213, July 2015.

[21] H. Nazem, "Modeling of j_{sc} and v_{oc} versus the grain size in CdTe, CZTS and Perovskite thin film solar cells," *Superlattices and Microstructures*, vol. 128, pp. 421–427, Apr. 2019.

[22] O. Shargaieva, "The influence of the grain size on the properties of $CH_3NH_3PbI_3$ thin films," *ACS Applied Materials Interfaces*, vol. 9, p. 38428–38435, 2017.

[23] A. A. Mamun, *Effect of Grain Size and Interface Engineering on the Photovoltaic Performance and Stability of Perovskite Solar Cells*. PhD thesis, Old Dominion University Libraries, 2021.

[24] R. Long, J. Liu, and O. V. Prezhdo, "Unravelling the Effects of Grain Boundary and Chemical Doping on Electron–Hole Recombination in $CH_3NH_3PbI_3$ Perovskite by Time-Domain Atomistic Simulation," *Journal of American Chemical Society*, vol. 138, pp. 3884–3890, Mar. 2016.

[25] A. A. Mamun, T. T. Ava, H. J. Jeong, M. S. Jeong, and G. Namkoong, "A deconvoluted PL approach to probe the charge carrier dynamics of the grain interior and grain boundary of a perovskite film for perovskite solar cell applications," *Physical Chemistry Chemical Physics*, vol. 19, pp. 9143–9148, Mar. 2017.

[26] G. Niu, W. Li, F. Meng, L. Wang, H. Dong, and Y. Qiu, "Study on the stability of $CH_3NH_3PbI_3$ films and the effect of post-modification by aluminum oxide in all-solid-state hybrid solar cells," *Journal of Material Chemistry A*, vol. 2, pp. 705–710, dec 2013.

[27] T. Leijtens, G. E. Eperon, N. K. Noel, S. N. Habisreutinger, A. Petrozza, and H. J. Snaith, "Stability of Metal Halide Perovskite Solar Cells," *Advanced Energy Materials*, vol. 5, p. 1500963, Oct. 2015.

[28] L. McGovern, I. Koschany, G. Grimaldi, L. A. Muscarella, and B. Ehrler, "Grain Size Influences Activation Energy and Migration Pathways in $MAPbBr_3$ Perovskite Solar Cells," *Journal of Physical Chemistry Letters*, vol. 12, pp. 2423–2428, Mar. 2021.

[29] D. T. Moore, H. Sai, K. W. Tan, D.-M. Smilgies, W. Zhang, H. J. Snaith, U. Wiesner, and L. A. Estroff, "Crystallization Kinetics of Organic–Inorganic Trihalide Perovskites and the Role of the Lead Anion in Crystal Growth," *J. Am. Chem. Soc.*, vol. 137, pp. 2350–2358, Feb. 2015.

[30] L. Zhang, M. He, and S. Shao, "Machine learning for halide perovskite materials," *Nano Energy*, vol. 78, p. 105380, Dec. 2020.

[31] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, and F. El-Mellouhi, "Learn-and-Match Molecular Cations for Perovskites," *Journal of Physical Chemistry A*, vol. 123, pp. 7323–7334, Aug. 2019.

[32] J. Yang, Y. Wang, T. Wu, and S. Li, "Correlating the Composition-Dependent Structural and Electronic Dynamics of Inorganic Mixed Halide Perovskites," *Chemistry of Materials*, vol. 32, pp. 2470–2481, Mar. 2020.

[33] R. Jacobs, G. Luo, and D. Morgan, "Materials Discovery of Stable and Nontoxic Halide Perovskite Materials for High-Efficiency Solar Cells," *Advanced Functional Materials*, vol. 29, p. 1804354, June 2019.

[34] W. A. Saidi, W. Shadid, and I. E. Castelli, "Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convo-

lutional neural network," *npj Computational Materials*, vol. 6, pp. 1–7, Apr. 2020.

[35] H. C. Herbol, W. Hu, P. Frazier, P. Clancy, and M. Poloczek, "Efficient search of compositional space for hybrid organic–inorganic perovskites via Bayesian optimization," *npj Computational Materials*, vol. 4, pp. 1–7, Sept. 2018.

[36] Ç. Odabaşı and R. Yıldırım, "Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools," *Nano Energy*, vol. 56, pp. 770–791, Feb. 2019.

[37] Z. Ren, F. Oviedo, H. Xue, M. Thway, K. Zhang, N. Li, J. D. Perea, M. Layurova, Y. Wang, S. Tian, T. Heumueller, E. Birgersson, F. Lin, A. Aberle, S. Sun, I. M. Peters, R. Stangl, C. J. Brabec, and T. Buonassisi, "Physics-guided characterization and optimization of solar cells using surrogate machine learning model," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, pp. 3054–3058, June 2019.

[38] A. Ali, H. Park, R. Mall, B. Aïssa, S. Sanvito, H. Bensmail, A. Belaidi, and F. El-Mellouhi, "Machine Learning Accelerated Recovery of the Cubic Structure in Mixed-Cation Perovskite Thin Films," *Chemistry of Materials*, pp. 2998–3006, mar 2020.

[39] Y. Zhang, D. Lu, M. Gao, M. Lai, J. Lin, T. Lei, Z. Lin, L. N. Quan, and P. Yang, "Quantitative imaging of anion exchange kinetics in halide perovskites," *Proceedings of the National Academy of Sciences*, vol. 116, pp. 12648–12653, June 2019.

[40] T. Chen, Y. Zhou, and M. Rafailovich, "Application of Machine Learning in Perovskite Solar Cell Crystal Size Distribution Analysis," *MRS Advances*, vol. 4, pp. 793–800, Mar. 2019.

[41] B. R. Jany, A. Janas, and F. Krok, "Retrieving the Quantitative Chemical Information at Nanoscale from Scanning Electron Microscope Energy Dispersive X-ray Measurements by Machine Learning," *Nano Letters*, vol. 17, pp. 6520–6525, Nov. 2017.

[42] J. Madsen, P. Liu, J. Kling, J. B. Wagner, T. W. Hansen, O. Winther, and J. Schiøtz, "A Deep Learning Approach to Identify Local Structures in Atomic-Resolution Transmission Electron Microscopy Images," *Advanced Theory and Simulations*, vol. 1, p. 1800037, Aug. 2018.

[43] Y. Liu, Q. Sun, W. Lu, H. Wang, Y. Sun, Z. Wang, X. Lu, and K. Zeng, "General Resolution Enhancement Method in Atomic Force Microscopy Using Deep Learning," *Advanced Theory and Simulations*, vol. 2, no. 2, p. 1800137, 2019.

[44] J. Hidalgo, A.-F. Castro-Méndez, and J.-P. Correa-Baena, "Imaging and Mapping Characterization Tools for Perovskite Solar Cells," *Advanced Energy Materials*, vol. 9, p. 1900444, 2019.

[45] W. Chen, Y. Wu, Y. Yue, J. Liu, W. Zhang, X. Yang, H. Chen, E. Bi, I. Ashraful, M. Grätzel, and L. Han, "Efficient and stable large-area perovskite solar cells with inorganic charge extraction layers," *Science*, vol. 350, pp. 944–948, Nov. 2015.

[46] W. Nie, H. Tsai, R. Asadpour, J.-C. Blancon, A. J. Neukirch, G. Gupta, J. J. Crochet, M. Chhowalla, S. Tretiak, M. A. Alam, H.-L. Wang, and A. D. Mohite, "High-efficiency solution-processed perovskite solar cells with millimeter-scale grains," *Science*, vol. 347, pp. 522–525, Jan. 2015.

[47] J. F. González Martínez, I. Nieto-Carvajal, J. Abad, and J. Colchero, "Nanoscale measurement of the power spectral density of surface roughness: How to solve a difficult experimental challenge," *Nanoscale Res Lett*, vol. 7, p. 174, Mar. 2012.

[48] P. R. Nayak, "Random Process Model of Rough Surfaces," *Journal of Lubrication Technology*, vol. 93, pp. 398–407, July 1971.

[49] W. M. Tong and R. S. Williams, "Kinetics of Surface Growth: Phenomenology, Scaling, and Mechanisms of Smoothening and Roughening," *Annual Review of Physical Chemistry*, vol. 45, no. 1, pp. 401–438, 1994.

[50] J. S. Bendat and A. G. Piersol, "Engineering applications of correlation and spectral analysis," *Wiley, New York*, 1980.

[51] Y. Gong, S. T. Misture, P. Gao, and N. P. Mellott, "Surface Roughness Measurements Using Power Spectrum Density Analysis with Enhanced Spatial Correlation Length," *The Journal of Physical Chemistry C*, vol. 120, pp. 22358–22364, Oct. 2016.

[52] A. Duparré, J. Ferre-Borrull, S. Gliech, G. Notni, J. Steinert, and J. M. Bennett, "Surface characterization techniques for determining the root-mean-square roughness and power spectral densities of optical components," *Applied Optics*, vol. 41, pp. 154–171, jan 2002.

[53] F. M. Mwema, E. T. Akinlabi, O. P. Oladijo, and O. P. Oladijo, "The Use of Power Spectrum Density for Surface Characterization of Thin Films," in *Photoenergy and Thin Film Materials* (X.-Y. Yang, ed.), pp. 379–411, Wiley, 1 ed., Mar. 2019.

[54] B. Bhushan, "Surface Roughness Analysis and Measurement Techniques," in *Modern Tribology Handbook, Two Volume Set*, pp. 79–150, CRC Press, Dec. 2000.

[55] R. Gavrila, A. Dinescu, and D. Mardare, "A Power Spectral Density Study of Thin Films Morphology Based on AFM Profiling," *Romanian Journal of Information Science and Technology*, vol. 10, pp. 291–300, Jan. 2007.

[56] C. J. Buchko, K. M. Kozloff, and D. C. Martin, "Surface characterization of porous, biocompatible protein polymer thin films," *Biomaterials*, vol. 22, pp. 1289–1300, June 2001.

[57] A. van der Schaaf and J. H. van Hateren, "Modelling the Power Spectra of Natural Images: Statistics and Information," *Vision Research*, vol. 36, pp. 2759–2770, Sept. 1996.

[58] B. B. Mandelbrot and J. A. Wheeler, "The Fractal Geometry of Nature," *American Journal of Physics*, vol. 51, pp. 286–287, Mar. 1983.

[59] A. Majumdar and B. Bhushan, "Role of Fractal Geometry in Roughness Characterization and Contact Mechanics of Surfaces," *Journal of Tribology*, vol. 112, pp. 205–216, Apr. 1990.

[60] D. Raoufi, "Fractal analyses of ITO thin films: A study based on power spectral density," *Physica B: Condensed Matter*, vol. 405, pp. 451–455, Jan. 2010.

[61] W. R. Bowen and T. A. Doneva, "Artefacts in AFM studies of membranes: Correcting pore images using fast fourier transform filtering," *Journal of Membrane Science*, vol. 171, pp. 141–147, June 2000.

[62] B. B. Mandelbrot, "Self-Affine Fractals and Fractal Dimension," *Physica Scripta*, vol. 32, pp. 257–260, Oct. 1985.

[63] J. C. Russ, *Fractal Surfaces*. Boston, MA: Springer US, 1994.

[64] C. C. Barton and C. H. Scholz, "The Fractal Size and Spatial Distribution of Hydrocarbon Accumulations," in *Fractals in Petroleum Geology and Earth Processes*, pp. 13–34, Springer, Boston, MA, 1995.

[65] M. Senthilkumar, N. K. Sahoo, S. Thakur, and R. B. Tokas, "Characterization of microroughness parameters in gadolinium oxide thin films: A study based on extended power spectral density analyses," *Applied Surface Science*, vol. 252, pp. 1608–1619, Dec. 2005.

[66] F. M. Mwema, O. P. Oladijo, T. S. Sathiaraj, and E. T. Akinlabi, "Atomic force microscopy analysis of surface topography of pure thin aluminum films," *Materials Research Express*, vol. 5, p. 046416, apr 2018.

[67] W. Kwaśny, L. A. Dobrzański, M. Pawlyta, and W. Gulbiński, "Fractal nature of surface topography and physical properties of the coatings obtained using magnetron sputtering," *Journal of Materials Processing Technology*, vol. 157–158, pp. 188–193, Dec. 2004.

[68] R. P. Yadav, S. Dwivedi, A. K. Mittal, M. Kumar, and A. C. Pandey, "Fractal and multifractal analysis of LiF thin film surface," *Applied Surface Science*, vol. 261, pp. 547–553, Nov. 2012.

[69] S. Ganti, "Generalized fractal analysis and its applications to engineering surfaces," *Wear*, vol. 180, pp. 17–34, Jan. 1995.

[70] S. J. Fang, S. Haplepete, W. Chen, C. R. Helms, and H. Edwards, "Analyzing atomic force microscopy images using spectral methods," *Journal of Applied Physics*, vol. 82, pp. 5891–5898, Dec. 1997.

[71] T. Babadagli and K. Develi, "Fractal characteristics of rocks fractured under tension," *Theoretical and Applied Fracture Mechanics*, vol. 39, pp. 73–88, Jan. 2003.

[72] S. Karan and B. Mallik, "Power spectral density analysis and photoconducting behavior in copper(II) phthalocyanine nanostructured thin films," *Physical Chemistry Chemical Physics*, vol. 10, pp. 6751–6761, Dec. 2008.

[73] S. J. Moody, M. R. Phillips, and M. Toth, "Assessment of SEM Image Quality using 1D Power Spectral Density Estimation," *Microscopy and Microanalysis*, vol. 15, pp. 48–49, July 2009.

[74] G. Palasantzas, "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," *Phys. Rev. B*, vol. 48, pp. 14472–14478, Nov. 1993.

[75] E. L. Church, "Fractal Surface Finish," in *Grazing Incidence Optics for Astronomical and Laboratory Applications*, vol. 0830, pp. 75–85, SPIE, Aug. 1988.

[76] E. L. Church, P. Z. Takacs, and T. A. Leonard, "The Prediction Of BRDFs From Surface Profile Measurements," in *33rd Annual Techincal Symposium* (J. C. Stover, ed.), (San Diego), p. 136, jan 1990.

[77] J. M. Bennett, "Measurement of the rms roughness, autocovariance function and other statistical properties of optical surfaces using a FECO scanning interferometer," *Appl. Opt., AO*, vol. 15, pp. 2705–2721, Nov. 1976.

[78] J. C. Stover, "Roughness characterization of smooth machined surfaces by light scattering," *Appl. Opt., AO*, vol. 14, pp. 1796–1802, Aug. 1975.

[79] J. M. Bennett, "Characterization of Surface Roughness," in *Light Scattering and Nanoscale Surface Roughness*, pp. 1–33, Springer, Boston, MA, 2007.

[80] C. Liu, Q. Dong, H. Li, Z. Li, X. Li, and C. Cheng, "Measurement of surface parameters from autocorrelation function of speckles in deep Fresnel region with microscopic imaging system," *Optics Express*, vol. 22, p. 1302, jan 2014.