PROCESS-STRUCTURE LINKAGES WITH MATERIALS KNOWLEDGE SYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

David B. Brough

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computational Science and Engineering

Georgia Institute of Technology
December 2016

PROCESS-STRUCTURE LINKAGES WITH MATERIALS KNOWLEDGE SYSTEMS

Approved by:

Professor Surya Kalidindi, Advisor
The George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Professor Srinivas Aluru
School of Computational Science and Engineering
Georgia Institute of Technology

Professor Hamid Garmestani
School of Materials Science and Engineering
Georgia Institute of Technology

Professor Martha A. Grover
School of Chemical and Biomoleculer Engineering
Georgia Institute of Technology

Professor Hongyuan Zha
School of Computational Science and Engineering
Georgia Institute of Technology

Date Approved: 23 Aug 2016

To my wife,

Kendree,

for her invaluable

support and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Process-Structure Linkages with
Materials Knowledge Systems

David B. Brough

124 Pages

Directed by Professor Surya Kalidindi

The search for optimal manufacturing process routes that results in the combination of desired properties for any application is a highly dimensional optimization problem due to the hierarchical nature of material structure. Yet, this problem is a key component to materials design. Customized materials data analytics provides a new avenue of research in the efforts to address the challenge described above, while accounting for the inherently stochastic nature of the available data. The analytics mine and curate transferable, high value, materials knowledge at multiple length and time scales. More specifically, this materials knowledge is cast in the form of Process-Structure-Property (PSP) linkages of interest to the design/manufacturing experts.

The extension of the novel Materials Knowledge Systems (MKS) framework to Process-Structure linkages holds the exciting potential to development full PSP linkages that can be can be leveraged by experts in data science, manufacturing and materials science and engineering communities. PSP linkages are an essential component in the to realize a modern accelerated materials innovation ecosystem. This work describes the methodologies used to extend the MKS framework to Process-Structure linkages and demonstrates their utility.

CHAPTER I

INTRODUCTION

Customized materials design has great potential for impacting virtually all emerging technologies, with significant economic consequences [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. However, materials design (including the design of a manufacturing process route) resulting in the combination of properties desired for a specific application is a highly challenging inverse problem due to the hierarchical nature of materials structure. Material properties are controlled by the hierarchical internal structure (over multiple length scales which spans from atomic to macroscopic) as well as coupled physical phenomena which can occur at different timescales at each of the hierarchical length scales. Characterization of the structure at each of these different length scales is often in the form of images which come from different experimental/computational techniques resulting in highly heterogeneous data as illustrated in Figure 1. As a result, tailoring the material hierarchical structure to yield desired combinations of properties or performance characteristics is enormously difficult.

Several recent national and international initiatives [15, 6, 7] have been launched with the premise that the adoption and utilization of modern informatics and data science offers a new opportunity to dramatically accelerate the design and deployment cycle of new advanced materials in commercial products. There is a critical need for customized data science approaches that take into account the stochastic nature of materials data at multiple length scales in order to extract relevant and transferable materials knowledge. Data driven and computationally cheap Process-Structure-Property (PSP) linkages can exact valuable materials knowledge from these heterogeneous datasets and provide a format that is usable and valuable for design

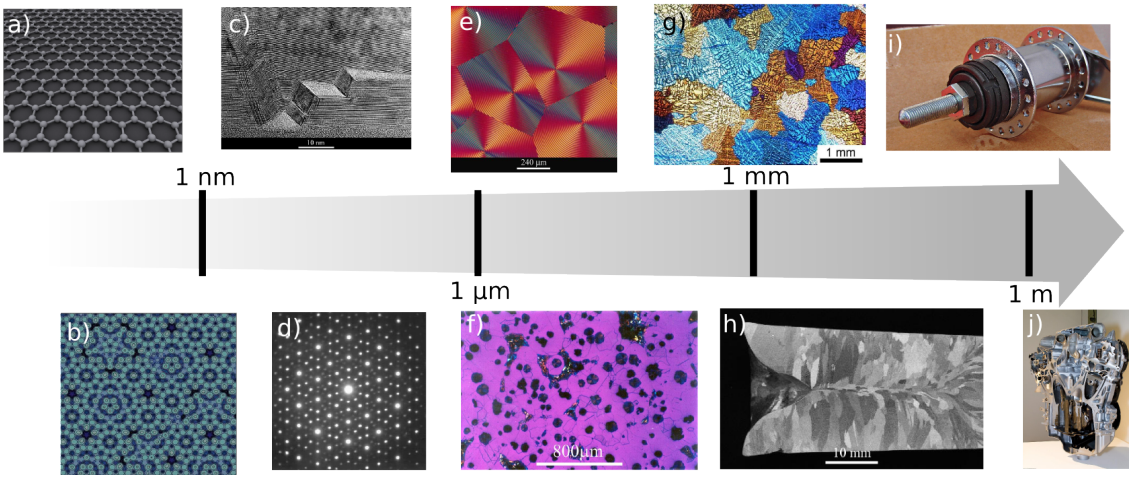


Figure 1: Heirarchical Materials Structure at Multiple Length Scales a). Idealized graphene crystalline structure. b). Atomic model for a fivefold icosahedral Al-Ag quasicrystals. c). Delamination cracks in h-BN particles subjected to compressive stress in the (0001) planes (within a silicon nitride particulate-reinforced silicon carbide composite. d). Electron diffraction pattern of an icosahedral Zn-Mg-Ho quasicrystal. e). Spherulites in in poly-3-hydroxy butyrate (PHB) f). Cast iron with magnesium induced spheroidised graphite. h). Cross-section of an aluminium casting g). As-cast wrought-grade aluminium alloy (Al-Mg-Fe-Si) i). Metal component. j). Automotive engine. Structure images courtesy of Core-Materials [1]

and manufacturing experts. Additionally, PSP linkages provide systemic, modular and hierarchical protocols for community engagement (i.e., several people making complementary or overlapping contributions to the overall curation of materials knowledge).

The Materials Knowledge Systems (MKS) is a novel data science approach for solving hierarchical materials science problems using techniques from physics, machine learning, regression analysis, signal processing, and spatial statistics. The MKS framework can be used to bridge multiple length scales, and provides a data driven approach to solving inverse material design problems. Previous development of the MKS framework provides systematic protocols to create high value Structure-Property linkages for hierarchical materials that can be leveraged by experts in the manufacturing, data science, and materials science and engineering communities.

This work builds on the existing framework and extends the previous definition of MKS from Structure-Property linkages defined by physics based kernels, to include Process-Structure linkages using similar kernels. The use of these kernel based methods is referred to as MKS localization linkages. Previous work related to the same framework uses spatial correlations to created Structure-Property linkages with effective material properties. This work also extends this framework to Process-Structure linkages using advanced regression methods. These linkages based on spatial correlations are referred to as MKS homogenization linkages. In summary this work contributes to the accelerated materials development community by i). creating a code base that allows for high level access to previously developed MKS framework to create Structure-Property linkages through the Python package PyMKS, and ii). the extension of the MKS framework to Process-Structure linkages which when combined with the original framework provides a method for the creation of full PSP linkages.

CHAPTER II

EXECUTIVE SUMMARY

This chapter provides a high level overview of the contributions of this work. Section 2.1 provides the necessary background on the previously developed core principles of MKS framework used to create Structure-Property linkages. The remaining sections discuss the implementation of the Structure-Property linkage framework into the Python package Materials Knowledge Systems in Python (PyMKS), and the extension of this framework to Structure-Property linkages.

2.1 Introduction to Materials Knowledge Systems

In this section the mathematical connections between between physics and machine learning used in the MKS framework are presented. Additionally, the workflows to create homogenization linkages (connecting a lower length scales to a higher length scales) and localization linkages (connecting higher length scales to lower length scales) in hierarchical materials are presented.

2.1.1 Microstructure Function and Local State Space

Two of the foundational concepts in the MKS framework are the microstructure function and local states. Local states, h, are the local thermodynamic state variables (or order parameters) needed to uniquely define local material structure. The space of all possible local states is called the Local State Space and is represented with the variable H. The microstructure function $m(h, x)$ is the probability density associated with finding the local state h at spatial position x. The expectation value obtained using this probability density (on the local state space) is the measured value of the

local state at location x denoted by $\Phi(x)$ as shown in Eq. (1) and (2).

$$\int_H m(h,x)dh = 1 \tag{1}$$

$$\Phi(x) = \int_H hm(h,x)dh \tag{2}$$

The introduction of the microstructure function as a probability density function provides two advantages: i). it casts material structure into a probabilistic framework that effectively connects to data science, and ii). it maps complex descriptions of local states (potentially could be a combination of several scalar and tensor thermodynamic state variables) into a consolidated continuous scalar-valued function that lends itself naturally to spectral representations [16, 17, 18, 19].

2.1.2 Derivation of MKS Series using Elastostatics

Although the derivation presented here for the MKS series is demonstrated with the elastostatics equation, the same series can be derived from differential equations that govern other phenomena such as thermoelectric, piezoelectric, diffusion, and conductivity for composite materials [20]. Similar derivations with the electrostatics can be found in published literature [21].

Let $C(x)$ be the local stiffness tensor for a two-phase material with stiffness tensors C_A and C_B. The stiffness tensor and strain field at location x can be represented as a perturbation from a pre-selected reference.

$$C(x) = C^R + C'(x) \tag{3}$$

$$\varepsilon(x) = \bar{\varepsilon} + \varepsilon'(x) \tag{4}$$

where C^R and $\bar{\varepsilon}$ are the reference stiffness tensor and average strain, and $C'(x)$ and $\varepsilon'(x)$ are the local perturbations from their respective references.

With these perturbed definitions, the constitutive equation for stress and strain is therefore

$$\sigma_{ij}(x) = \big(C^R_{ijlk} + C'_{ijlk}(x)\big)\big(\bar{\varepsilon}_{lk} + \varepsilon'_{lk}(x)\big) \tag{5}$$

and the equilibrium condition is

$$\sigma_{ij,j}(x) = \Big[\big(C^R_{ijlk} + C'_{ijlk}(x)\big)\big(\bar{\varepsilon}_{lk} + \varepsilon'_{lk}(x)\big)\Big]_{,j} = 0 \tag{6}$$

$$\sigma_{ij,j}(x) = C^R_{ijlk}\varepsilon'_{lk,j}(x) + C'_{ijlk,j}(x)\bar{\varepsilon}_{lk} + \Big[C'_{ijlk}(x)\varepsilon'_{lk}(x)\Big]_{,j} = 0 \tag{7}$$

Eq. (7) can be recasted into an inhomogenous differential equation with terms that only contain the reference tensor and the perturbed strain as shown in Eq. (8) and (9).

$$F_i(x) = C'_{ijlk,j}(x)\bar{\varepsilon}_{lk} + \Big[C'_{ijlk}(x)\varepsilon'_{lk}(x)\Big]_{,j} \tag{8}$$

$$C^R_{ijlk}\varepsilon'_{lk,j}(x) + F_i(x) = 0 \tag{9}$$

The perturbed strain can be represented in terms of a perturbed displacement field $u'(x)$.

$$\varepsilon'_{kl}(x) = \frac{\big(u'_{k,l}(x) + u'_{l,k}(x)\big)}{2} \tag{10}$$

$$C^R_{ijkl}\frac{\big(u'_{k,lj}(x) + u'_{l,kj}(x)\big)}{2} + F_i(x) = 0 \tag{11}$$

The solution to the displacement field can be found for any inhomogeneous term (and therefore microstructure) using Green's functions as shown in Eq. (13). Note that the $G(r)$ terms depend on the reference medium C^R.

$$C^R_{ijkl}G_{km,lj}(r) + \delta_{im}\delta(x-r) = 0 \tag{12}$$

$$\begin{aligned}\varepsilon'_{kl}(x) &= \int_V \frac{\big(G_{ik,l}(r) + G_{il,k}(r)\big)}{2}F_i(x-r)dr \\ &= \int_V \frac{\big(G_{ik,l}(r) + G_{il,k}(r)\big)}{2}\Big[C'_{ijlk}(x-r)\bar{\varepsilon}_{lk} + \big[C'_{ijlk}(x-r)\varepsilon'_{lk}(x-r)\big]\Big]_{,j}dr\end{aligned} \tag{13}$$

Assuming periodic boundary conditions, the derivatives in the terms which define $F(x)$ can be moved to the Green's functions using integration by parts, and the

perturbed strain can then be written in the form shown in Eq. (14)

$$\varepsilon'_{kl}(x) = \int_V \Gamma_{ikjl}(r)\Big[C'_{ijlk}(x-r)\bar{\varepsilon}_{lk} + \big[C'_{ijlk}(x-r)\varepsilon'_{lk}(x-r)\big]\Big]dr \tag{14}$$

In Eq. (14), the $\Gamma_{ikjl}(r)$ term contains the Green's functions and is independent of the microstructure. Eq. (14) contains $\varepsilon'(x)$ on both sides. A series can be created by recursively inserting the right hand side into itself. The resulting series is shown in Eq. (15). For the remainder of the derivation the indices will be dropped for economy of notation.

$$\varepsilon'(x) = \int_V \Gamma(r)C'(x-r)\bar{\varepsilon}dr + \int_V\int_V \Big[\Gamma(r)C'(x-r)\bar{\varepsilon}\Big]\Big[\Gamma(r')C'(x-r')\bar{\varepsilon}\Big]dr'dr + ... \tag{15}$$

The microstructure function $m(h,x)$ and local states h can be introduced into the series found in Eq. (15) using the relationships described in Eq. (16) as reults in the series shown in Eq. (17).

$$\begin{aligned} C'(x-r) &= \int_H hm(h,x-r)dh \\ \alpha(h,r) &= \Gamma(r)h\bar{\varepsilon} \\ \alpha(h,h',r,r') &= \Gamma(r)h\bar{\varepsilon}\Gamma(r')h'\bar{\varepsilon} \end{aligned} \tag{16}$$

$$\begin{aligned} p(x) = &\int_V\int_H \alpha(h,r)m(h,x-r)drdh + \\ &\int_V\int_V\int_H\int_H \alpha_(h,h',r,r')m(h,x-r)m(h',x-r')dr'drdhdh' + ... \end{aligned} \tag{17}$$

In Eq. (17), the perturbed strain $\varepsilon'(x)$ as been replaced with $p(x)$ as a more general term that can be used to represent some local property field of interest. The discretized version of Eq. (17) is the MKS series and can be found in Eq. (18).

$$\begin{aligned} p[s] = &\sum_{l=0}^{L-1}\sum_{r=0}^{S-1} \alpha[l,r]m[l,s-r] + \\ &\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\sum_{r=0}^{S-1}\sum_{r'=0}^{S-1} \alpha[l,l',r,r']m[l,s-r]m_[l',s-r'] + ... \end{aligned} \tag{18}$$

The transition from round brackets "()" to square brackets "[]" is used to indicate the transition from continuous to discrete variables. In Eq. (18), s, S, l and L are the discrete variables for x, V, h, and H respectively.

2.2 MKS Structure-Property Linkages

MKS Structure-Property linkages are based on Eq. (18). The framework used to create linkages between materials structure and effective properties referred to as MKS Homogenization Structure-Property linkages can be derived from this equation by using constitutive relationships (e.g. relating the strain ε to the stress σ through the stiffness tensor C) an additional microstructure function can be multiplied to all of the terms in the series. By taking the volume average of the series with the additional microstructure function, it changes the first term of the series into 2-point spatial correlations and the higher order terms become higher spatial correlations. More detailed derivation of this transition can be found in [22]. The MKS Homogenization workflow uses spatial correlations to create features for machine learning models that connect effective properties and structure.

The MKS Localization workflow uses regression to calibrate microstructure independent α terms to connect local property field with structure. The high level steps for each of the workflows are shown in Fig. 2.2. The first step in both of the workflows is the discretization of the microstructure function. The remaining steps in the homogenization workflow are the computation of 2-point statistics, dimensionality reduction, and the application of machine learning models to calibrate the linkage. The second and final step of the localization workflow is the calibration of the α terms from Eq. (18). The remainder of this section will describe the workflows in more detail.

2.2.1 MKS Homogenization Structure Property Linkages

2.2.1.1 Step 1: Discretize the Microstructure Functions

$m(h, x)$ exhibits a dependence on both local state and spatial variables. The spatial variables are discretized by either sampling or averaging over a finite domain. The local state variable can also be discretized with one of two strategies. The simplest

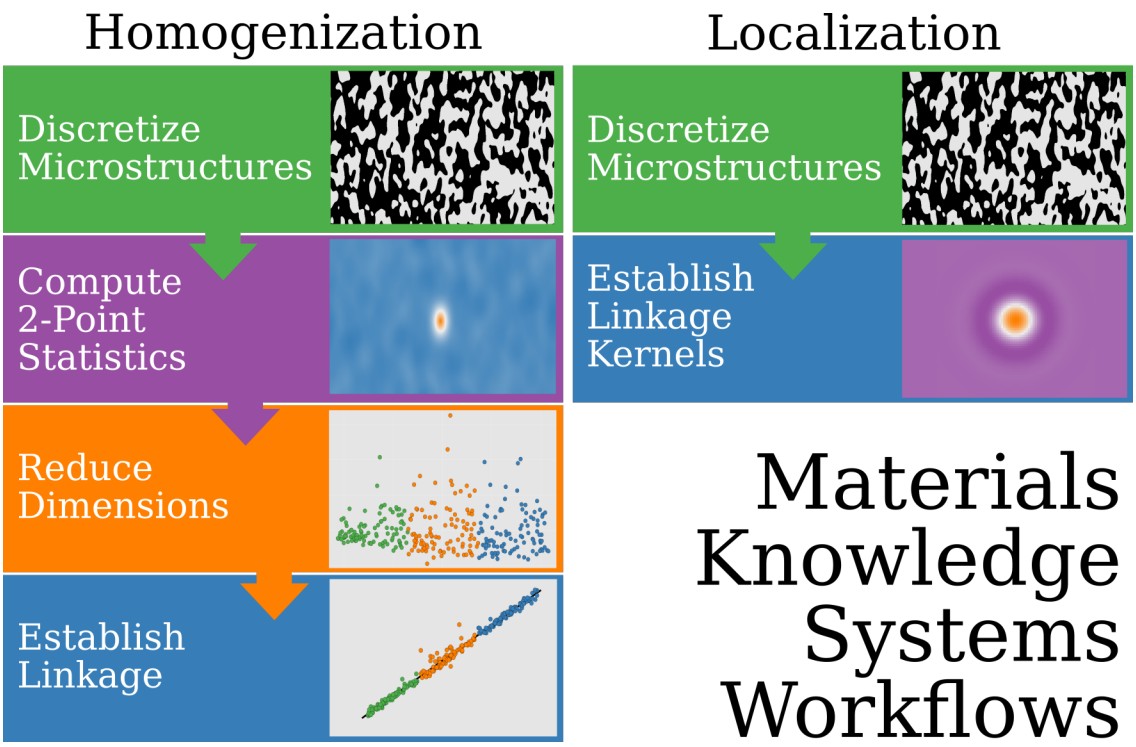


Figure 2: The MKS Homogenization workflow (left) consists of four steps. 1. Discretize the raw microstructure with the microstructure function. 2. Compute 2-point statistics to create a natural origin. 3. Create low dimensional microstructure descriptors using dimensionality reduction. 4. Calibrate structure-property linkage with low dimensional microtructure descriptors and effective properties using machine learning. The MKS Localization workflow (right) consists of 2 steps. 1. Discretize the raw microstructure with the microstructure function. 2. Calibrated physics-based kernels using regression.

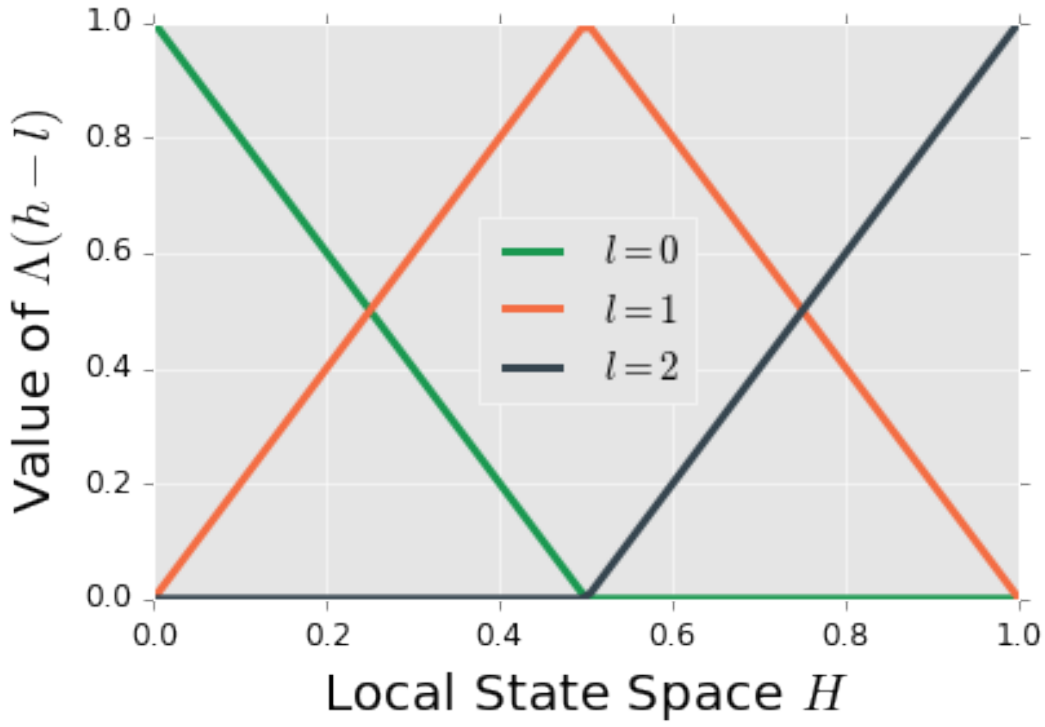


Figure 3: Three hat basis functions Λ used to discretize the local state space H. This method is referred to as Primitive basis functions.

approach is to discretize the local state space H using triangle or hat basis functions $\Lambda(h-l)$ to divide the local state space into intervals. The hat basis functions are defined in Eq. (19).

$$\Lambda(h-l) = max\left(1 - \left|\frac{h(L-1)}{H} - \frac{Hl}{L-1}\right|, 0\right) \tag{19}$$

In Eq. (19), H and h maintain their definitions as the local state and local state space variable respectively, L is the total number of hat basis functions used to span the local state space and l enumerates the hat functions. The hat functions are placed along the local state space such that the maximum and minimum values of the local state space domain fall on the peak values of the hat functions associated with the largest and smallest values of l as descript in Eq. (20). As an example, hat functions with $L = 3$ and $h \in [0, 1]$ can be found in Fig. 3. The discretized microstructure function $m[l, s]$ using the hat functions as basis function is shown in Eq. (20).

$$m[l,s] = \int_H \int_V \delta(x-s)\Lambda(h-l)m(h,x)dxdh \tag{20}$$

In Eq. (20), $\delta(x)$ is a dirac delta function. In previous work, this approach has been referred to as primitive binning using the Primitive basis functions [23, 24, 25, 26, 27, 3].

As an alternative, functions on the local state space $f(h)$ can be represented using orthogonal basis functions ξ as shown in Eq. (21) and (22).

$$f(h) = \sum_{l=-\infty}^{\infty} c_l \xi_l(h) \tag{21}$$

$$\frac{1}{N_l} \int_a^b \xi_{l'}(h)\xi_l(h)w(h)dh = \delta_{ll'} \tag{22}$$

In Eq. (22) ξ_l is the l^{th} order orthogonal basis function, $w(h)$ is the weighting function, $\delta_{ll'}$ is the Kronecker delta and N_l is a normalization constant that depends on the order and type of the basis functions. The set of coefficients c_l defined in Eq. (21) provide a unique discrete representation of the function $f(h)$. This approach is useful when only a small number of c_l dominate the representation and the weighting function $w(h)$ equals one.

Applying this discretization approach to capture the h dependence in $m(h,x)$ using an orthogonal basis functions with $w(h) = 1$ and mapping the local state space to the orthogonal domain leads to the discretization found in Eq. (23).

$$\int_V \delta(x-s)m(h,x)dx = \sum_{l=0}^{L-1} m[l,s]\xi_l(h) \tag{23}$$

In both discretization methods found in Eq. (20) and (23), $m[l,s]$ are associated with a bases functions that represent the local state dependence of $m(h,x)$.

2.2.1.2 Step 2: Compute 2-Point Statistics

n-point spatial correlations provide a way statistical framework to rigorously quantify material structure. 1-point statistics $f_j[l]$ are the probability that a specified local

state will be found in any randomly selected spatial bin in a microstructure j as shown in Eq. (24).

$$f_j[l] = \frac{1}{S} \sum_s m_j[l, s] \tag{24}$$

While 1-point statistics provide information on the relative amounts of the different local states, information about how those local states are spatially arranged in the microstructure is not provided. 2-point spatial correlations $f_j[l, l', r]$ (also known as a 2-point statistics) [28, 29, 30] capture the probability of finding the head of a randomly placed vector r in local state l, while the tail of the vector is in local state l' in the microstructure j, and is defined in Eq. (25).

$$f_j[l, l', r] = \frac{1}{\Omega_j[r]} \sum_s m_j[l, s] m_j[l', s + r] \tag{25}$$

In Eq. (25), $\Omega_j[r]$ is a normalization factor that depends on the vector r. The application of this concept to a two-phase composite material with example discretized vectors can be found in Fig. 4. When the 2 local states are the same $l = l'$, the set of 2-point statistics are called an autocorrelation, otherwise they are called a cross-correlation.

Higher order spatial statistics are similar to 2-point statistics, in that they can be thought of in terms of probabilities of finding specified local states separated by a prescribed set of vectors. 3-point statistics are the probability of finding three specified local states at the ends of a triangle (defined by 2 vectors), placed randomly in the material structure. 4-point statistics describes the probability of finding 4 local states at 4 locations (defined, using 3 vectors) and so on.

While higher order statistics provide more information about the material structure, the general time complexity to compute n-point statistics is $O(N^n)$ making the computation of higher order statistics prohibitive. Additionally, using the properties of Fourier transforms [31], the time complexity for 2-point statistics can be reduced to $O(N \log N)$. For this reason, the general protocol for MKS Homogenization only

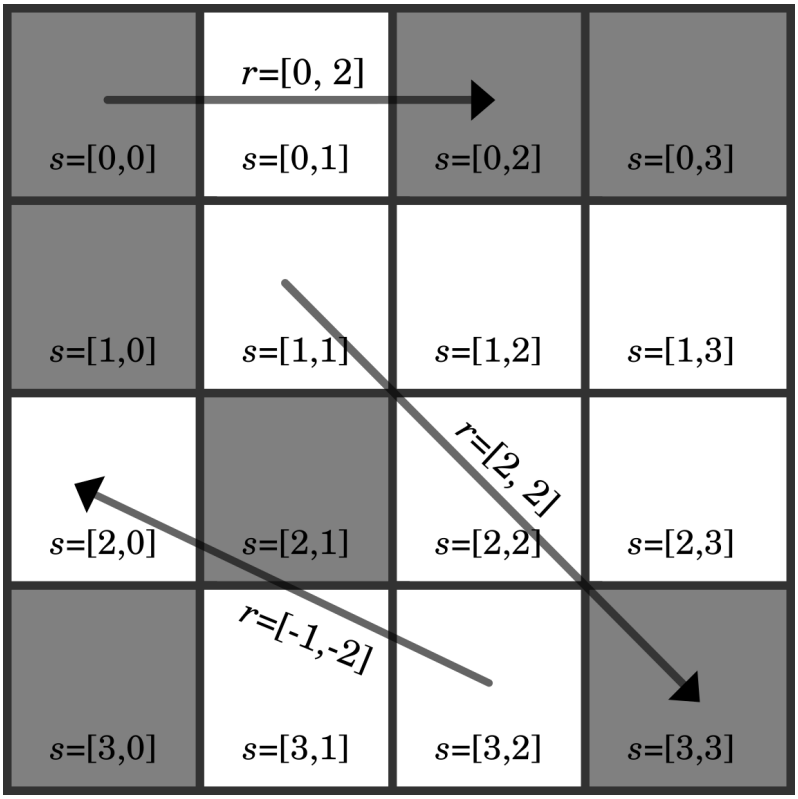


Figure 4: The discretization scheme for both the microstructure function and the vector space needed to define the spatial correlations, illustrated on a simple two-phase composite material. The discretized vectors r describe the relative positions between different spatial locations.

uses 2-point statistics.

2.2.1.3 Step 3: Dimensionality Reduction with PCA

The total number of combinations of 2-point statistics defined by the two local states creates an extremely large feature space with collinear features. Dimensionality reduction can be used to reduce the feature space to a reasonable size. Eq. (18), shows that the 2-point statistics are linearly related to effective property, therefore a linear dimensionality reduction technique is advantageous. Principal Component Analysis (PCA) is a global linear dimensionality reduction technique that maps high dimensional data into a low dimensional space through Principal Components (PCs) while retaining as much of the variance in the data as possible [32]. The application of PCA to 2-point statistics is shown in Eq. (26).

$$f_j\left[\nu\right] \approx \sum_{k \in K} \mu_j\left[k\right] \phi\left[k, \nu\right] + \overline{f[\nu]} \tag{26}$$

In Eq. (26), $f_j\left[\nu\right]$ is a set of 2-point statistics. The variable ν uniquely represents a combination of l, l', and r in Eq. (25). $\phi[k, \nu]$ are PCs and can be thought of as calibrated bases that can be used to represent $f_j\left[\nu\right]$. $\mu_j\left[k\right]$ are the Principal Component scores (PC scores) and are unique low dimensional representation of $f_j[\nu]$. $\overline{f[\nu]}$ are the average values over all ν used to calibrate $\phi[k, \nu]$.

2.2.1.4 Step 4: Calibration of Linkage

The low dimensional microstructure descriptors $\mu_j\left[k\right]$ and effective property, p_j^{eff} serve as inputs and outputs for a machine learning model $\mathcal{F}$ as show in Eq. (27).

$$p_j^{\text{eff}} = \mathcal{F}(\mu_j[k]) \tag{27}$$

Previous studies have used polynomial regression as $\mathcal{F}$ [22, 2, 33].

2.2.2 Localization

The homogenization and localization workflows share the same first step which is described in Sec. 2.2.1.1.

2.2.2.1 Step 2: Calibration of Linkage

The second and final step of the MKS localization workflow is to calibrate the Green's function based α terms in Eq. (28) using regression.

$$p_j[s] = \sum_{l,r} \alpha[l,r] m_j[l, s-r] + \sum_{l,l',r,r'} \alpha[l,l',r,r'] m_j[l, s-r] m_j[l', s-r'] + ... \quad (28)$$

Prior work [23, 24, 25, 27, 3] has shown that the first term is dominant for problems with low to moderate contrast. The calibration of the first-order MKS localization kernels can be done efficiently by taking advantage of discrete Fourier transforms and the convolution theorem [34, 35]. This transformation leads to

$$P[k] = \sum_{l} \beta[l,k] M[l,k] \quad (29)$$

In Eq. (29) $P[k]$, $\beta[l,k]$ and $M[l,k]$ are the discrete Fourier transforms of $p[s]$, $\alpha[l,s]$ and $m[l,s]$ from the first term in Eq. (28) respectively.

With the uncoupled spatial frequency representation shown in Eq. (29), the β terms can be calibrated easily using multiple linear regression techniques using the known values for $P[k]$ and $M[l,k]$. The discretization method found in Eq. (20) is subject to the constraint that the discretized microstructure function sums to one at any instance in space as shown in Eq. (30).

$$\sum_{l} m[l,s] = 1 \quad (30)$$

Due to this constraint, multiple linear regression with categorical variables as outlined in previous studies is used [23, 24, 25, 27]. The discretization method found using Eq. (23) is not subject to the constraint shown in Eq. (30) and therefore standard multiple linear regression is used.

In prior work, information from higher order kernels has been calibrated by re-defining the local state space to included information about nearest neighbors and using the same regression approach for a first term [36].

2.3 Contribution 1: Materials Knowledge Systems in Python

This section presents an overview of the Python project PyMKS and its implementation of high level APIs to create Structure-Property linkages.

2.3.1 Background

In the materials development community, common practices for developing tools and infrastructure are highly localized within organizations resulting in major inefficiency (unnecessary duplication of codes, inadequate verification and validation of multiple instantiations of code, not engaging the right talent for the right task, etc.). Although it is well know that the pace of discovery, development and deployment of knowledge and technology significantly increases with effective collaboration [37, 38, 39, 40], historically scaling effective collaborations to large communities has been difficult.

The modern advent of information technology has facilitated massive electronic collaborations (e-collaborations) that have lead to significant advances in several domains including the discover of the Higg's boson [41], the sequencing of the human genome [42], the Polymath project [43], the monitoring of species migration [44, 45] and numerous open source software projects. E-collaborations allow experts from complementary domains to create close collaborations regardless of spatial and temporal distances. E-collaborations require cyber-infrastructure that allow members to generated, analyzed, disseminated, and consumed information [46]. Open source cyber-infrastructure tools eliminate collaboration hurdles due to software licenses, and can help foster truly massive e-collaborations. In other words, even with collaborations involving proprietary data, open source cyber-infrastructure provide a common

language that can facilitate e-collaborations among large groups of practitioners.

The Materials Knowledge Systems in Python project (PyMKS) is the first open source materials data science framework that can be used to create high value Structure-Property linkages for hierarchical materials that can be leveraged by experts in materials science and engineering, manufacturing, and data science communities, and is an essential component in the cyber-infrastructure needed to realize modern accelerated materials innovation ecosystems.

The programming language Python was chosen because of it's portability, human readable syntax, holistic design, optimized libraries and community. PyMKS is licensed under the permissive MIT license which allows for unrestricted distribution in commercial and non-commercial systems.

2.3.2 Underlying Technologies

PyMKS is built on highly optimized Python packages NumPy [47], SciPy [48], and Scikit-learn [49]. NumPy arrays are the primary data structure used throughout PyMKS and provide basic algorithmic functionality. SciPy's signal processing and numerical linear algebra functions are used to calibrate models and generate synthetic data. PyMKS is highly integrated with Scikit-learn and mimics its simple API in order to leverage models for dimensionality reduction, regression, classification and model selection. In addition PyMKS uses the Python testing framework Pytest for unit-tests and doc-tests.

2.3.3 Code and API Design

PyMKS contains 5 major modules that include microstructure function basis classes, models to objectively quantify structure and create PSP linkages, functions to compute 2-point statistics, data generation tools and plotting functions. The basis classes use methods to discretize the raw structure data and introduce the local state variable l. The four basis classes are designed efficiently discretize different types of local state

variables. The Primitive Basis uses indicator (or hat) functions and it well suited for microstructures that have discrete local states. The Legendre Polynomial Basis and Fourier Series Basis create spectral representation of non-periodic and periodic continuous local state distributions respectively. Lastly Generalized Spherical Harmonics basis create compact spectral representations of orientation distribution functions in the local state space [50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61].

PyMKS has three models which implement the approaches to quantify differences in materials structure, and create homogenization and localization Structure-Property linkages. The Structure Analysis model creates objective low dimensional structure descriptors using 2-point spatial statistics and dimensionality reduction to quantify materials structure. The default dimensionality reduction technique is PCA, but any model from Scikit-learn that has a `transform_fit` method can be used. The Homogenization inherits the Structure Analysis model, but adds the additional step of creating a linkage between the low dimensional microstructure descriptors and a effective property to class label using an estimator form Scikit-learn. The default estimator is a polynomial regression, but any estimator from Scikit-learn that has `fit` and `predict` methods can be used to either predict material properties or classify the microstructures. The Localization model requires a set of microstructures and their local field responses to calibration influence coefficients that can be used to predict local materials responses as well as structure evolutions that are independent of the microstructure. The Localization model has `fit` and `predict` methods to follow the standard API for an estimator from Scikit-learn.

PyMKS also contains modest data generation tools that are used in both examples and unit tests. A Microstructure Generator class uses filters of different shapes to create synthetic microstructures. A Cahn-Hilliard simulation with an optimized semi-implicit spectral scheme with periodic boundary conditions [62] is used to generate data with a continuous local state. A finite element simulation with periodic boundary

conditions [25] is used to create both local and effective material properties with a microstructure containing only discrete local states. Lastly PyMKS contains some plotting tools that are used throughout the examples. A more detailed overview of the PyMKS project can be found in Ch. 3 and on pymks.org.

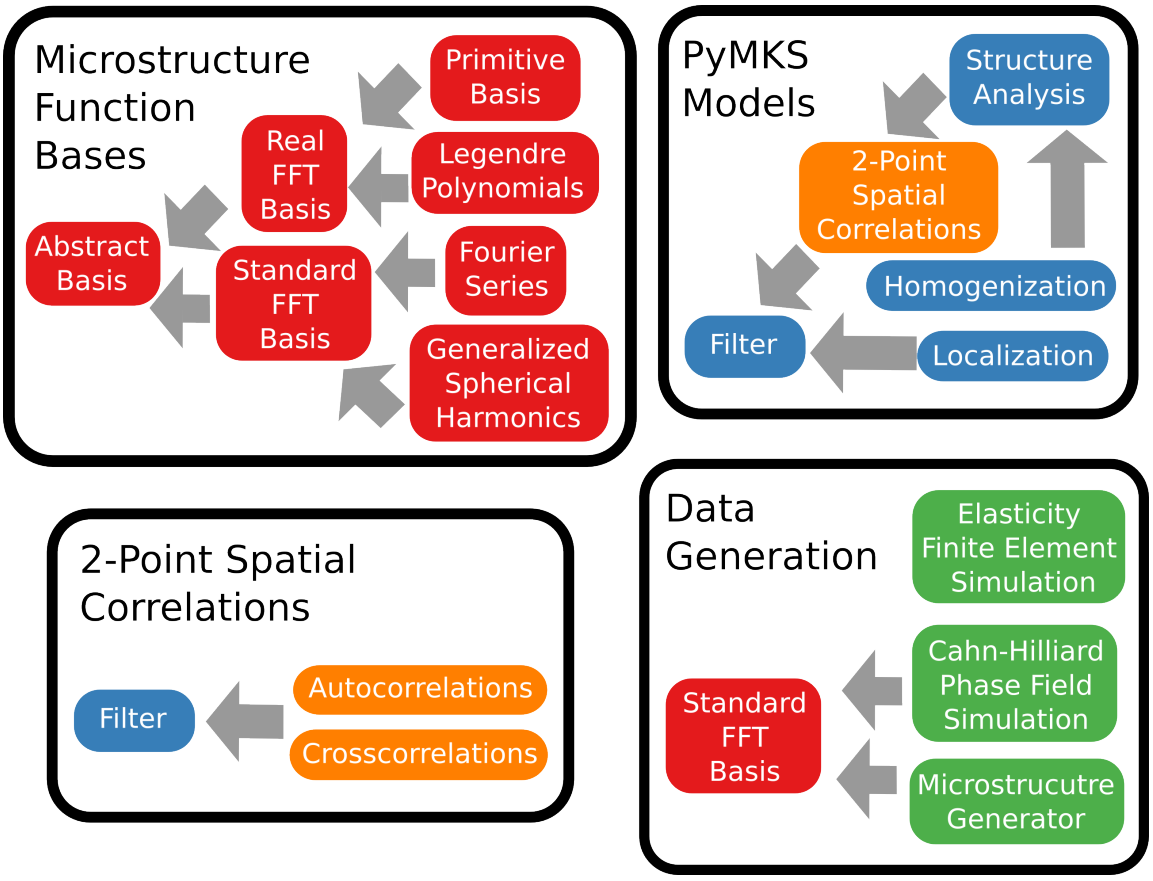


Figure 5: An illustration of the design for PyMKS. The four blocks represent the main functionalists in PyMKS. The modules listed on the far left are exposed in the API. The arrows indicate imported modules.

2.4 Contribution 2: Process-Structure Homogenization Linkages

This section provides a high level description the contribution to development of the process-structure homogenization linkage for the MKS framework with an example using synchrotron X-ray scattering data of polyethylene.

2.4.1 Structure Evolution and Time Series Analysis

Process-structure homogenization linkages aim to predict the evolution of low dimensional structure descriptors. To do so, the introduction of time into the microstructure function is needed $m[l, s, n]$, and results in changes to Eq. (25) and (26) to their time dependent versions found in Eq. (31) and (32) respectively.

$$f_j[l, l', r, n] = \frac{1}{\Omega_j[r, n]} \sum_{s \in S} m_j[l, s, n] m_j[l', s + r, n] \tag{31}$$

$$f_j[\nu, n] \approx \sum_{k=1}^{K} \mu_j[k, n] \phi[k, \nu] + \overline{f[\nu]} \tag{32}$$

In Eq. (31), $f_j[l, l', r, n]$ is the 2-point statistic for the local states l and l', the vector r and time n. $\Omega_j[r, n]$ is a normalization factor that depends on the sample j and the vector r and time n. In Eq. (32) $f_j[\nu, n]$ is a vectorized format of the 2-point statistics and ν represents a unique combination of l, l' and r at time n. K is the total number of principal components (PCs). $\mu_j[k, n]$ is a time dependent low dimensional microstructure descriptors for PC k at time n. A model is desired to predict future microtructure descriptors $\mu[k, n + i]$ using previous values and time series analysis.

A non-parametric regression method called Time Series Multivariate Adaptive Regression Splines (TSMARS) has been shown to work well with a moderate number of dimensions and medium size data [63]. A major concept in the TSMARS model are hinge functions. Hinge functions equal the value of its argument when the argument is positive and zero otherwise and shown in Eq. (33). Two hinge functions can create disjoint subregions in a high dimensional space using a data point referred to as a know κ in their argument as shown in Fig. 6.

$$g(x) = \begin{cases} x, & \text{if } x > 0. \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

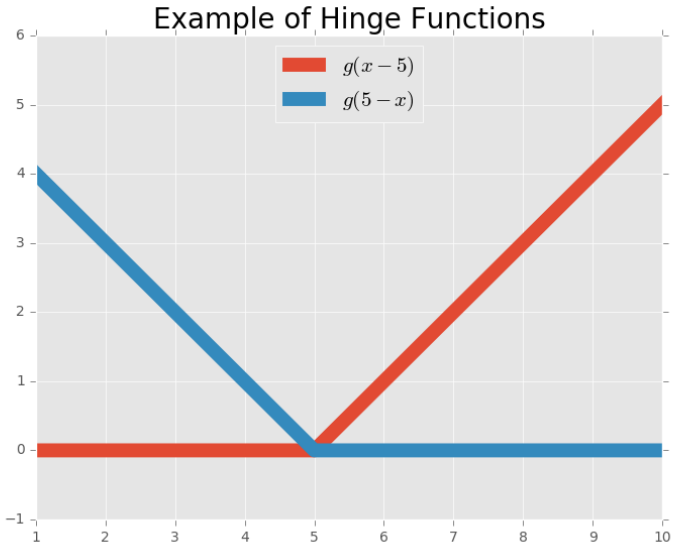


Figure 6: An example of two hinge functions meeting at a knot with value of five ($\kappa = 5$).

The calibration of the TSMARS generates non-linear functions in disjoint subregions of PC space for previous values of microstructure descriptors using products and summations of hinge functions as shown in Eq. (34).

$$\hat{\mu}[k,n] = a[0] + \sum_{i=1}^{P} a[n-i] \prod_{q=1}^{Q_i} b[q,i] g(\mu[k', n-i] - \kappa[q,i])] + \varepsilon \tag{34}$$

In Eq. (34), $\hat{\mu}[k,n]$ is the estimated microstructure descriptor for PC k at time n. $a[0]$ and $a[n-i]$ are coefficients that are calibrated using ordinary least squares regression. P is the number of previous time steps used in the model. $\mu[k', n-i]$ is a microstructure descriptor for PC k' and time $n-i$. g is the hinge function described in Eq (33). $\kappa[q,i]$ is the knot for subregion q and time $n-i$. Lastly $b[q,i]$ is a constant that can only take the values of ± 1.

2.4.2 Process-Structure Homogenization Linkage with X-ray Scattering

In this study X-ray scattering images are used in place of the 2-point statistics in Eq. (32). The two are indeed related and, it has been shown that the X-ray scattering data corresponds to the Fourier transform of the autocorrelation of the difference

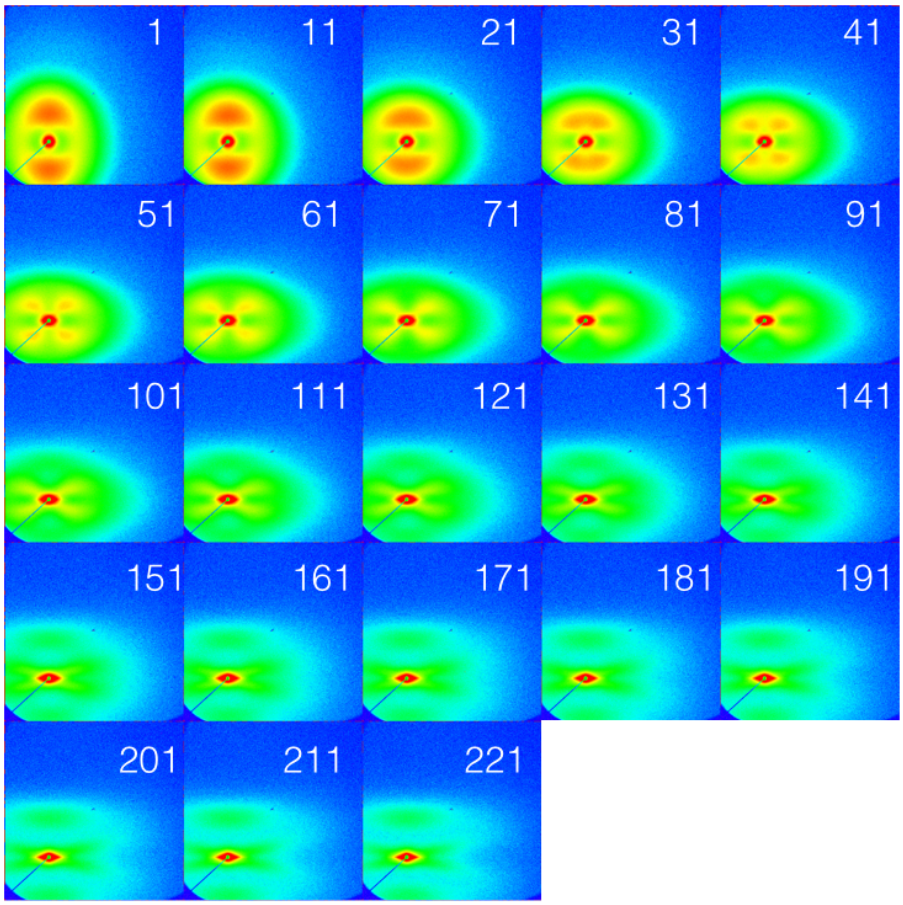


Figure 7: An example of the typical evolution in 2D SAXS patterns with increasing strain for a tensile specimen of LLDPE2.2a. Strain is applied in the vertical direction. The numbers indicate the SAXS pattern number; an interval of 3 seconds between consecutive patterns is strictly maintained. Every tenth image is displayed for clarity. The intensity is log scaled to highlight the characteristic features in the SAXS evolution with time

in electron density [64]. In this example 12 different samples of polyethylene with systematically different processing conditions using three parameters (density, blow up ratio (BUR) and thickness) were used. Each sample had a uniaxially strain applied while a synchrotron X-ray beam measured the scattering pattern as a function of time. Fig 7 shows the evolution of once of the X-ray scattering patterns.

Principal component analysis was used on all 2224 images from all 12 samples. The process steam lines found in image on top of Fig. 8 illustrate the evolution of

Polymer	Film Label	BUR	Thickness (μm)	Images
6*LLDPE1	LLDPE1.1a	2.5	20	182
	LLDPE1.1b	2.5	30	191
	LLDPE1.1c	2.5	75	195
	LLDPE1.2a	3	20	191
	LLDPE1.2b	3	30	195
	LLDPE1.2c	3	75	191
6*LLDPE2	LLDPE2.1a	2.5	20	191
	LLDPE2.1b	2.5	30	195
	LLDPE2.1c	2.5	75	61
	LLDPE2.2a	3	20	227
	LLDPE2.2b	3	30	199
	LLDPE2.2c	3	75	206

Table 1: Labeling of tensile specimens made from blown films of the two LLDPE polymers.

the structure due to the strain and the percent variance as a function of PC is shown in lower image of Fig. 8

A leave one sample out cross-validation method was used to optimize the number of principal components K and autoregressive order P, and was found to be $K = 20$ and $P = 1$. The multistep prediction process starts with proving the initial P microstructure descriptor values and recursively uses predictions for previous time steps to march forward in time. The most accurate model (based on predicted mean sqaure error values) was used to predicted microstructure descriptors up to the final time step. These final predictions in turn were used to construct a low rank approximation of the final X-ray scattering image. The results for this process can be found in Fig. 9. In Fig. 9, the top images show the predicted and actual low dimensional microstructure descriptors for all time steps for the first eight PCs. The lower two images show the final image from the experiment and the final prediction images from the model.

This example demonstrates the utility of this approach. With this models only the microstructure descriptors for the initial X-ray scattering image are needed to predict the final X-ray scattering image and still maintain the important scattering

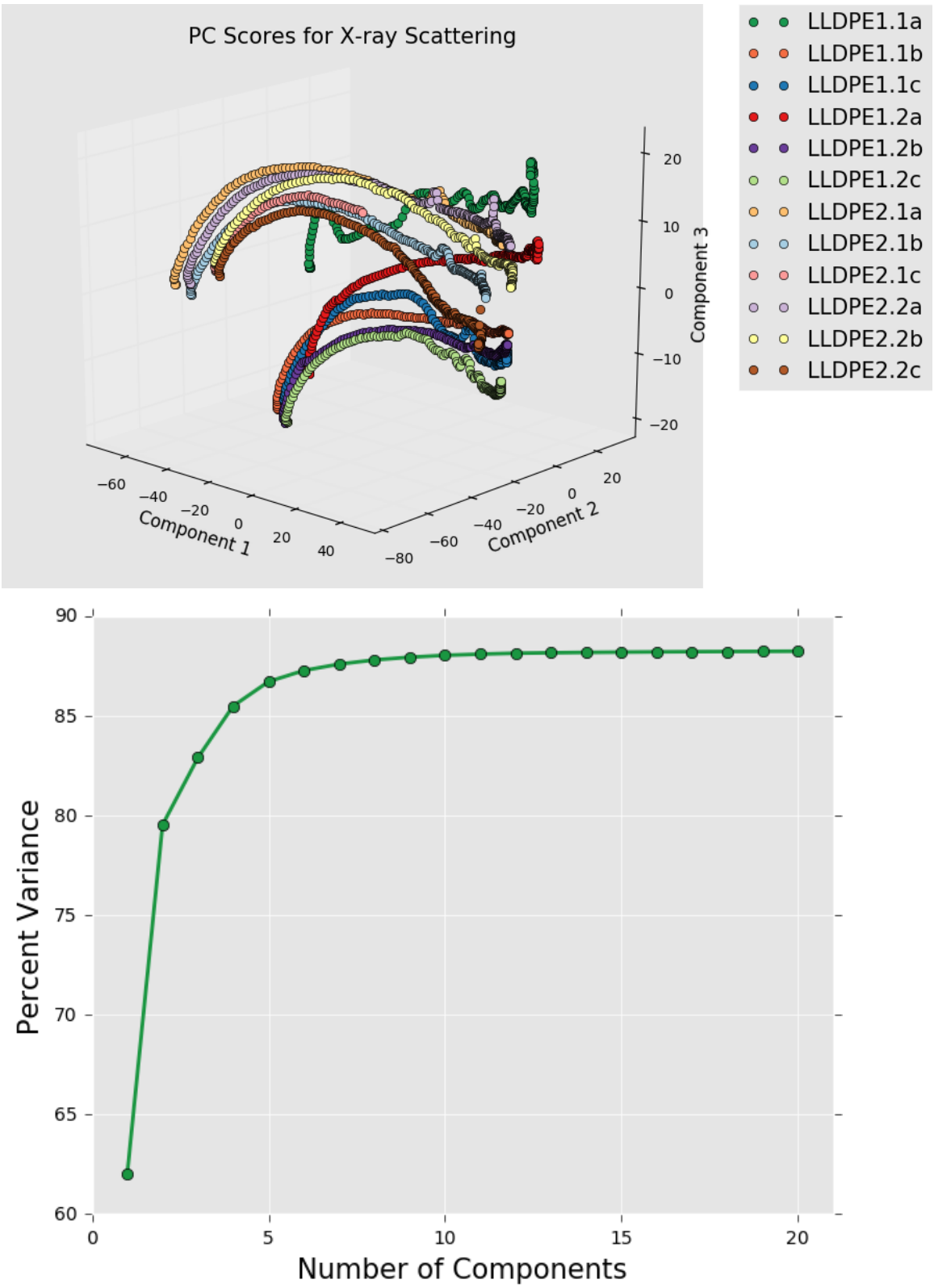


Figure 8: Principal component scores for X-ray scattering images from the 12 different samples and there percentage of the variance captured as a function of the number principal components.

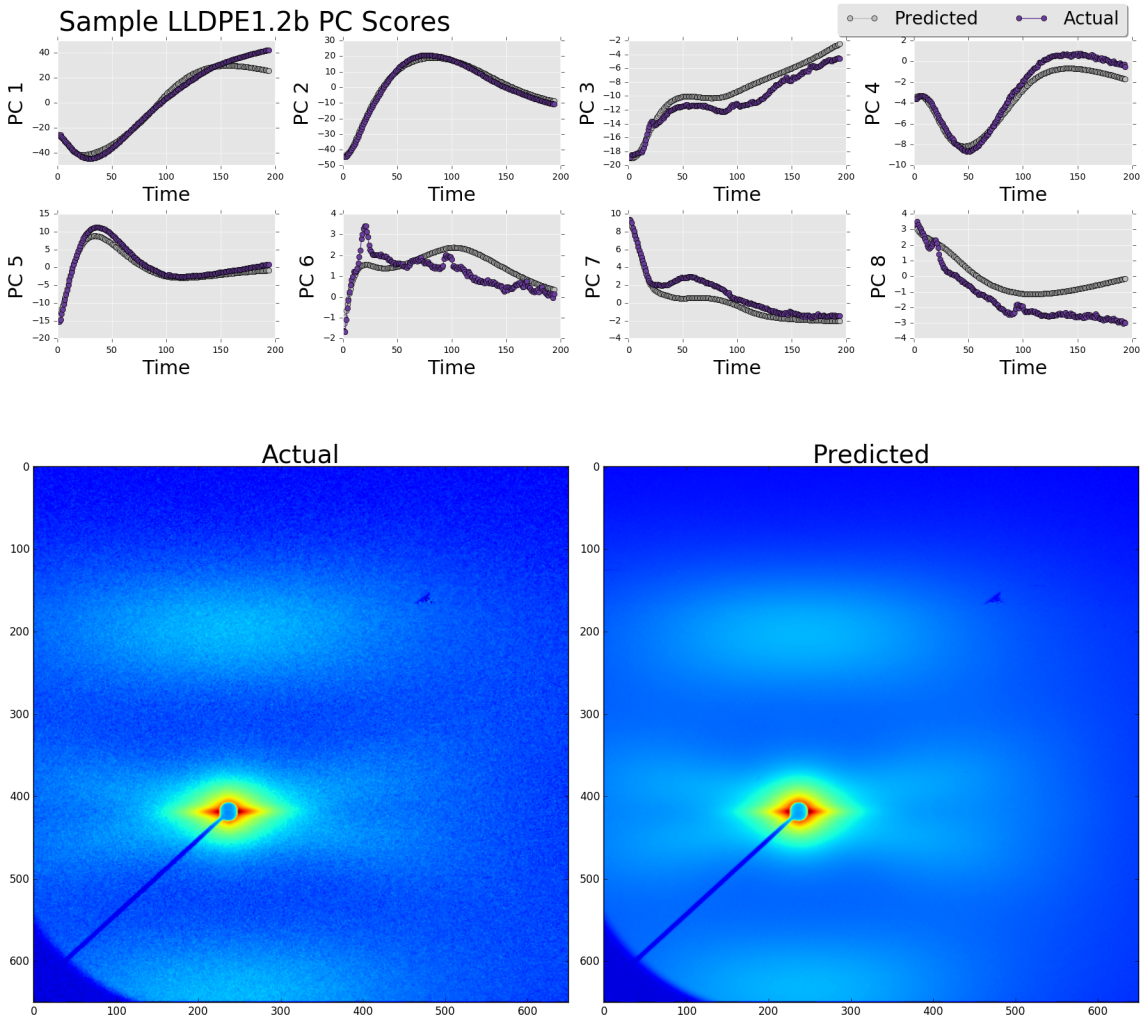


Figure 9: Predicted and actual principal component scores for sample LLDPE1.2b (above). The original image (bottom left) and the predicted image (bottom right). The mean squared error value over the predicted principal component scores had a value 2.11.

features. Althought not displayed in this study, the addtion of a Structure-Property linkage could be added to this study to create a full PSP linkage. A more detailed review of this methodology and case study can be found in Ch. 4.

2.5 Contribution 3: Process-Structure Localization Linkages

This section discusses the contribution of Process-Structure linkage to the MKS localization framework and demonstrates the methods viability using the Cahn-Hilliard equation as a case study.

2.5.1 Derivation of the MKS Series for Process-Structure Localization Linkages

The derivation of the MKS Localization Process-Structure Linkages can be viewed as introducing time dependence into the Structure-Property Linkage derivation in Sec. 2.1.2, but the starting point differs in two ways. i). The field of interest is the change in structure rather than a propertyfield, and ii). rather than using constitutive relations to introduce the microstructure terms, we begin with a non-linear differential equation that governs the evolution of the microstructure. In this example the Cahn-Hilliard equation is used [65]. The equation models the solidification process called spinodal decomposition where a single liquid phase separates into two solid phases and can be found in Eq. (35).

$$\frac{\partial \phi(x,t)}{\partial t} = D\nabla^2\Big(\big[\phi(x,t)(1-\phi(x,t)^2)\big] + \gamma\nabla^2\phi(x,t)\Big) \tag{35}$$

In Eq. (35), $\phi(x,t)$ is an order parameter used to represent the concentration field at location x and time t, $\sqrt{\gamma}$ represents the interface width between the two phases, and D is the diffusivity. The equation uses a double well potential with minima at -1 and 1 [66, 67]. We start by representing the concentration field $\phi(x,t)$ in terms of a reference quantity $\bar{\phi}$ and a local perturbation $\phi'(x,t)$ and introduce these terms into

Eq. (35) as shown in Eq. (36).

$$\begin{aligned}\frac{\partial \phi'}{\partial t} - D\nabla^2\phi' &= D\nabla^2\Big(\gamma\nabla^2\phi' - \left[3\phi'\bar{\phi}^2 + 3\phi'^2\bar{\phi} - \phi'^3\right]\Big) \\ &= D\nabla^2\Big(\gamma\nabla^2\phi' + \psi(x,t)\Big)\end{aligned} \tag{36}$$

In this new form, the Green's function approach can be used to find the solution to the equation on the left hand side as shown in Eq. (37).

$$\begin{aligned}\phi'(x,t) = &-\int_V G(r,t)\phi'(x-r,0)dr + \\ &\int_T\int_V G(r,\tau)D\nabla_r^2\Big(\gamma\nabla_r^2\phi'(x-r,t-\tau) + \psi(x-r,t-\tau)\Big)drd\tau\end{aligned} \tag{37}$$

In Eq. (37), $\phi'(x,0)$ is the initial value of the perturbed concentration and $G(r,t)$ is the Green's function. By assuming periodic boundary conditions, the operators ∇_r can be moved from concentration terms to the Green's functions. This allows $\phi'(x,t)$ to be recursively substituted forming the series expansion shown in Eq. (38).

$$\phi'(x,t) = -\int_V \Big[1 - \int_T\int_V \tilde{G}(r,r',t,\tau)dr'd\tau\Big]G(r,t)\phi'(x-r,0)dr + ... \tag{38}$$

with

$$\tilde{G}(r,r',t,\tau) = D\nabla_r^2\Big[\nabla_r^2 G(r',\tau)\gamma - 3\bar{\phi}^2 G(r',\tau)\Big] \tag{39}$$

Introducing the time-dependent microstructure function $m(h,x,t)$ and influence kernels $\alpha(h,r,t)$ as shown in Eq. (40) into Eq. (38) leads to the continuous version of the continuous time dependent MKS series as shown in Eq. (41).

$$\begin{aligned}\phi'(x,t) &= \int_H hm(h,x,t)dh \\ \alpha(h,r,t) &= -h\Big[1 - \int_T\int_V \tilde{G}(r,r',t,\tau)dr'd\tau\Big]G(r,t)\end{aligned} \tag{40}$$

$$\phi'(x,t) = \int_V\int_H \alpha(h,r,t)m(h,x-r,0)dhdr + ... \tag{41}$$

Eq. (41) is the key develop to the MKS Process-Structure localization relationship by demonstrating that the initial microstructure function $m(h,x,0)$ can be linked to

the evolution of the structure at an arbitrary time t through the time dependant, but microstructure independent influence kernels $\alpha(h, r, t)$. This results shows that with calibrated $\alpha(h, r, t)$, the evolution of new initial structure can be mapped to it's evolution at time t. Additionally, Eq. (41) is the exact analog of the structure-property localization linkages established previously in the MKS framework [23, 24, 25, 27, 3]. Therefore, the extension presented here now makes it possible to explore the complete set of Process-Structure-Property linkages in a consistent MKS framework in both space and time.

Following the discretization method outlined in Sec. 2.2.1.1, the discrete version of the MKS process-structure-localization series is shown in Eq. (42).

$$\psi[s, n] = \sum_{r=0}^{S-1} \sum_{l=0}^{L-1} \alpha[l, r, n] m[l, s - r, 0] + ... \tag{42}$$

In Eq. (42), $\psi[s, n]$ is the evolved microstructure up to time n. All other variables are the time dependent equivalents to those found in Eq. (18). The next section demonstrates the utilization of this approach using data from Cahn-Hilliard simulation.

2.5.2 Process-Structure Localization Linkage with Cahn-Hilliard Simulations

In order to calibrate the influence kernel $\alpha(h, x, t)$, 250 random initial microstructures of size (100, 100) were generated using normal distributions with standard deviations of 10^{-2}. The means of the distributions were randomly samples from an uniform distribution between [-0.5, 0.5]. The evolution of each of the samples were found by numerically solving Eq. (35) for 500 time steps. The initial microstructures and their evolved compliments served as the input and output to calibrate $\alpha(h, x, t)$ respectively. As example of one pair can be found in Eq. (10). The influence kernels were calibrated using regression techniques developed previously [24, 25, 23, 27], and images of the significant kernels can be found in Fig. 11.

In order to validate the process-structure localization linkage, an additional 250

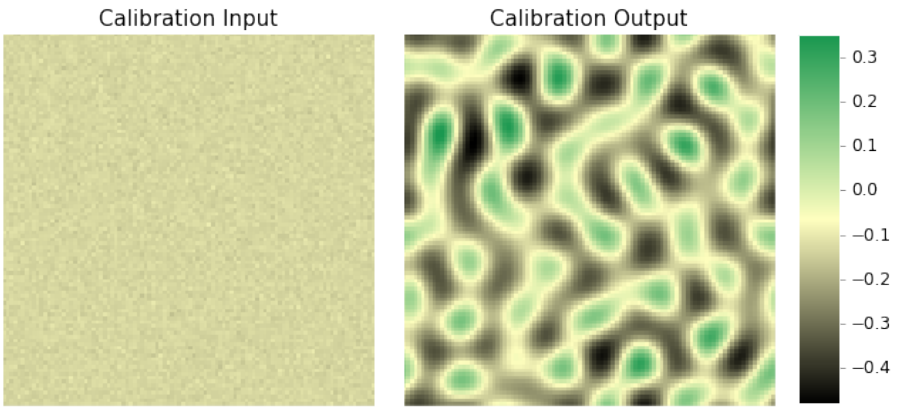


Figure 10: One instance of an initial microstructure (100 x 100) and its corresponding microstructure after 500 small time steps, which were used to calibrate the influence kernels for the MKS process-structure localization linkage.

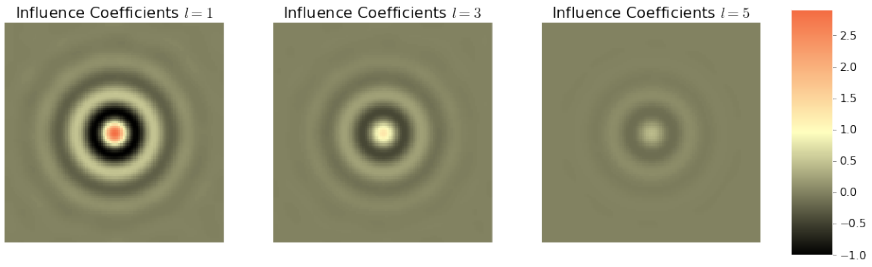


Figure 11: Significant influence kernels for the Cahn-Hillard simulation at 500 time steps.

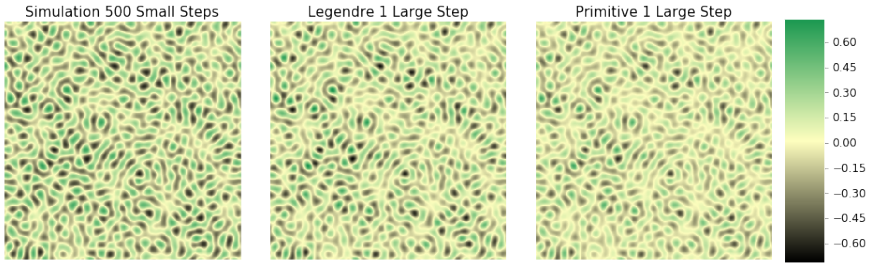


Figure 12: Predicted concentration fields by simulations using 500 time steps (left) as well as the concentration fields predicted by the two MKS localization linkages using one large time step.

micrstructures were generated with the inital structure from a noraml distrbution with the same variance as the calbration data, but the mean values were sampled from a uniform distribution between [-1, 1]. The evolution of these structures were predicted using the calibrated influence coefficients and the evolutions were also computed by solving the Cahn-Hilliard model for 500 time steps as ground truth. The average run time for the simulation to reach 500 time steps was 1.33 secs, while the average run time using the Process-Structure localization linkage was 4.94 x 10^{-3} an a root mean squared error of 3.36 x 10^{-2}. An examples of the output from the simulation and process-structure linkages with two different basis function can be found in Fig. 12.

This example demonstrates the viability of the MKS Process-Structure localization linkage approach to solve non-linear differential equation for a range of initial conditions. This approach may be extremely useful when numerical integration schemes are difficult to optimize. A more detailed overview of the approach and case study can be found in Ch. 5.

2.6 Conclusion

The major contributions of the work are:

1. The development of an open source project Materials Knowledge Systems in Python (PyMKS):

- PyMKS allows high level access the MKS Structure-Property localization and homogenization linkages. PyMKS is an open source project with a permissive license that provides simple high level APIs to access the MKS framework and has been launched with the aim to nucleate and grow an emergent community focused on establishing data-driven homogenization and localization Process-Structure-Property linkages for hierarchical materials.

2. The extension of the MKS framework to Process-Structure homogenization linkages.

    - An extension of the MKS framework to Process-Structure homogenization linkages using Time Series Multivariate Adaptive Regression Splines was presented. The Process-Structure linkage is generated using a data driven approach to calibrate non-linear relationships between low dimensional microstructure in PCA space. The approach was validated by predicting a low rank representation of the final step of Small Angle X-ray Diffraction data. This extensions allows for the complete Process-Structure-Property homogenization linkages to be created to solve inverse materials design problems and expedite the development of new materials.

3. Extension of the MKS framework to Process-Structure localization linkages.

    - An extension to the MKS framework to Process-Structure localization linkages was presented. The extension presents an alternate method to learn the underlying embedded physics in a numerical model. These Green's function based influence kernels may provide certain computational advantages in the case for problems where traditional numerical integration schemes have been difficult to optimize.

Overall, it was demonstrated that the MKS localization kernels extracted were insensitive to the details of the initial microstructure and the same kernel can be applied to any initial microstructure in the selected material system.

The methods described here have laid a strong foundation for future developments addressing a broad range of materials systems with richer microstructures and more complex governing physics. These approaches now make it possible to create complete Process-Structure-Property linkages for both homogenization and localization workflows which can be used to significantly expedite the develop of new materials.

CHAPTER III

MATERIALS KNOWLEDGE SYSTEMS IN PYTHON - A DATA SCIENCE FRAMEWORK FOR ACCELERATED DEVELOPMENT OF HIERARCHICAL MATERIALS

3.1 Introduction

Current practices for developing tools and infrastructure used in multiscale materials design, development, and deployment are generally highly localized (sometimes even within a single organization) resulting in major inefficiencies (duplication of effort, lack of code review, not engaging the right talent for the right task, etc.). Although it is well known that the pace of discovery and innovation significantly increases with effective collaboration [37, 38, 39, 40], scaling such efforts to large heterogeneous communities such as those engaged in materials innovation has been very difficult.

The advent of information technology has facilitated massive electronic collaborations (generally referred to as e-collaborations) that have lead to significant advances in several domains including the discovery of the Higg's boson [41], the sequencing of the human genome [42], the Polymath project [43], the monitoring of species migration [44, 45] and numerous open source software projects. E-collaborations allow experts from complementary domains to create highly productive collaborations that transcend geographical, temporal, cultural, and organizational distances. E-collaborations require a supporting cyber-infrastructure that allows team members to generate, analyze, disseminate, access, and consume information at dramatically increased pace and/or quantity [46]. A key element of this emerging cyber-infrastructure is open source software, as it eliminates collaboration hurdles due to software licenses and

can help foster truly massive e-collaborations. In other words, even with collaborations involving proprietary data, open source cyber-infrastructure provides a common language that can facilitate e-collaborations with large numbers of team members.

Several recent national and international initiatives [15, 6, 7] have been launched with the premise that the adoption and utilization of modern data science and informatics toolsets offers a new opportunity to accelerate dramatically the design and deployment cycle of new advanced materials in commercial products. More specifically, it has been recognized that innovation cyber-ecosystems [68] are needed to allow experts from the materials science and engineering, design and manufacturing, and data science domains to collaborate effectively. The challenge in integrating these traditionally disconnected communities comes from the vast differences in how knowledge is captured, curated, and disseminated in these communities [69]. More specifically, knowledge systems in the materials field are rarely captured in a digital form. In order to create a modern materials innovation ecosystem, it is imperative that we design, develop, and launch novel collaboration platforms that allow automated distilling of materials knowledge from large amounts of heterogeneous data acquired through customized protocols that are necessarily diverse (elaborated next). It is also imperative that this curated materials knowledge is presented to the design and manufacturing experts in highly accessible (open) formats.

Customized materials design has great potential for impacting virtually all emerging technologies, with significant economic consequences [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. However, materials design (including the design of a manufacturing process route) resulting in the combination of properties desired for a specific application is a highly challenging inverse problem due to the hierarchical nature of the internal structure of materials. Material properties are controlled by the materials' hierarchical internal structure as well as physical phenomena with timescales that vary at each of the hierarchical length scales (from the atomic to the macroscopic length scale).

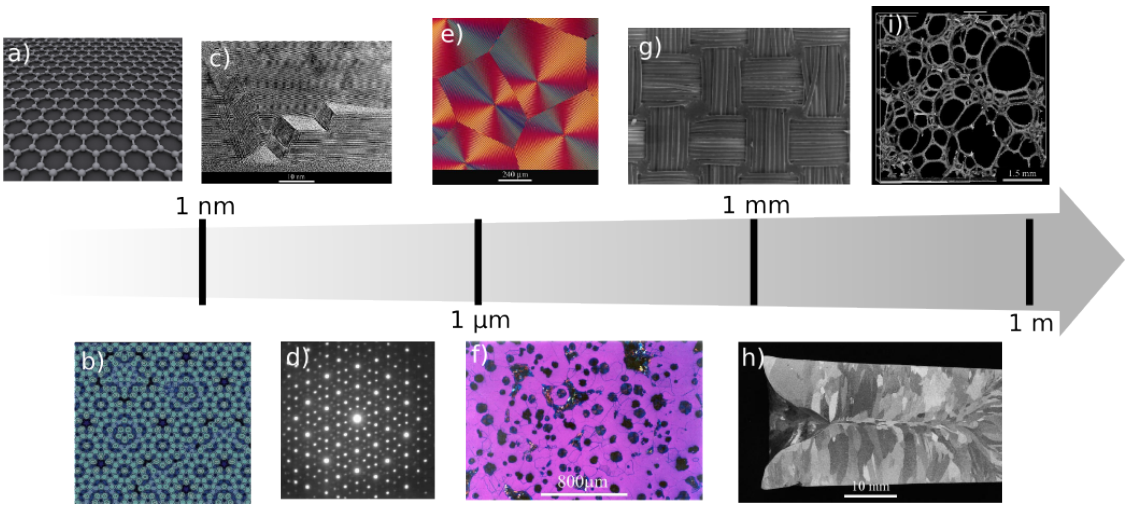


Figure 13: Heirarchical Materials structure at multiple length scales a). Simulated graphene crystalline structure. b). Simulated fivefold icosahedral Al-Ag quasicrystals. c).High resolution electron microscopy image of delamination cracks in h-BN particles subjected to compressive stress in the (0001) planes (within a silicon nitride particulate-reinforced silicon carbide composite). d). Electron diffraction pattern of an icosahedral Zn-Mg-Ho quasicrystal. e). Cross-polarised light image of spherulites in poly-3-hydroxy butyrate (PHB) f). Cast iron with magnesium induced spheroidised graphite. g). SEM micrograph of a taffeta textile fragment. h). Optical microscopy image of a cross-section of an aluminium casting. i). X-ray tomography image of open cell polyurethane foam. Images courtesy of Core-Materials [1].

Characterization of the structure at each of these different length scales is often in the form of images which come from different experimental/computational techniques resulting in highly heterogeneous data. As a result, tailoring the material hierarchical structure to yield desired combinations of properties or performance characteristics is enormously difficult. Figure 13 provides a collection of materials images depicting material structures at different length scales, which are generally acquired using diverse protocols and are captured in equally diverse formats.

While the generation (from experiments and computer simulations) and dissemination of datasets consisting of heterogeneous images are necessary elements in a modern materials innovation ecosystem, there is an equally critical need for customized analytics that take into account the stochastic nature of these data at multiple length

scales in order to extract high value, transferable, knowledge. Data-driven Process-Structure-Property (PSP) linkages [70] provides a systemic, modular, and hierarchical framework for community engagement (i.e., several people making complementary or overlapping contributions to the overall curation of materials knowledge). Computationally cheap PSP linkages also communicate effectively the curated materials knowledge to design and manufacturing experts in highly accessible formats.

The Materials Knowledge Systems in Python project (PyMKS) is the first open source materials data analytics toolkit that can be used to create high value PSP linkages for hierarchical materials in large scale efforts driven and directed by an entire community of users. In this regard, it could be a foundational element of the cyber-infrastructure needed to realize a modern materials innovation ecosystem.

3.2 Current Materials Innovation Ecosystem

Open access materials databases and computational tools are critical components of the cyber-infrastructure needed to curate materials knowledge through effective e-collaborations [71]. Several materials science open source computational toolsets and databases have emerged in recent years to help realize the vision outlined in the Materials Genome Initiative (MGI) and the Integrated Computational Materials Engineering (ICME) paradigm [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Yet, the creation and adoption of a standard materials taxonomy and database schema has not been established due to the unwieldy size of material descriptors and heterogeneous data. Additionally, the coupled physical phenomena that govern material properties are too complex to model all aspects of a material simultaneously using a single computational tool. Consequently, current practices have resulted in the development of computation tools and databases with a narrow focus on specific length/structure scales, material classes, or properties.

The NIST Data Gateway contains over 100 free and paid query-able web-based

materials databases. These databases contain atomic structure, thermodynamics, kinetics, fundamental physical constants, x-ray spectroscopy, among other features [72]. The NIST DSpace provides a curation of links to several materials community databases [73]. The NIST Materials Data Curation Systems (MDCS) is a general online database that aims to facilitate the capturing, sharing, and transforming of materials data [74]. The Open Quantum Materials Database (OQMD) is an open source data repository for phase diagrams and electronic ground states computed using density functional theory [75]. MatWeb is a database containing materials properties for over 100,000 materials [76]. Atomic FLOW of Materials Discovery (AFLOW) databases millions of materials and properties and hosts computational tools that can be used for atomic simulations [77]. The Materials Project (and the tool pyMatgen) [78, 79] provides open web-based access to computed information on known and predicted materials as well as analysis tools for electronic band structures. The Knowledgebase of Interatomic Models (OpenKIM) hosts open source tools for potentials for molecular simulation of materials [80]. PRedictive Integrated Structural Materials Science (PRISMS) hosts a suite of ICME tools and datastorage for the metals community focused on microstructure evolution and mechanical properties [81].

SPPARKS Kinetic Monte Carlo Simulator (SPPARKS) is a parallel Monte Carlo code for on-lattice and off-lattice models [82]. MOOSE is a parallel computational framework for coupled systems of nonlinear equations [83]. Dream3D is a tool used for synthetic microstructure generation, image processing and mesh creation for finite element [84].

While there exits a sizable number of standard analytics tools [85, 86, 49, 87, 88, 89, 90, 91, 92, 93], none of them are tailored to create PSP linkages from materials structure image data and their associated properties. PyMKS aims to seed and

nurture an emergent user group in the materials data analytics field for establishing homogenization and localization (PSP) linkages by leveraging open source signal processing and machine learning packages in Python. An overview of the PyMKS project accompanied with several examples is presented here. This paper is a call to others interested in participating in this open science activity.

3.3 Theoretical Foundations of Materials Knowledge Systems

Material properties are controlled by their internal structure and the diverse physical phenomena occurring at multiple time and length scales. Generalized composite theories [94, 95] have been developed for hierarchical materials exhibiting well separated length scales in their internal structure. Generally speaking, these theories either address homogenization (i.e., communication of effective properties associated with the structure at a given length scale to a higher length scale) or localization (i.e., spatiotemporal distribution of the imposed macroscale loading conditions to the lower length scale). Consequently, homogenization and localization are the essential building blocks in communicating the salient information in both directions between hierarchical length/structure scales in multiscale materials modeling. It is also pointed out that localization is significantly more difficult to establish, and implicitly provides a solution to homogenization.

The most sophisticated composite theory available today that explicitly accounts for the full details of the material internal structure (also simply referred as microstructure) comes from the use of perturbation theories and Green's functions [96, 94, 97, 98, 99, 100, 101, 102, 103, 104, 105, 21, 106]. In this formalism, one usually arrives at a series expansion for both homogenization and localization, where the individual terms in the series involve convolution integrals with kernels based on Green's functions. This series expansion was refined and generalized by Adams and co-workers [107, 21, 108] through the introduction of the concept of a microstructure function,

which conveniently separates each term in the series into a physics-dependent kernel (based on Green's functions) and a microstructure-dependent function (based on the formalism of n-point spatial correlations [100, 101, 102, 103, 104, 105]).

Materials Knowledge Systems (MKS) [25, 24, 27, 23, 22, 2, 33] complement these sophisticated physics-based materials composite theories with a modern data science approach to create a versatile framework for extracting and curating multiscale PSP linkages. More specifically, MKS employs a discretized version of the composite theories mentioned earlier to gain major computational advantages. As a result, highly adaptable and templatable protocols have been created and used successfully to extract robust and versatile homogenization and localization metamodels with impressive accuracy and broad applicability over large microstructure spaces.

The MKS framework is based on the notion of a microstructure function. The microstructure function provides a framework to represent quantities that describe material structure such as phase identifiers, lattice orientation, chemical composition, defect types and densities, among others (typically referred to as local states). The microstructure function, $m_j\left(h,s\right)$, represents a probability distribution for the given local state, $h \in H$, at each position, $s \in S$, in a given microstructure, j [109, 110, 111, 30]. The introduction of the local state space L (i.e., the complete set of all potential local states) provides a consolidated variable space for combining the diverse attributes (often a combination of scalar and tensor quantities) needed to describe the local states in the material structure. The MKS framework requires a discretized description of m_j, which is denoted here as $m_j\left[l,s\right]$, where the $[\cdot,\cdot]$ represent the discretized space (in contrast to $(\cdot,\cdot)$, which defines the continuous space). In most applications, S is simply tessellated into voxels on a regular (uniform) grid so that the position can be denoted by $s \rightarrow i,j,k$ in three dimensions.

As noted earlier, the local state space in most advanced materials is likely to demand sophisticated representations. In prior work [27, 3, 112], it was found that

spectral representations on functions on the local state space offered many advantages both in compact representation as well as in reducing the computational cost. In such cases, l indexes the spectral basis functions employed. The selection of these functions depends on the nature of local state descriptors. Examples include: (i) the primitive basis (or indicator functions) used to represent simple tessellation schemes [25, 24, 23, 22, 2, 33, 109, 110, 113], (ii) generalized spherical harmonics used to represent functions over the orientation space [27, 3], and (iii) Legendre polynomials used to represent functions over the concentration space [112].

3.3.1 Homogenization

Comparing different microstructures is quite difficult even after expressing them in convenient discretized descriptions mainly due to the lack of a reference point or a natural origin for the index s in the tessellation of the microstructure volume. Yet the relative spatial distributions of the local states provide a valuable representation of the microstructure that can be used effectively to quantify the microstructure and compare it with other microstructures in robust and meaningful ways [110, 30, 109, 33, 113]. The lowest order of spatial correlations with relative spatial information comes in the form of 2-point statistics and can be computed as a correlation of a microstructure function such that

$$f_j[l, l', r] = \frac{1}{\Omega_j[r]} \sum_s m_j[l, s] m_j[l', s + r] \tag{43}$$

where r is a discrete spatial vector within the voxelated domain specified by s, $f_j[l, l', r]$ is one set of 2-point statistics for the local stats l and l' and $\Omega_j[r]$ is a normalization factor that depends on r [113]. The subscript j refers to a sample microstructure used for analysis (i.e., each j could refer to a microstructure image). The physical interpretation of the 2-point statistics is explained in Fig. 26 with a highly simplified two-phase microstructure (the two phases are colored white and gray). If the primitive basis is used to discretize both the spatial domain and the

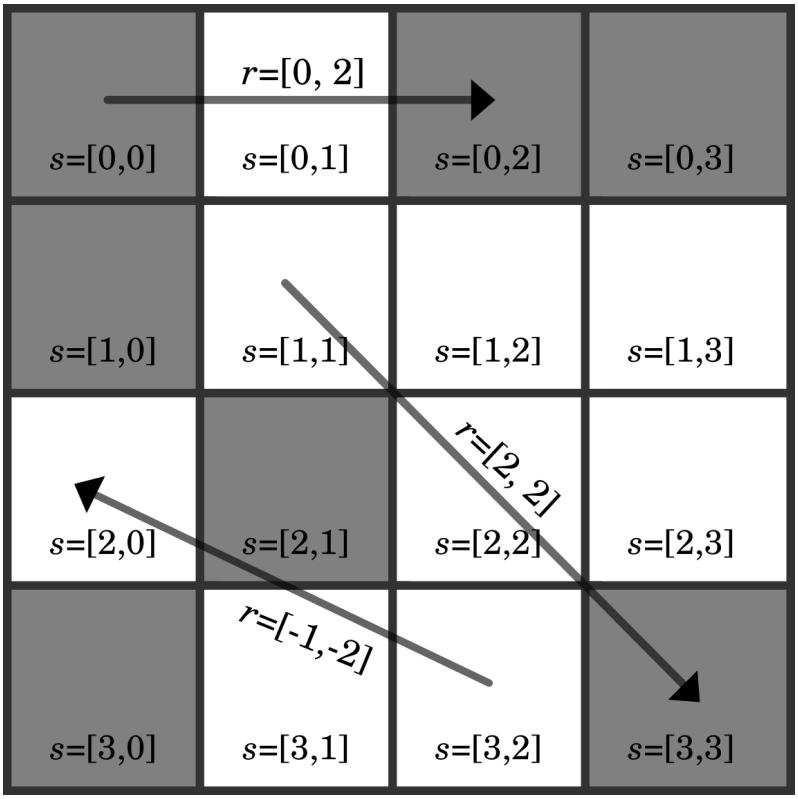


Figure 14: The discretization scheme for both the microstructure function and the vector space needed to define the spatial correlations, illustrated on a simple two-phase composite material. The discretized vectors r describe the relative positions between different spatial locations.

local state space then $f_j[l, l', r]$ can be interpreted as the probability of finding local states l and l' at the tail and head, respectively, of a randomly placed vector r.

2-Point statistics provide a meaningful representation of the microstructure, but create an extremely large feature space that often contains redundant information. Dimensionality reduction can be used to create low dimensional microstructure descriptors from the sets of spatial correlations (based on different selections of l and l') with principal component analysis (PCA). The PCA dimensionality reduction can be mathematically expressed as

$$f_j\left[\nu\right] \approx \sum_{k \in K} \mu_j\left[k\right] \phi\left[k, \nu\right] + \overline{f[\nu]} \tag{44}$$

In Eq. 44, $f_j\left[\nu\right]$ is a contracted representation of $f_j\left[l, l', r\right]$ as a large vector (i.e., ν maps uniquely to every combination of l, l' and r deemed to be of interest in the analyses). The $\mu_j[k]$ are low dimensional microstructure descriptors (the transformed 2-point statistics) or principal component scores (PC scores). The $\phi\left[k, \nu\right]$ are the calibrated principle components (PCs) and the $\overline{f[\nu]}$ are the mean values from the calibration ensemble of $f_j\left[\nu\right]$ for each ν. The $k \in K$ indices refer to the $\mu_j\left[k\right]$ in decreasing order of significance and are independent of l, l' and r. The main advantage of this approach is that the $f_j\left[\nu\right]$ can be reconstructed to sufficient fidelity with only a small subset of $\mu_j\left[k\right]$ [32].

After obtaining the needed dimensionality reduction in the representation of the material structure, machine learning models can be used to create homogenization PSP linkages of interest. As an example, a generic homogenization linkage can be expressed as

$$p_j^{\text{eff}} = \mathcal{F}(\mu_j[k]) \tag{45}$$

In Eq. 45, p_j^{eff} is the effective materials response (reflecting an effective property in structure-property linkages or an evolved low dimensional microstructure descriptor in process-structure linkages), and $\mathcal{F}$ is a machine learning function that links $\mu_j[k]$

to p_j^{eff}.

3.3.2 Localization

MKS Localization linkages are significantly more complex than the homogenization linkages. These are usually expressed in the same series forms that are derived in the general composite theories, while employing discretized kernels based on Green's functions [96, 94, 97, 98, 99, 100, 101, 102, 103, 104, 105, 21, 106]. Mathematically, the MKS localization linkages are expressed as

$$p_j[s] = \sum_{l,r} \alpha[l,r] m_j[l, s-r] + \sum_{l,l',r,r'} \alpha[l,l',r,r'] m_j[l, s-r] m_j[l', s-r'] + ... \quad (46)$$

In Eq. 46, $p_j[s]$ is the spatially resolved (localized) response field (e.g., a response variable such as stress or strain rate in a structure-property linkage, or an evolved microstructure function in a process-structure linkage), and $\alpha[l,r]$ are the Green's function based discretized influence kernels. These digital kernels are calibrated using regression methods [23, 24, 25, 27, 3, 112].

Fig. 3.3.2 provides schematic overviews of the MKS homogenization and localization workflows. More detailed explanations on the MKS homogenization and localization linkages can be found in prior literature [25, 24, 27, 23, 22, 2, 112, 33, 109, 110, 113].

3.4 Materials Knowledge Systems in Python

PyMKS is an object-oriented numerical implementation of the MKS theory developed in the literature [24]. It provides a high-level, computationally efficient, framework to implement data pipelines for classification, cataloging and quantifying materials structures for PSP relationships. PyMKS is written in Python, a natural choice for scientific computing due to its ubiquitous use among the data science community as well as many other favorable attributes [114]. PyMKS is licensed under the permissive

MIT license [115] which allows for unrestricted distribution in commercial and non-commercial systems.

3.4.1 Core Functionality

PyMKS consists of four main components including a set of tools to compute 2-point statistics, tools for both homogenization and localization linkages and tools for discretizing the microstructure. In addition, PyMKS has modules for generating data sets using conventional numerical simulations and a module for custom visualization of microstructures. PyMKS builds on Scikit-learn's pipelining methodology to create materials specific machine learning models. This is a high level system for combining multiple data and machine learning transformations into a single customizable pipeline with only minimal required code. This approach makes cross-validation and parameter searches simple to implement and avoids the complicated book keeping issues associated with training, testing and validating data pipelines in machine learning.

The starting point for an MKS homogenization analysis is to use 2-point statisics as outlined in Eq. 48 and provided in PyMKS by the `MKSStructureAnalysis` object, which calculates the objective low dimensional structure descriptors, $\mu_j[k]$. The default dimensionality reduction technique is PCA, but any model that uses the `transform_fit` or a "transformer" object can be substituted. After calculating the descriptors, the `MKSHomogenizationModel` is used to create linkages between the $\mu_j[k]$ and the effective material response, p_j^{eff}, as indicated in Eq. 45. The default machine learning algorithm is a polynomial regression, but any estimator with the `fit` and `predict` methods can be substituted to create the linkages between $\mu_j[k]$ and p_j^{eff}.

The `MKSLocalizationModel` object provides the MKS localization functionality. It calibrates the first order influence kernels $\alpha[l, r]$ used to predict local materials responses, $p_j[s]$, as outlined in Eq. 46. The calibration of the influence kernels is

achieved using a variety of linear regression techniques described in numerous previous studies [25, 24, 27, 112]. The `MKSLocalizationModel` object uses `fit` and `predict` methods to follow the standard interface for a Scikit-learn estimator object.

To use either the homogenization or the localization models in PyMKS, the microstructure first needs to be represented by a microstructure function, $m_j[l, s]$. The `bases` module in PyMKS contains four transformer objects for generating the $m_j[l, s]$ using a varietly of discretization methods [25, 24, 27, 23, 22, 2, 33, 112]. These four objects can be thought of as materials specific extension to the feature extraction module in Scikit-learn. A `PrimitiveBasis` object uses indicator (or hat) functions and is well suited for microstructures that have discrete local states (e.g., distinct thermodynamic phases). The `LegendreBasis` and `FourierBasis` objects create spectral representations of microstucture functions defined on nonperiodic and periodic continuous local state spaces, respectively. For example, functions over a range of chemical compositions can be described using `LegendreBasis`, while functions over orientations in two-dimensional space can be described using `FourierBasis`. Furthermore, `GSHBasis` creates compact spectral representations for functions over lattice orientation space (such as those needed to describe polycrystalline microstructures) [50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61].

PyMKS contains modest data generation tools (in the `datasets` module) that are used in both the PyMKS examples and the PyMKS test suite. The `MicrostructureGenerator` object creates stochastic microstructures using digital filters. This assists users in creating PyMKS workflows even when data is unavailable. PyMKS has objects for generating sample data from both a spinodal decomposition simulation (using the `CahnHilliardSimulation` object) and a linear elasticity simulation (using the `ElasticFESimulation` object). PyMKS comes with custom functions for visualizing microstructures in elegant ways (in the `tools` module). These are used extensively

in the PyMKS example notebooks to minimize incidental code associated with visualization.

3.4.2 Underlying Technologies

PyMKS is built upon the highly optimized Python packages NumPy [47], SciPy [48], and Scikit-learn [49]. NumPy arrays are the primary data structure used throughout PyMKS and provide the basic vector and matrix manipulation operations. SciPy's signal processing and numerical linear algebra functions are used to calibrate models and generate synthetic data. PyMKS is highly integrated with Scikit-learn and mimics its simple API in order to leverage from Scikit-learn's data pipeling methodology for machine learning and data transformations. In addition, PyMKS uses the Pytest framework to automate execution of the test suite [116].

Optional packages that can be used with PyMKS include Simple Finite Elements in Python (SfePy) [117], the python wrapper for the FFTW library (pyFFTW) [118] and the plotting package Matplotlib [119]. SfePy is used to simulate linear elasticity to create sample response field data. PyFFTW is a hightly optimized Fast Fourier Transform library that enhances the efficiency of PyMKS and enables parallel computations in PyMKS. Matplotlib is used to generate custom microstructure visualizations.

3.4.3 Development Practices

PyMKS leverages from existing tools, standards and web resources wherever possible. In particular the developers are an open community that use GitHub for issue tracking and release management (see `https://github.com/materialsinnovation/pymks`). Additionally a Google group is used as a public forum to discuss the project development, support and announcements (see `pymks-general@googlegroups.com`). The Travis CI continuous integration tool is used to automate running the test suite for branches of the code stored on GitHub. Code standards are maintained by following

the Python PEP8 standards and by reviewing code using pull requests on GitHub. Detailed administrative guidelines are outlined in the `ADMINISTRATA.md` document, and potential developers are encouraged to follow them.

3.5 Examples of Homogenization and Localization with PyMKS

A demonstration of the MKS homogenization and localization workflows as shown in Fig. 3.3.2 are presented in this section using PyMKS. Additional workflow examples can be found on the PyMKS website `pymks.org`.

3.5.1 Prediction of Effective Stiffness with Homogenization

3.5.1.1 Generation of Calibration Data

In this example the `MKSHomogenizationModel` is used to create a structure-property linkage between a 2-phase composite material and effective stiffness C_{xx}.

Multiple classes of periodic microstructures and their effective elastic stiffness values can be generated by importing the `make_elastic_stiffness` function from `pymks.datasets`.

This function has several arguments. `n_samples` is a list indicating the number of microstructures for each class. `grain_size` and `volume_fraction` are also lists that specify the average grain features and mean volume fractions for each of the microstructure classes. Variance in the volume fractions for each class can be controlled using `percent_variance` which specifies a range of volume fractions centered about the mean values (i.e. `volume_fration` $\pm$ `percent_variance`). `size` indicates the dimensions of all the microstructures. `elastic_modulus` and `poissons_ratio` are used to indicate the material properties for each of the phases. Lastly, `seed` is used as the seed for the random number generator.

In this homogenization example, 50 samples from 16 different microstructures classes with dimensions 21 x 21 and their effective stiffness values were created totaling to 800 samples. Each of the 16 classes have different sized microstructure features and

volume fractions. The `make_elastic_stiffness` function returns the microstructures `X` and their associated stiffness values `y`.

```
from pymks.datasets import make_elastic_stiffness
import numpy as np

sample_size = 50
n_samples = [sample_size] * 16

grain_size = [(8, 8), (8, 6), (6, 8), (6, 6),
              (10, 4), (4, 10), (4, 4), (10, 10),
              (12, 2), (2, 12), (2, 2), (12, 12),
              (14, 1), (1, 14), (1, 1), (14, 14)]

volume_fraction = [(0.8, 0.2), (0.7, 0.3), (0.6, 0.4), (0.5,
   0.5),
                   (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), (0.5,
   0.5),
                   (0.8, 0.2), (0.7, 0.3), (0.6, 0.4), (0.5,
   0.5),
                   (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), (0.5,
   0.5)]

percent_variance = 0.15
elastic_modulus = (300, 200)
poissons_ratio = (0.28, 0.3)
size = (21, 21)
seed = 1

X, y = make_elastic_stiffness(n_samples=n_samples,
                              volume_fraction=volume_fraction,
                              grain_size=grain_size,  size=
   size,
                              percent_variance=
   percent_variance,
                              elastic_modulus=elastic_modulus,
                              poissons_ratio=poissons_ratio,
                              seed=seed)
```

An example microstructure from each of the 16 classes can be visualized by

importing `draw_microstructures` function from `pymks.tools`. The output from `draw_microstructures` can be found in Fig. 16.

```
from pymks.tools import draw_microstructures

X_examples = X[1::sample_size]
draw_microstructures(X_examples, figsize=(4, 4))
```

3.5.1.2 Calibration of Homogenization Model

Before an instance of the `MKSHomogenizationModel` can be made, an instance of a basis class is needed to specify the discretization method for the microstructure functions (see Fig. 3.3.2). For this particular example, there are only 2 discrete phases numerated by 0 and 1. It has been shown that the primitive basis provides the most compact representation of discrete phases [109, 110, 30, 113, 33, 25, 25, 23]. In PyMKS the class `PrimitiveBasis` from `pymks.bases` can be used with `n_states` equal to `2` and the `domain` equal to `[0, 1]`.

The periodic axes as well as the set(s) of spatial correlations need to be specified in addition to the basis class for the `MKSHomogenizationModel`. This is done using the arguments `periodic_axes` and `correlations` respectively . In practice the set of spatial correlations are a hyper parameter of our model that could be optimized, but for this example only the two autocorrelations will be used.

```
from pymks import MKSHomogenizationModel
from pymks import PrimitiveBasis

prim_basis = PrimitiveBasis(n_states=2, domain=[0, 1])
model = MKSHomogenizationModel(basis=prim_basis,
                               periodic_axes=[0, 1],
                               correlations=[(0, 0), (1, 1)])
```

The default pipeline used to create the homogenization linkage uses PCA and polynomial regression objects for Scikit-learn. Using `GridSearchCV` from Scikit-learn, cross validation is used on the testing data to find the optimal number of

principal components and degree of polynomial (based on the R-squared values) within a defined subspace for the hyper parameters for our model. A dictionary `params_to_tune` defines the subspace. For this example `n_components` will be varied between 1 to 13 and `degree` of the polynomial regression will be varied between 1 to 3. `StratifiedKFold` is used to ensure that microstructures from each of the classes are used for each fold during cross validation. The array `labels` is used to label each of the classes.

```
from sklearn.cross_validation import StratifiedKFold
from sklearn.grid_search import GridSearchCV

flat_shape = (X.shape[0],) + (X[0].size,)
params_to_tune = {'degree': np.arange(1, 4),
                  'n_components': np.arange(1, 13)}
labels = np.repeat(np.arange(16), 50)
skf = StratifiedKFold(labels, n_folds=5)

fit_params = {'size': X[0].shape}
gs = GridSearchCV(
    model, params_to_tune, cv=skf,
    \usepackage{courier}fit_params=fit_params).fit(X.reshape(
   flat_shape), y)
```

The results of our parameter grid search can be examined by either printing or creating visualizations. The parameters and score of the best estimator can be printed as shown below.

```
from __future__ import print_function

print('Order of Polynomial', gs.best_estimator_.degree)
print('Number of Components', gs.best_estimator_.n_components)
print('R-squared Value', gs.score(X, y))
```

```
Order of Polynomial 2
Number of Components 11
R-squared Value 0.999960653331
```

Two different visualizations of the results from `GridsearchCV` with can be created using `draw_gridscores_matrix` and `draw_gridscores` from `pymks.tools`.

`draw_gridscores_matrix` provides a visualization of two matrices for both the mean R-squared values and their standard deviation. The output from `draw_gridscores_matrix` can be found in Fig. 17.

```
from pymks.tools import draw_gridscores_matrix

draw_gridscores_matrix(gs, ['n_components', 'degree'],
                       score_label='R-Squared',
                       param_labels=['Number of Components',
                                     'Order of Polynomial'])
```

`draw_gridscores` provides another view of the same information with the mean values indicated by the points and the standard deviation indication by the shared regions. The output from `draw_gridscores` can be found in Fig. 18.

```
from pymks.tools import draw_gridscores

gs_deg_1 = [x for x in gs.grid_scores_ \
            if x.parameters['degree'] == 1]
gs_deg_2 = [x for x in gs.grid_scores_ \
            if x.parameters['degree'] == 2]
gs_deg_3 = [x for x in gs.grid_scores_ \
            if x.parameters['degree'] == 3]

draw_gridscores([gs_deg_1,  gs_deg_2, gs_deg_3], 'n_components
   ',
                data_labels=['1st Order',
                             '2nd Order', '3rd Order'],
                param_label='Number of Components',
                score_label='R-Squared')
```

For the specified parameter range, the model with the highest R-squared value was found to have a 2nd order polynomial with 11 principal components. This model is calibrated using the entire training dataset and is used for the rest of the example.

```
model = gs.best_estimator_

model.fit(X, y)
```

3.5.1.3 Prediction of Effective Stiffness Values

In order to validate our model, additional data is generated using the `make_elastic_stiffness` function again with the same parameters with the exception of the number of samples and the seed used for the random number generator. The function returns the new microstructure `X_new` and their effective stiffness values `y_new`.

```
test_sample_size = 10
n_samples = [test_sample_size] * 16

seed = 0

X_new, y_new = make_elastic_stiffness(
    n_samples=n_samples, size=size,
    grain_size=grain_size,
    volume_fraction=volume_fraction,
    percent_variance=percent_variance,
    elastic_modulus=elastic_modulus,
    poissons_ratio=poissons_ratio,
    seed=seed)
```

Effective stiffness values predicted by the model for the new data are generated using the `predict` method.

```
y_pred = model.predict(X_new)
```

A visualization of the PC scores for both the calibration and the validation data can be created using `draw_components_scatter` from `pymks.tools`. The output from `draw_components_scatter` can be found in Fig. 19

Because both the validation and the calibration data were generated from the `make_elastic_stiffness` function with the same parameters both sets of data are

different samples from the same distribution. Similar visualizations can provide insights on differences between different data sources.

```
from pymks.tools import draw_components_scatter

draw_components_scatter([model.reduced_fit_data[:, :2],
                         model.reduced_predict_data[:, :2]],
                        ['Training Data', 'Test Data'],
                        legend_outside=True)
```

To evaluate our model's predictions, a goodness-of-fit plot can be generated by importing `draw_goodness_of_fit` from `pymks.tools`. The results from `draw_goodness_of_fit` can be found in Fig. 20. Additionally the R-squared value for our predicted data can be printed.

```
from pymks.tools import draw_goodness_of_fit

fit_data = np.array([y, model.predict(X)])
pred_data = np.array([y_new, y_pred])
draw_goodness_of_fit(fit_data, pred_data,
                     ['Training Data', 'Test Data'])
```

```
print('R-squared value', model.score(X_new, y_new))
```

```
R-squared value 0.999949544961
```

3.5.2 Prediction of Local Strain Field with Localization

3.5.2.1 Generation of Calibration Data

In this example the `MKSLocalizationModel` is used to predict the local strain field for a three phase microstructure with elastic moduli values of 80 MPa, 100 MPa and 120 MPa; Poisson's ratio values all equal to 0.3 and a macroscopic imposed strain equal to 0.02. The model is calibrated using delta microstructures (analogous to using a unit impulse response to find the kernel of a system in signal processing) [25]. The the material parameters specified above are used in a finite element simulation using

the make_elasticFEstrain_delta function from pymks.datasets. The number of Poisson's ratio and elastic moduli values indicates the number of phases.

```
from pymks.datasets import make_elastic_FE_strain_delta
import numpy as np

n = 21
n_phases = 3

elastic_modulus = (80, 100, 120)
poissons_ratio = (0.3, 0.3, 0.3)
macro_strain = 0.02
size = (n, n)

X_delta, strains_delta = make_elastic_FE_strain_delta(
    elastic_modulus=elastic_modulus,
    poissons_ratio=poissons_ratio,
    size=size, macro_strain=macro_strain)
```

Delta microstructures are composed of only two phases with the center of the microstructure being a different phase from the rest. All permutations of the delta microstructures and their associated strain fields ε_{xx} are needed to calibrate the localization model. A delta microstructure and it's strain field can be visualized using draw_microstructure_strain from pymks.tools. The output from draw_microstructure_strain can be found in Fig. 21.

```
from pymks.tools import draw_microstructure_strain

draw_microstructure_strain(X_delta[0], strains_delta[0])
```

3.5.2.2 Calibration of the Localization Model

In order to make an instance of the MKSLocalizationModel, an instance of a basis class must first be created to specify the discretization method for the microstructure function (see Fig. 3.3.2). For this particular example, there are 3 discrete phases,

therefore the `PrimitiveBasis` from `pymks.bases` will be used. The phases are enumerated by 0, 1 and 2, therefore we have three local states with a domain from 0 to 2. An instance of the `PrimitiveBasis` with these parameters can be used to create an instance of the `MKSLocalizationModel` as follows.

```
from pymks import MKSLocalizationModel
from pymks import PrimitiveBasis

p_basis =PrimitiveBasis(n_states=3, domain=[0, 2])
model = MKSLocalizationModel(basis=p_basis)
```

With the delta microstructures and their strain fields, the influence kernels can be calibrated using the `fit` method. A visualization of the influence kernels can be generated using the `draw_coeff` function from `pymks.tools`. The results from `draw_coeff` can be found in Fig. 22.

```
from pymks.tools import draw_coeff

model.fit(X_delta, strains_delta)
draw_coeff(model.coef_)
```

3.5.2.3 Prediction of the Strain Field for a Random Microstructure

Model validation is done by comparing strain fields computed using a finite element simulation and our localization model for the same random microstructure. The `make_elasticFEstrain_random` function from `pymks.datasets` generates a random microstructure and its strain field results from finite element analysis. The output from `make_elasticFEstrain_random` are visualized using `draw_microstructure_strain` can be found in Fig. 23.

```
from pymks.datasets import make_elastic_FE_strain_random

np.random.seed(101)

X, strain = make_elastic_FE_strain_random(
    n_samples=1, elastic_modulus=elastic_modulus,
```

```
    poissons_ratio=poissons_ratio, size=size,
    macro_strain=macro_strain)

draw_microstructure_strain(X[0] , strain[0])
```

The localization model predicts the strain field by passing the random microstructure to the `predict` method. A visualization of the two strain fields from both the localization model and finite element analysis can be created using `draw_strains_compare` from `pymks.tools`. The output from `draw_strains_compare` can be found in Fig. 24.

```
from pymks.tools import draw_strains_compare

strain_pred = model.predict(X)
draw_strains_compare(strain[0], strain_pred[0])
```

These examples demonstrate the high level code that creates accurate and computationally efficient homogenization structure-property linkages using `MKSHomogenizationModel` and localization linkages using `MKSLocalizationModel` with PyMKS.

3.6 Conclusion

The MKS framework offers a practical and computationally efficient approach for distilling and disseminating the core knowledge gained from physics-based simulations and experiments using emerging concepts in modern data science. PyMKS is an open source project with a permissive license that provides simple high level APIs to access the MKS framework by implementing pipelines from Scikit-learn with customized objects for data from hierarchical materials. PyMKS has been launched with the aim to nucleate and grow an emergent community focused on establishing data-driven homogenization and localization Process-Structure-Property linkages for hierarchical materials.

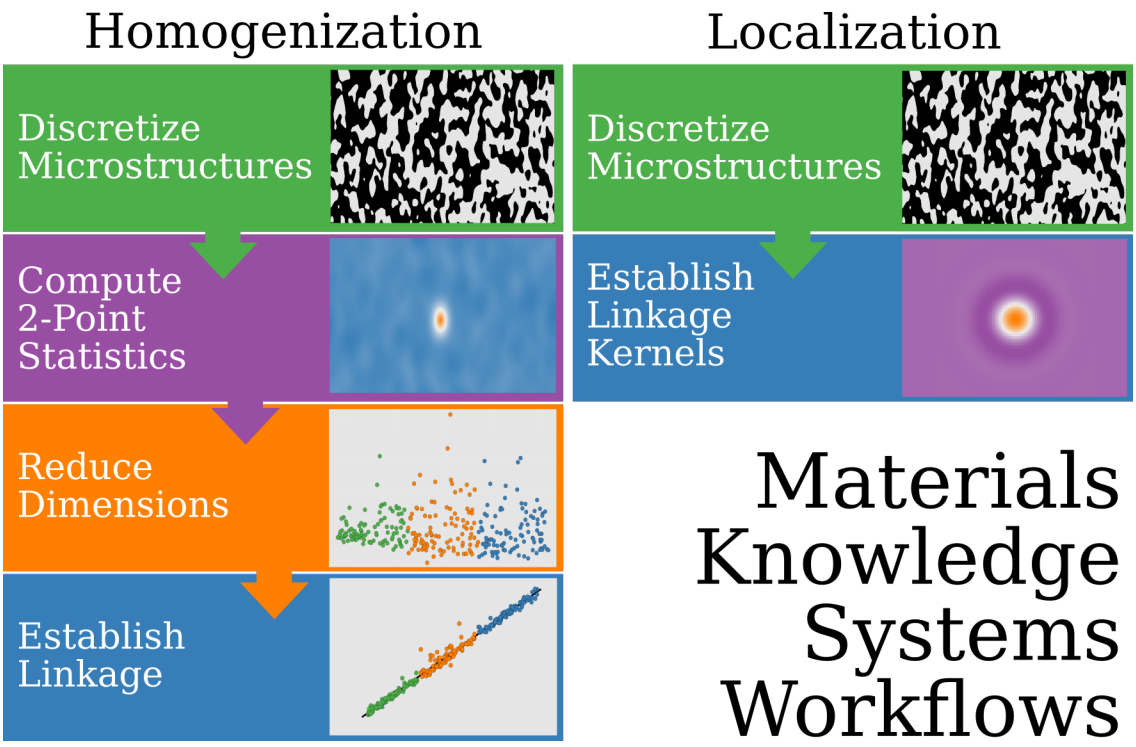


Figure 15: The MKS Homogenization workflow (left) consists of four steps. 1. Discretize the raw microstructure with the microstructure function. 2. Compute 2-point statistics using local states (Eq. 48). 3. Create low dimensional microstructure descriptors using dimensionality reduction techniques (Eq. 44). 4. Establish a linkage with low dimensional microtructure descriptors using machine learning. (Eq. 45). The MKS Localization workflow (right) consists of 2 steps. 1. Discretize the raw microstructure with the microstructure function. 2. Calibrated physics-based kernels using regression methods (Eq. 46).

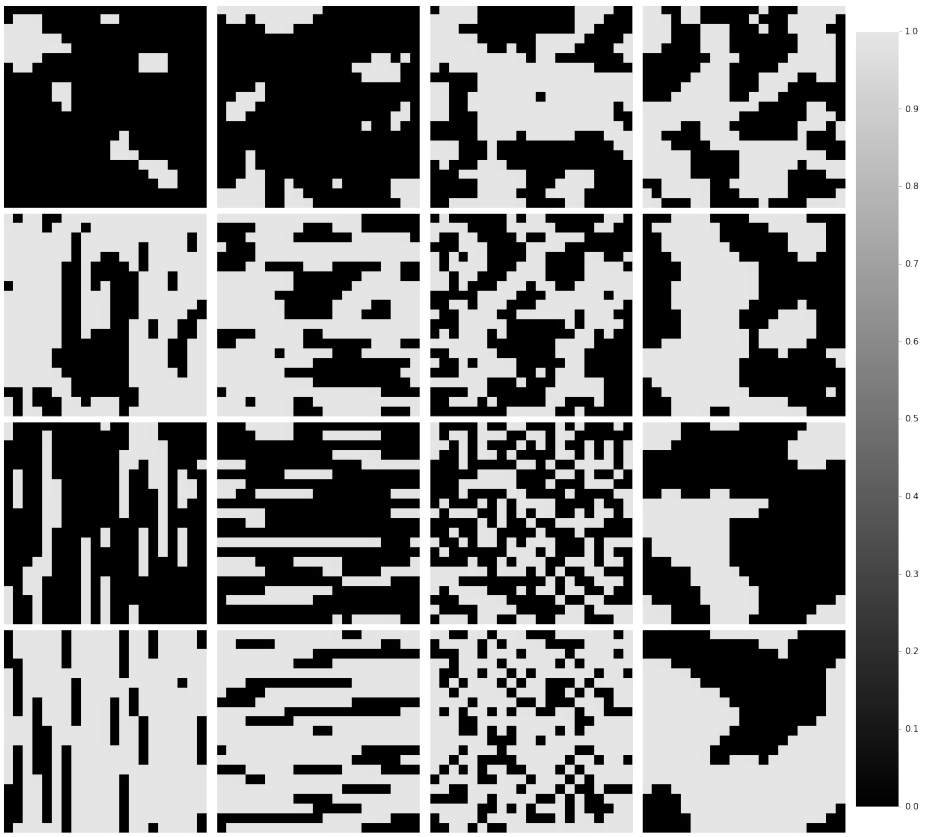


Figure 16: One sample from each of the 16 difference microstructure classes used for calibration of the homogenization model.

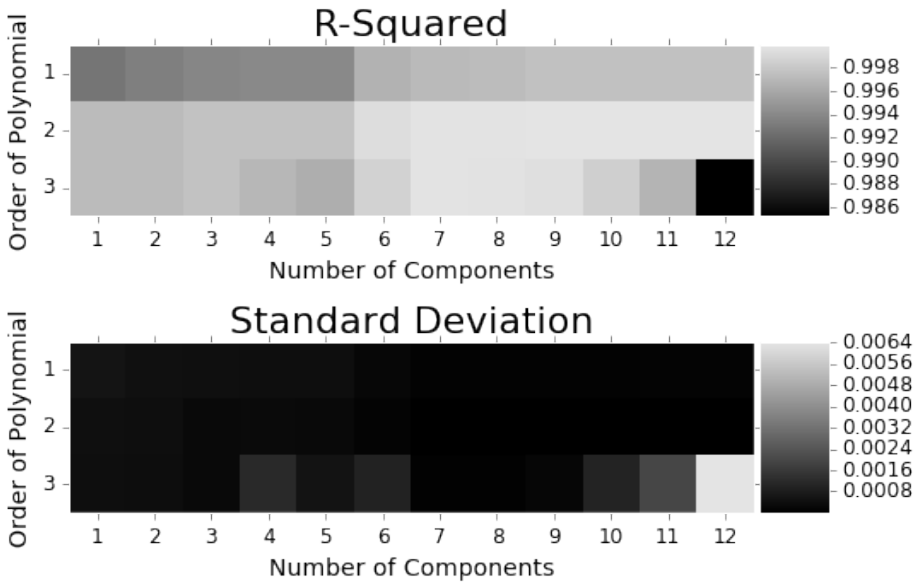


Figure 17: Mean R-Squared values and Standard deviation as a function of the order of the polynomial and the number of principal components.

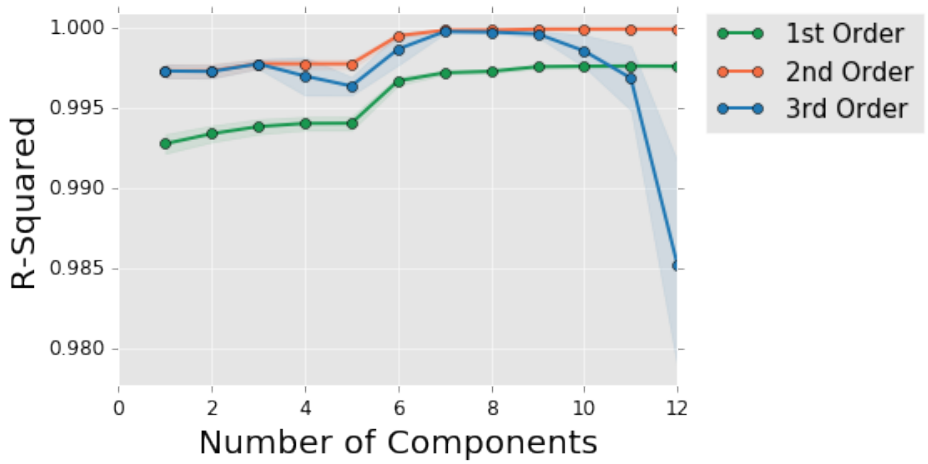


Figure 18: The mean R-Squared values indicated by the points and the standard deviation indication by the shared regions as a function of the number of principal components for the first three orders of a polynomial function.

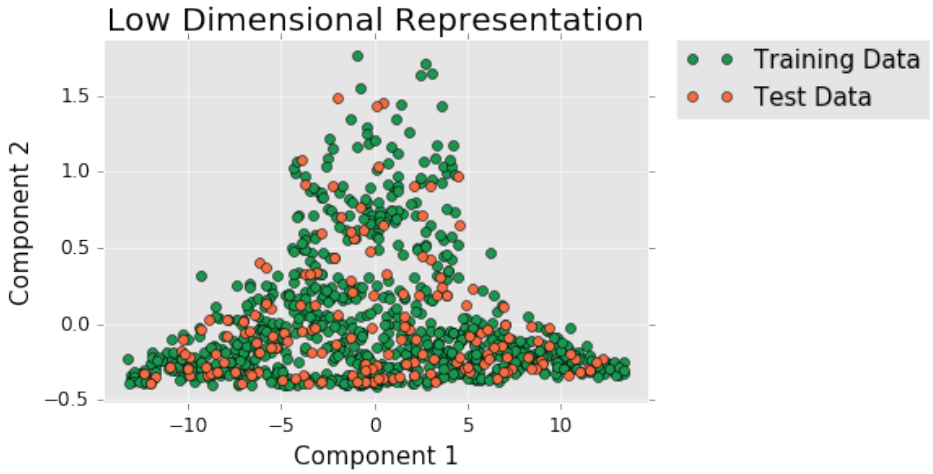


Figure 19: Low dimensional microstructure distributions ($\mu_j[k]$ from Eq. 44) for both the calibration and validation datasets.

Figure 20: Goodness-of-Fit plot for effective stiffness C_{xx} for the homogenization model.

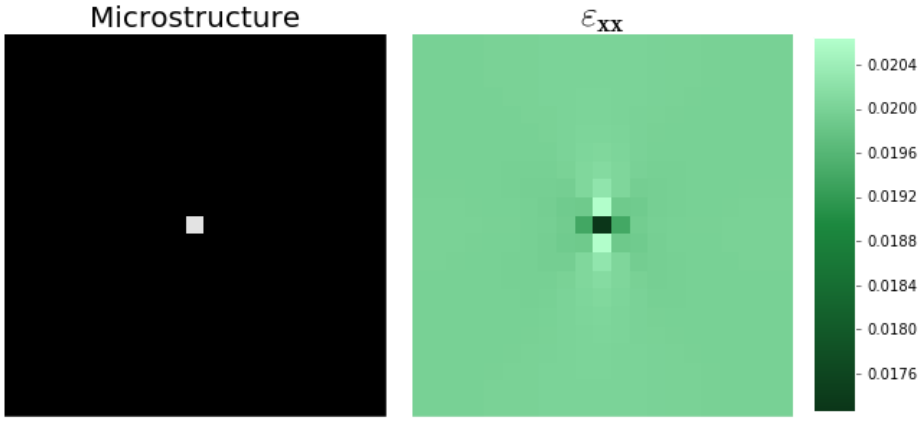


Figure 21: Delta microstructure (right) and its associated strain field (left). The delta microstructures and their local response fields are used to calibrated the localization model.

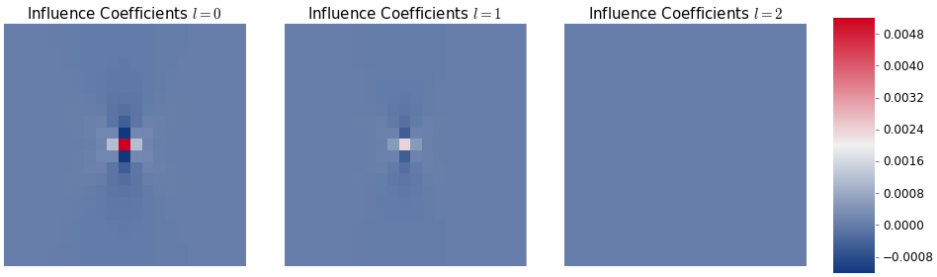


Figure 22: Calibrated influence kernels for the localization model.

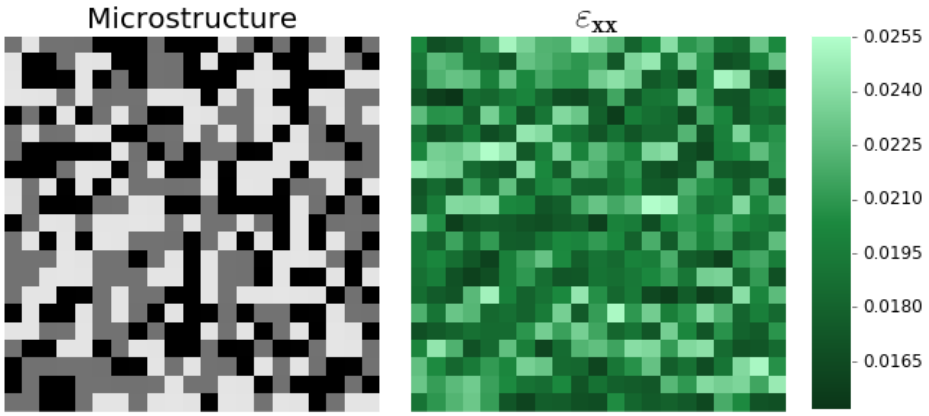


Figure 23: Random microstructure and its local strain field found using finite element analysis.

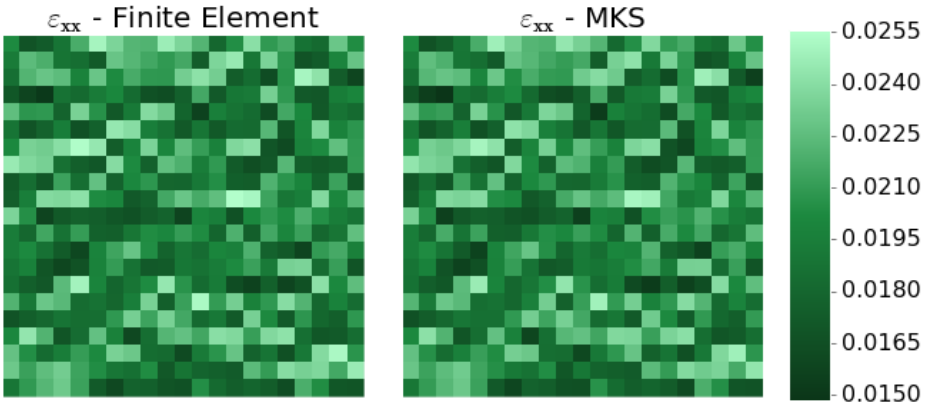


Figure 24: A comparison between the local strain field computed using finite element (left) and the prediction from the localization model (right).

CHAPTER IV

PROCESS-STRUCTURE EVOLUTION LINKAGE USING DIMENSIONALITY REDUCTION AND TIME SERIES ANALYSIS

4.1 Introduction

The discovery and curation of Process-Structure-Property (PSP) linkages in advanced materials and their efficient communication towards the design and manufacturing experts are the main limiting steps in the accelerated exploitation of new materials in emerging technologies. While it has been recognized that an accelerated design cycle for advanced materials can have a significant economic impact [6, 7, 13, 8, 9, 10, 11, 12], in practice the design cycle often takes decades. Some of the difficulties encountered include materials property dependence on extreme values of microstructure distributions, metastablity of microstructures during processing and/or use, variations in data collection protocols, and uncertainty in data, models and model parameters [120, 69]. Additionally, the multiscale (or hierarchical) nature of materials structure necessitates a high dimensional representation and poses the central challenge in establishing high value PSP linkages [121, 112, 22, 69]. The large descriptor space needed to capture the salient details of the material structure creates a major challenge that, in turn, demands a significant amount of data analysis in extracting reliable and useful PSP linkages.

This existing challenge is exacerbated in the case of materials such as polymers where the added complexity arising from minute differences in chemistry and chemical composition can produce dramatic effects in properties across members of the same

polymer family. Consider for example, the case of the commodity polymer polyethylene resin (PE) chemically summarized as $(CH_2\text{-}CH_2)_n$. The same chemical formula represents a broad family of materials that span an application range from grocery bags to bulletproof vests.

In reality, the simple chemical structure actually represents a family of materials which can be subclassified into a variety of grades on the basis of factors inculding but not restricted to density, crystallinity, average molecular chain length and extent of chain branching. The choice of processing conditions under which grades of PE resin are converted from raw material to finished product influence the hierarchy of structural assembly from nanoscale to microscale domains and ultimately have a profound impact on the macroscopic properties. Constructing the necessary PSP linkages for polymers to accelerate materials development is non trivial. Firstly, the quantification of the structure requires detailed experimental data at the corresponding hierarchical length scales. Secondly, dynamic datasets capable of characterizing the changes in properties with corresponding changes in structure are required.

The recently developed Materials Knowledge Systems (MKS) combines concepts from sophisticated physics-based composite theories [97, 98], signal processing [122], and machine learning [123, 124, 125] to establish a new framework for pursuing PSP linkages. These linkages can be established at separable time and length scales relevant to the materials hierarchical structure in order to communicate the salient information for both the top-down (referred to as localization) and the bottom-up (referred to as homogenization) scale-bridging. The Python package PyMKS [126] provides the code base to efficiently establish these linkages.

The MKS framework has thus far been applied largely to capturing structure-property linkages from data generated by multiscale models [25, 24, 27, 23, 22, 2].

These structure-property linkages are, in general, less complex than the process-structure linkages as they do not require a rigorous treatment of the structure evolution over time. In other words, the extension of the MKS framework to process-structure linkages necessitates the introduction of time series analysis.

The extension to process-structure relationships is an important and critical component of the MKS framework. Only with this extension, is it actually possible to formulate a complete and comprehensive set of PSP linkages needed in a materials innovation effort. With the complete formulation of PSP linkages (typically in the form of metamodels or surrogate models), it is possible to address inverse problems in materials and/or process design, where the goal is to identify a process recipe capable of producing a material with improved combination of properties or performance metrics. Furthermore, such surrogate models lend themselves to an integrated community effort for curating and sharing material knowledge (in the form of PSP linkages) at different length and time scales which can also be effectively communicated and digested by the manufacturing experts.

4.2 Background

The MKS framework provides templatable protocols that can be used to create PSP linkages. These protocols start by introducing the concepts of the microstructure function and the local state space. The simplest interpretation of the local state is the thermodynamic state variables (or order parameters) needed to uniquely define a material system such as (orientation, chemical concentration, crystal structure, phase, etc). The local state space defines the space of all possible local states used to define a material system for a given problem. The microstructure function is represented by $m_j[l, s, n]$ and is a function of discretized representations of space s, time n and local state l. The subscript j denotes a particular sample. The introduction of the

microstructure function provides two benefits. It introduces a probabilistic interpretation of the microstructure by converting the structure into a probability distribution where the expectation value is the measured material structure as shown in Eq. (47).

$$E_j[l|s,n] = \sum_{l \in L} l m_j[l,s,n] \tag{47}$$

Additionally the local state space introduces a consolidated descretized variable space L where both tensoral and scalar quantities used to define the materials structure can be mapped into and are represented by a scalar value l.

The homogenization protocol starts by converting the raw structure information into the microstructure function (also referred to as digitizing the microstructure) based on the local states (e.g., phase identifier, chemical composition, lattice orientation). An idealized example of discretized microstructure with two discrete local states can be found in Fig. 25. In this example, the image is segmented such that each voxel is assigned to a particular local state.

With the microstructure function, we can now look at computing spatial correlations between the local states. A 2-point statistic is the probability of finding the head of the vector r in one local state l and the tail of the vector in another local state l' over the entire image. A set of 2-point statistics is defined by all possible vectors that can be defined within an image [22, 110, 109, 113]. 2-point statistics $f_j[l,l',r,n]$ are efficiently computed in the form of spatial correlations using Discrete Fourier Transforms as shown in Eq. (48).

$$f_j[h,h',r,n] = \frac{1}{\Omega_j[r,n]} \sum_{s \in S} m_j[h,s,n] m_j[h',s+r,n] \tag{48}$$

In Eq. (48), $\Omega_j[r,n]$ is a normalization factor. The transformation of the materials structure information into 2-point statistics provides the benefit of creating a natural origin or a point of reference that can be used to objectively compare differences in the relative spatial arrangements of the local stats in each microstructure as shown

in Fig. 26.

Creating structure property linkages or classification models with the raw 2-point statistics is difficult due to the large feature space it creates as well as the collinear features that are present due to redundant information. Low dimensional microstructure descriptors can be created using dimensionality reduction techniques such as Principal Component Analysis (PCA) or one of its variants [32, 127]. PCA creates low dimensional microstructure descriptors using linear combinations of the 2-point statistics as shown in Eq. (49).

$$f_j[\nu, n] \approx \sum_{k=1}^{K} \mu_j[k, n]\phi[k, \nu] + \overline{f[\nu]} \tag{49}$$

In Eq. (49) $f_j[\nu, n]$ is a feature vector that includes all 2-point statistics deemed important for the problem (this includes all different combination of l, l', and r, which are all mapped onto ν), n is a discrete time variable, and j identifies a particular microstructure. $\phi[k, \nu]$ and $\overline{f[\nu]}$ are the calibrated PCs and the mean values of the selcted features respectively. K is the total number of PCs. $\mu_j[k, n]$ are the PC scores at time n, and are taken as the low dimensional microstructure descriptors for sample j. Often most of the variance in a dataset can be captured with a small number of low dimensional descriptors compared to the number of features. Previous studies have used these protocol to create structure-property linkages with effective properties or microstructure classification models [33, 113, 22, 109]. A related protocol for localization has also been applied to variety of material systems [26, 112, 27, 25, 24, 23].

The extension of the MKS framework from structure-property to process-structure linkages introduces the need to represent the evolution of the microstructure as time series data. Time series modeling approaches can be separated into methods that work in the frequency domain such as spectral analysis [128, 129, 130] and wavelets [131, 132, 133], and methods that work in the time domain. The time domain methods are

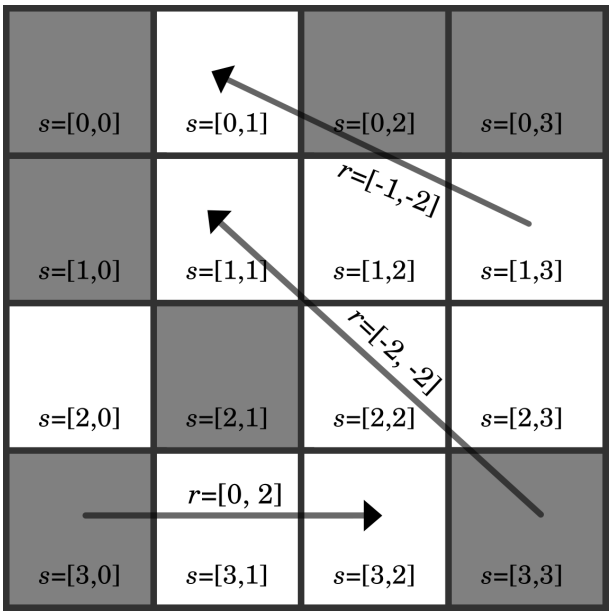


Figure 25: Idealized microstructure with two local states shown by the white and gray voxels. The relative spatial locations for two voxels is described by a discretized vector r.

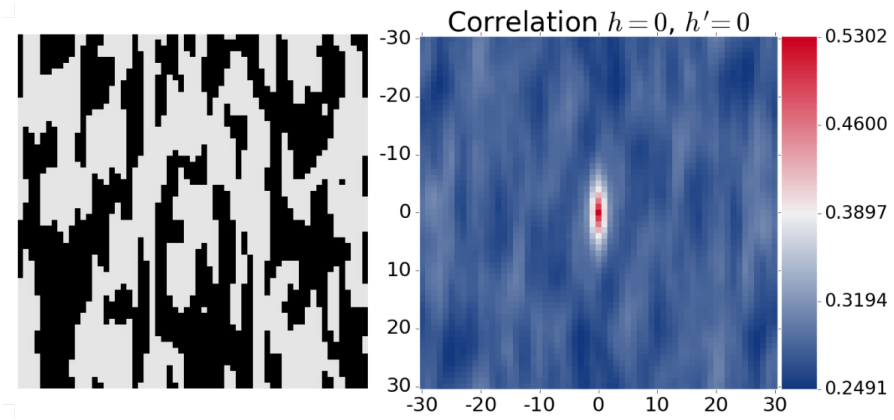


Figure 26: Discretized microstructure (left) and its 2-point statistics (right). Each pixel in the 2-point statistics image depicts the probability of finding the selected (ordered) local states at the ends of a vector whose tail is at the origin and the head is at the pixel itself.

to be further separated into three main categories: Autoregresive Integrated Moving Average (ARIMA) models, State Space models, and Neural Networks.

ARIMA models were developed by Box and Jenkins [134] and predict the evolution of time series data based on previous values and previous errors. The advantages of ARIMA models are that (i) the number of model parameters is linked to the number of previous values and previous errors as well as the number of times the time series is differenced, (ii) they can be solved with Ordinary Least Squares Regression, and (iii) they are intuitive. The model requires that nonlinear trends be removed from the data, and the residuals are normally distributed [134].

State space models estimate a joint probability over latent state variables and observed measurements. Kalman Filter [135] is used when the latent state variable is assumed to be continuous, while Hidden Markov Models are used with discrete latent variables [136, 137, 138]. State space models can be viewed as recursive Bayesian estimation [139], and are well suited for streaming noise data. The models create a linear function using a Markov assumption (only the current state is needed to predict the next), although extensions of the models have been made for nonlinearities [140, 141, 142]. The draw backs from this method are that (i) the parameter estimation is non-convex, and (ii) the dynamics of the system must be well understood a priori to create transition models to update the latent state variables.

Neural Networks have also been successfully applied to time series analysis as well as other sequential learning problems. The most notable model is Long Short Term Memory Neural Network (LSTM) [143]. LSTM introduces the concept of a memory block which contains gates that control the flow of information into and exiting the memory block as well as information carried into the next time step [144]. This model has been shown to outperform the previous two methods with non-stationary data [145], but optimization of neural network parameters is non-convex and typically requires a significant amount of data [144].

In this study, an extension to the MKS homogenization framework is presented using a non-parametric extension to ARIMA models using time series data gathered from synchrotron based *in-situ* X-ray scattering measurements. In experiments of this type, the material or a material system under investigation is exposed to an X-ray beam which allows the internal structure to be probed. Simultaneously, the corresponding changes in the properties of the material or material system are investigated by perturbing the varibles of state. These include variables such as temperature, pressure, and electric fields or their combinations. By combining the measurements of structure and function along with a suitable design of experiments, a detailed investigation into the structure-properties-processing-performance can be performed. Although our work towards establishing PSP linkages is demostrated on X-ray scattering data, the approach can be extended to data from *in-situ* experiments utilising a varitety of microscopy, tomography, neutron scattering, and spectroscopic techniques.

4.3 Extension of MKS to Process-Structure Linkages

State space models enjoy certain advantages in handling noisy data, and can be adapted to in-line learning from streaming data. But in order to avoid divergence and minimize error, these models require a priori knowledge of the dynamics of the system to create state transition matrices. Although some work has been done to empirically calibrate these transition matrices [146], the dynamics of low dimensional microstructure descriptors is generally not well understood. While LSTM have shown significant predictive power when optimized with large dataset, in many practical applications, materials datasets are not large [147]. For these reasons, LSTM and State Space models are not generally applicable for the evolution of low dimensional microstructure descriptors.

ARIMA models require that the model errors are known. This makes multistep

predictions impossible with a moving average component leaving only autoregressive model as a possible method. A non-parametric regression method call Multivariate Adaptive Regression Splines (MARS) was developed by Friedmen [63], and later applied to time series analysis by Lewis and others [148, 149]. In time series applications the method is referred to as Time Series Multivariate Adaptive Regression Splines (TSMARS). TSMARS has been shown to work well with moderate number of dimension (in our case PCs ≤ 20) and moderate sized data (between 50 and 1000)[63].

TSMARS looks to fit a function, $\mathcal{F}$, to connect the current microstructure descriptors $\mu[k,n]$ with it previous values as well as potentially with processing parameters $\eta[n]$ as shown in Eq. (50).

$$\hat{\mu}[k,n] = \mathcal{F}(\mu[k,n-1], \mu[k,n-2], ...\eta[n], \eta[n-1], ...) + \varepsilon \tag{50}$$

In Eq. (50), $\mu[k,n-i]$ is the microstructure descriptor for PC k at a discrete time value $n-i$ and $\eta[n-i]$ are processing parameters at time $n-i$. The function $\mathcal{F}$ is estimated in a piece-wise manner in disjoint subregions defined by hinge functions g using ordinary least squares regression. The borders of the disjoint regions are data points called knots, κ. The hinge function is defined in Eq. (51), and an example of two hinge functions meeting at a knot equal to 5 can be found in Fig. 27.

$$g(x) = \begin{cases} x, & \text{if } x > 0. \\ 0, & \text{otherwise.} \end{cases} \tag{51}$$

For economy of notation, the estimate of the function $\mathcal{F}$ will only be written in terms of the previous values of the microstructure descriptors $\mu[k,n-i]$, but the processing parameters can be added in the same way.

The TSMARS calibration consists of three major steps.

1. Forward pass which greedily adds hinge functions in pairs to minimize the local mean squared error (MSE) for subregions until some stopping criteria is reached.

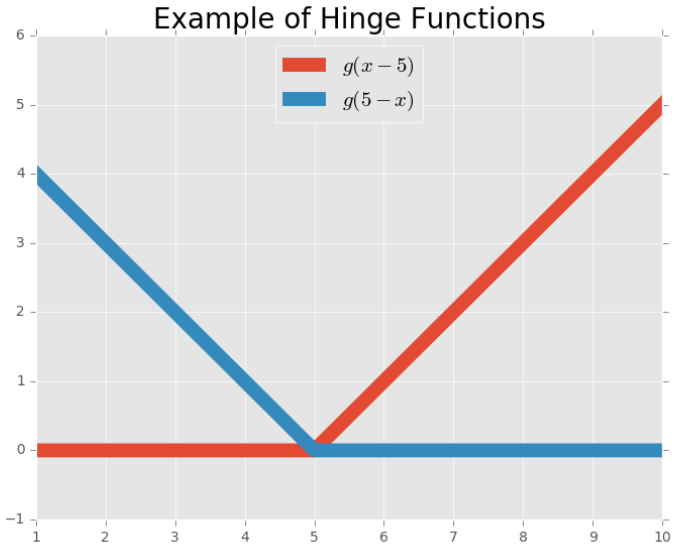


Figure 27: An example of two hinge functions meeting at a knot with value of five ($\kappa = 5$).

2. Pruning or backward pass to remove hinge functions to avoid over-fitting using Generalized Cross Validation (GCV) and MSE.

3. Coefficients for the bases function in each disjoint sub regions are calibrated using ordinary least squares regression.

The function $\mathcal{F}$ is estimated using linear combinations of the product of hinge functions as shown in Eq. (52).

$$\mathcal{F} = a[0] + \sum_{i=1}^{P} a[n-i] \prod_{q=1}^{Q_i} b[q,i] g(\mu[k', n-i] - \kappa[q,i])] \tag{52}$$

In Eq. (52), P is the number of previous time steps of values of microstructure descriptors that will be used in the model and is called the autoregressive (AR) order. Q_i is the set of disjoint subregions for time $n-i$, $\mu[k', n-i]$ the low dimensional microstructure descriptor for PC k' at time $n-i$, $\kappa[q,i]$ is a knot in PC k', $b[q,i]$ is a constant that can only take the values of ± 1, $a[n-i]$ and $a[0]$ are coefficients found using ordinary least squares regression. More detailed explanations of MARS and TSMAR can be found in published literature [63, 148, 149].

In this study the iterations of the Forward pass step of the calibration was stopped if i). the R-squared value is greater that .999 or ii). the change in the R-squared value was less than 0.001. The number of PCs K and the autoregressive order P were optimized for the Process-Structure linkage using a leave one sample out cross-validation approach. When selecting the range over which to search, three practical issues need to be considered. i). The AR order determines the number of initial images that need to be provided to the model at the time of prediction. As a result a low AR order that provides an accurate model is desired. ii). As the number of PCs increases, the amount of variance retained in the low dimensional microstructure descriptors $\mu_j[k,n]$ also increases. iii). While the time complexity for prediction is $O(NPK)$ where N is the number of time steps, the time complexity for calibration is $O\big(N^2(PK)^3\big)$. The last 2 issues require an optmized value of PCs where a sufficient amount of the variance is captured without the run time becoming too large.

During the cross-validation process, each of the samples was systematically left out during the calibration of the model. After each of the calibrations, the first P images from the sample not used during the calibration are provide as initial conditions for the model. The prediction process recursively used predictions from previous time steps in the model (i.e. if $P = 1$ then $\hat{\mu}_j[k,n+i]$ is used to predict $\hat{\mu}_j[k,n+i+1]$) to move forward in time. The MSE values, as defined in Eq. (53), were computed over the predicted time steps. This metric will be referred to as predicted MSE value through the remainder of the paper. During the calibration, the union of the calibration samples was used as a single time series which resulted in discontinuities at the edges between samples. To basis introduced by the discontinuities, the last value 50 steps in the sample used at padding at the end of each sample prior to forming a single calibration time series.

$$\frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}(\hat{\mu}_j[k,n]-\mu_j[k,n])^2 \tag{53}$$

In this work, the TSMARS model has been used to extend the MKS homogenization framework to Process-Structure Linkages. TSMARS uses a data driven approach to predict the evolution of the low dimensional microstructure descriptors by creating nonlinear functions of previous microstructure desciptor values in subregions in the PC space. In this paper, we explore the first application of this approach using X-ray scattering data.

4.4 Application

For this case study, X-ray scattering data is used as a surrogate for 2-point statistics. Indeed, it has been shown that the X-ray scattering data corresponds to the Fourier transform of the autocorrelation of the difference in electron density [64].

4.4.1 Dataset

In an X-ray scattering measurement, the specimen under investigation is irradiated by a coherent beam of monochromatic X-rays. The incident X-rays interact non-destructively with the electrons in the specimen. This interaction results in a fraction of the X-rays deviating from their original collimated path, i.e., results in scattering. The scattering of X-rays due to the distribution of electron density within the microstructure of a material is characteristic of the material. The scattered X-rays, which are captured on a two dimensional detector plate create a scattering pattern which contains the microstructural information characteristic to the specimen that caused the scattering. The scattering pattern can be mathematically related to the electron density distribution within the specimen by way of the Fourier Transform. Consequently, the inverse relationship between length scales in real and Fourier space is maintained analogous to time and frequency in a traditional temporal Fourier transform.

Small Angle X-ray Scattering (SAXS) is a subset of scattering techniques wherein

microstructural features at the mesoscopic scale between one to hundreds of nanometers can be probed in a specimen. In this work, we use SAXS to investigate the mesoscopic structure of semi-crystalline polymer films of linear low density polyethylene (LLDPE), a grade of PE. Semi-crystalline polymers comprise of a microstructure wherein the polymer chains can organize into crystalline and noncrystalline domains. The crystalline domains consist of tightly packed polymer chains that have become regularly ordered to form lamellae while amorphous domains are formed from loose disordered arrangements of the polymer chains. Consequently, the electron density in crystalline domains ρ_c is greater than ρ_a, the electron density in the amorphous domains.

The three dimensional Fourier transform of the electron density distribution within the scattering volume is captured by the X-Ray detector plate in the form of a two dimensional projection i.e. a SAXS pattern. In the specific case of a semi-crystalline microstructure the electron density distribution varies spatially between ρ_c and ρ_a.

A single SAXS pattern therefore provides an average description of all the spatial arrangements of crystalline and amorphous domains within the scattering volume at that instant of time. The time series data required for the application of the MKS homogenization approach are image sequences of such SAXS patterns obtained while simultaneously recording the stress and strain data of the individual specimens during uniaxial tensile stretching. The dataset therefore provides insight into the evolution of the semi-crystalline microstructure at the mesoscale for different LLDPEs under uniaxially applied stress and strain.

4.4.2 Materials and Methods

Melt blown films of two LLDPE polymers, henceforth referred to as LLDPE1 and LLDPE2 were supplied by ExxonMobil Chemical Company (EMCC). Both the LLDPEs were statistical copolymers of ethylene and hexene, where the hexene comonomer

Polymer	Film Label	BUR	Thickness (μm)	Images
6*LLDPE1	LLDPE1.1a	2.5	20	182
	LLDPE1.1b	2.5	30	191
	LLDPE1.1c	2.5	75	195
	LLDPE1.2a	3	20	191
	LLDPE1.2b	3	30	195
	LLDPE1.2c	3	75	191
6*LLDPE2	LLDPE2.1a	2.5	20	191
	LLDPE2.1b	2.5	30	195
	LLDPE2.1c	2.5	75	61
	LLDPE2.2a	3	20	227
	LLDPE2.2b	3	30	199
	LLDPE2.2c	3	75	206

Table 2: Labeling of tensile specimens made from blown films of the two LLDPE polymers.

was incorporated along the backbone chain in the form of butyl short chain branching (SCB). The densities for LLDPE1 and LLDPE2 were 0.912 g/cc and 0.923 g/cc. The density variation between the two polymers arose from the differing levels of SCB incorporation. The melt flow indices for each of the polymers was 1.0 (ASTM-D1238) and the molecular weight distributions for these LLDPEs were also similar. Both polymers were converted into films by the method of film blowing into two series of blown films. The first series of films had a blow up ratio (BUR) of 2.5 while the second series had a BUR of 3. The BUR is a standard processing parameter which describes the manufacture of blown films. Within each series, three films were fabricated with average thicknesses of 20μm, 30μm and 75μm, thereby totaling twelve films. The labeling scheme followed in the current work to describe tensile specimens for *in-situ* testing is described in Table 2

SAXS experiments were performed at beamline 12-IDC of the Advanced Photon Source (APS) at the Argonne National Laboratory (ANL). In these experiments, the X-ray beam had an energy of 12 keV (i.e., a wavelength of 1.0332Å) and the beam dimensions were 200 μm x 200 μm. X-rays scattered by the LLDPE film specimens were detected by a MAR CCD detector situated at a distance of 2426 mm from the

LLDPE specimen. The detector pixel size was 175 μm. A fixed exposure time of 0.1 seconds was utilised while taking SAXS snapshots. SAXS patterns were collected every 3 seconds. This time interval was determined based on the minimum detector readout time per pattern. A portable tensile stage made by Linkam Scientific Instruments, was utilised for the tensile measurements. The Linkam stage was operated at a tensile deformation rate of 25.4 mm per minute. The collection of SAXS data and deformation data was synchronized such that the first SAXS pattern in any of the image sequences was always obtained from an unstrained pristine specimen at t_0.

4.4.3 Data Processing

Prior to analysis, the contrast X-ray scattering images was enhance by taking the log of the intensity. In order to normalize the difference in intensity due to film thickness, each of the images were normalized by their mean intensity value.

PCA was done on all 2224 images from the 12 samples. The PC scores as well as the percent variance as a function of the number of PCs is shown in Fig. 29. In the image on the top in Fig. 29, the microstructure evolution of the 12 samples is represented by the process streamlines ($\mu_j[k, n]$ as a function of time), each indicated by different colors in the image on the bottom.

4.4.4 Model Selection, Calibration and Validation

The results from the leave one sample out cross-validation process can be found on the top image in Fig. 30. It was found that as the number of PCs increased, the mean predicted MSE value decreased. When K is small the mean predicted MSE value decreases as P increases, but this tread reverses as the K gets large. In general the value of P had less affect on the predicted MSE value as the number of PCs increased. The model with the lowest mean predicted MSE value was found to have $P = 1$ and $K = 20$ and had a value of 6.8. The general trend indicates that the model accuracy would continue as the number of PCs increases but with diminishing

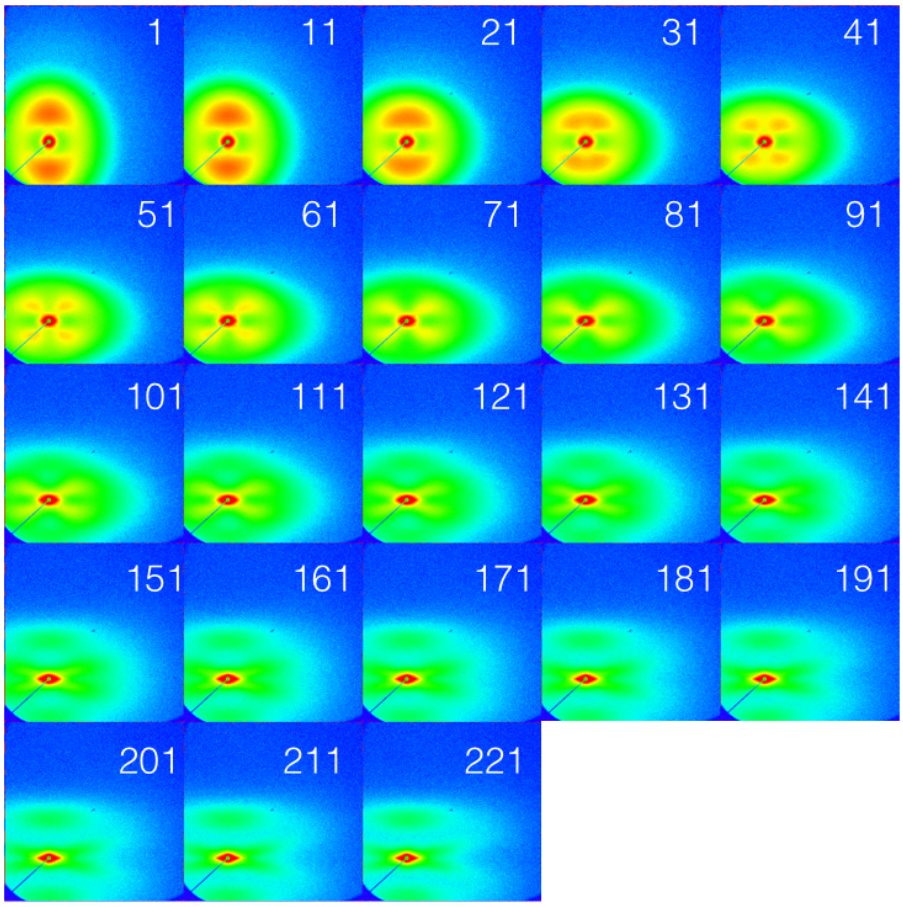


Figure 28: An example of the typical evolution in 2D SAXS patterns with increasing strain for a tensile specimen of LLDPE2.2a. Strain is applied in the vertical direction. The numbers indicate the SAXS pattern number; an interval of 3 seconds between consecutive patterns is strictly maintained. Every tenth image is displayed for clarity. The intensity is log scaled to highlight the characteristic features in the SAXS evolution with strain

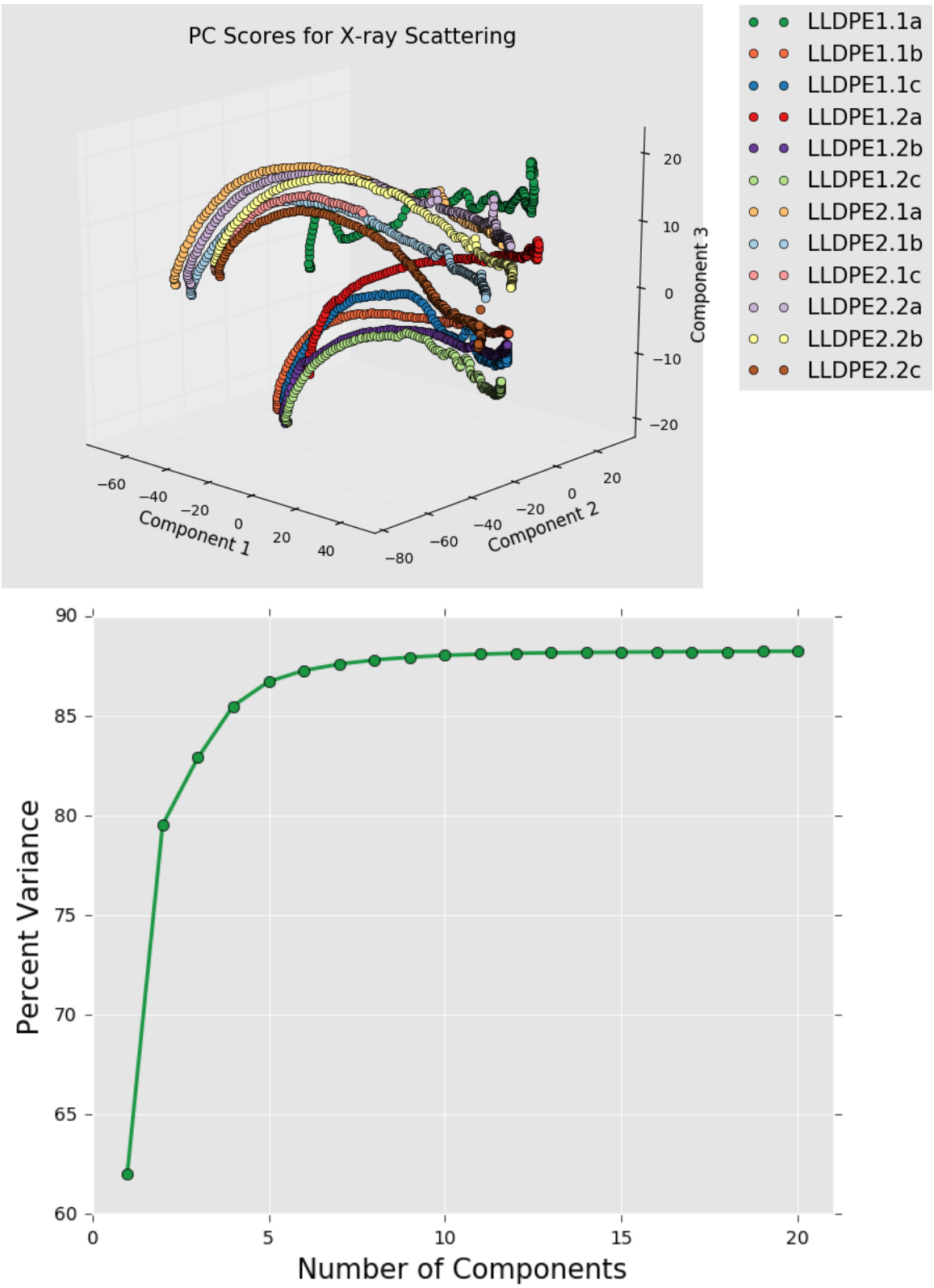


Figure 29: Principal component scores for X-ray scattering images from the 12 different samples and there percentage of the variance captured as a function of the number principal components.

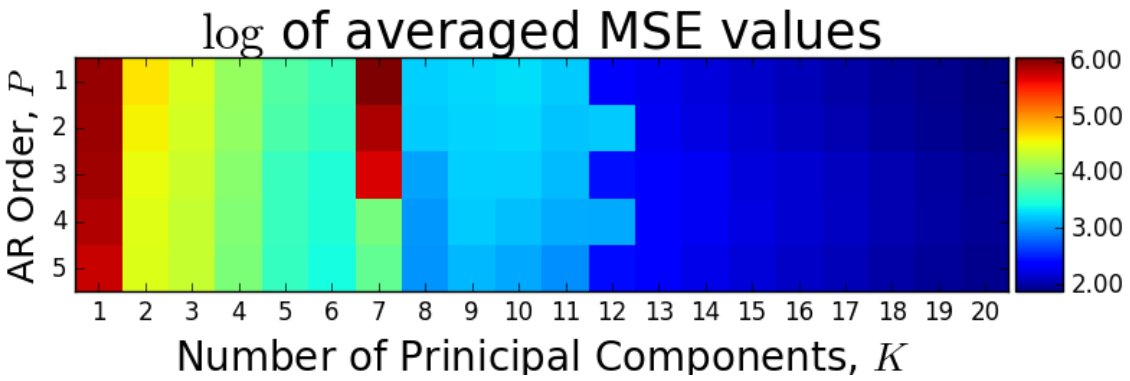
returns and higher computational costs. The bottom images in Fig. 30 qualitatively show that the predictions for PCs 1 and 2 improve as the total number of PCs K increases.

In order to demonstrate the utility of this method, the models with the maximum and minimum predicted MSE values found during the cross-validation process for $P = 1$ and $K = 2$ are used for prediction. Using the predicted low dimensional microstructure descriptors, a low rank approximate of the X-ray scattering images was create (as shown in Eq. (49)). The model used to predict sample LLDPE1.2b had the lowest predicted MSE value of 2.11. Fig. 31 shows the actual final image and the final predicted low rank image. The model used to predict sample LLDPE2.2b had a predicted MSE value of 19.9. The similar results can be found in Fig 32. The predictions from both models retain the relevant scattering information.

4.5 Conclusion

In this paper, an extension of the MKS framework to process-structure homogenization linkages using Time Series Multivariate Adaptive Regression Splines was presented. The process-structure linkage was generated using a data driven approach to calibrate non-linear relationships between low dimensional microstructure in PCA space containing non-linear functions, and the approach was validated by predicting a low rank representation of the final step of Small Angle X-ray Diffraction data. This extensions allows for the complete process-structure-property homogenization linkages to be created which can be used to solve inverse materials design problems and expedite the development of new materials.

0.95
(a)
b



0.95
(b)
b

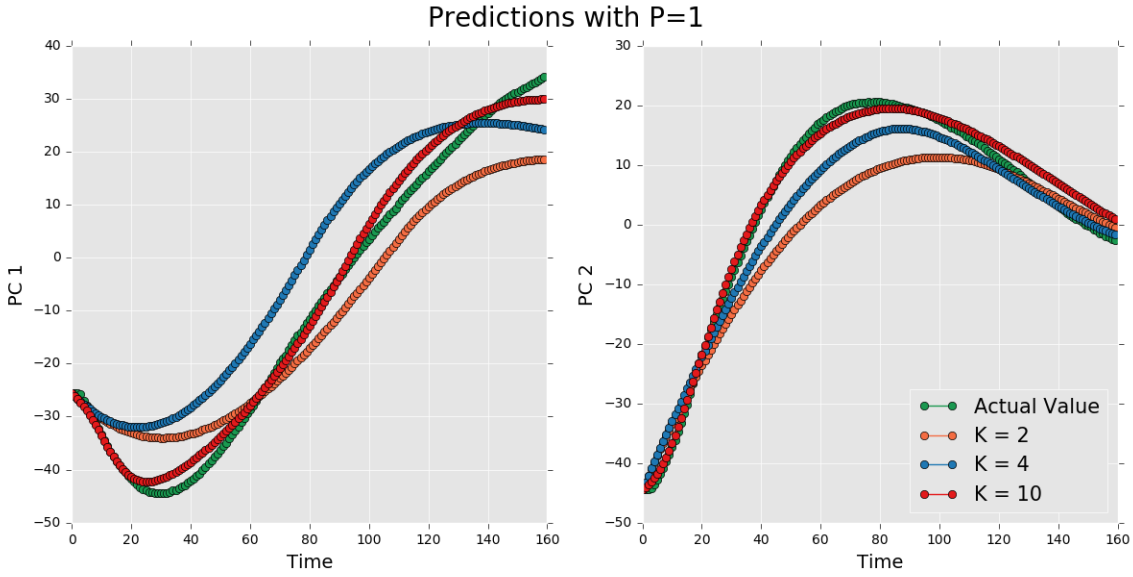
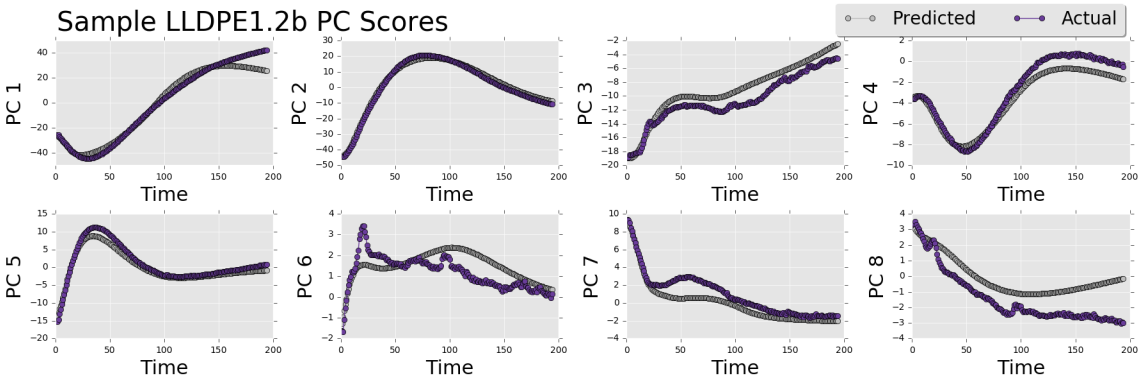


Figure 30: Log of the mean predicted MSE values created using leave one sample out cross-validation as a function of the autoregressive order P and total number of Principal components K (above). Mean predicted MSE values for PCs 1 and 3 decrease as the number of the total number of principal components K increases (below).

0.95

(a)

b



0.95

(b)

b

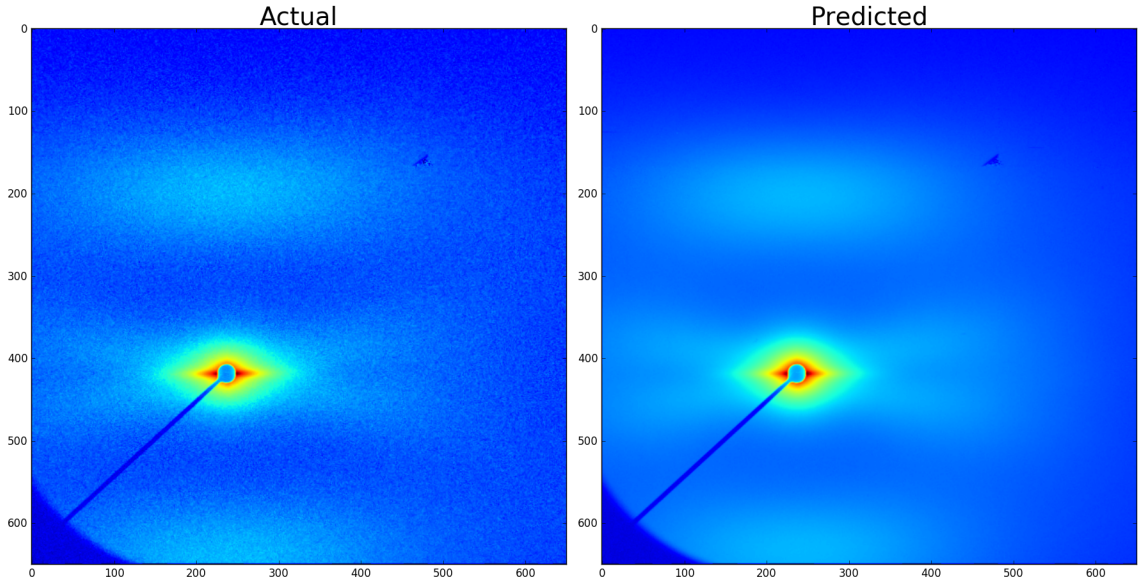
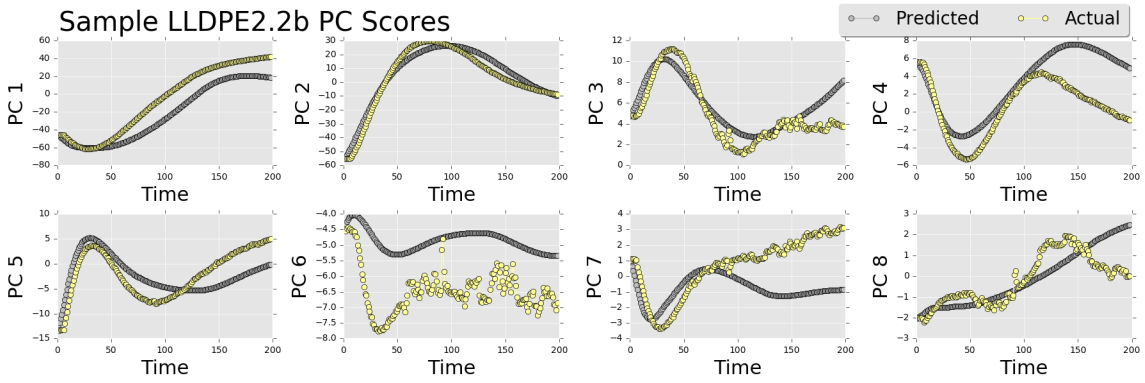


Figure 31: Predicted and actual principal component scores for sample LLDPE1.2b (above). The experimental image (bottom left) and the predicted image (bottom right). The mean squared error value over the predicted principal component scores had a value 2.11.

0.95

(a)

b



0.95

(b)

b

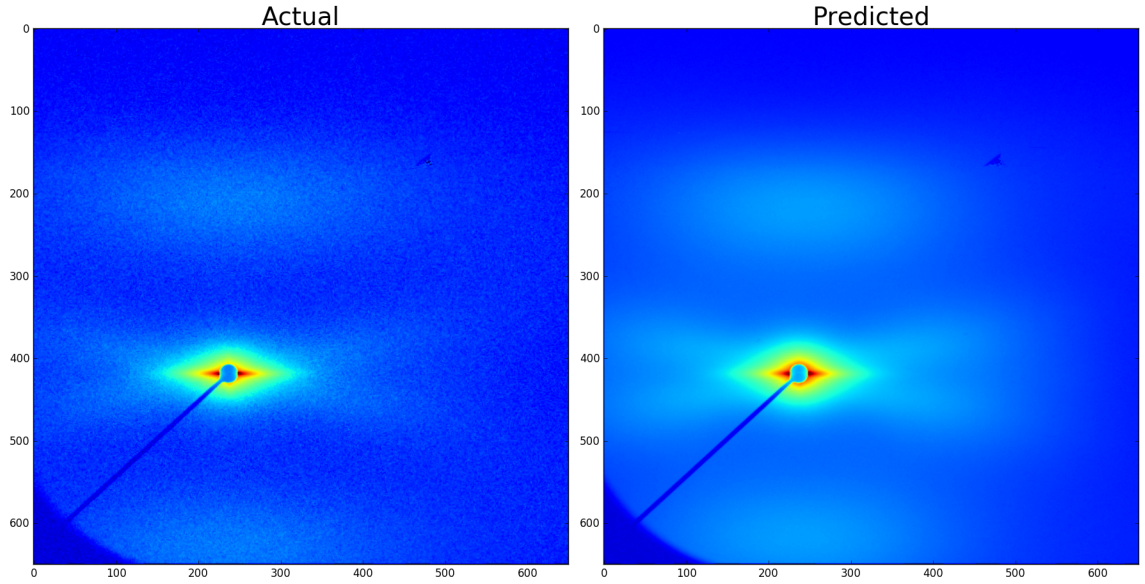


Figure 32: Predicted and actual principal component scores for sample LLDPE2.2b (above). The experimental image (bottom left) and the predicted image (bottom right). The mean squared error value over the predicted principal component scores had a value 19.9.

CHAPTER V

MICROSTRUCTURE-BASED KNOWLEDGE SYSTEMS FOR CAPTURING PROCESS-STRUCTURE EVOLUTION LINKAGES

5.1 Introduction

Customized materials design (including the design of a manufacturing process route) resulting in the combination of properties desired for a specific application is a highly challenging inverse problem, owing mainly to the extremely large parameter space involved in defining the hierarchical internal structure of the material. However, this endeavor has great potential for impacting virtually all emerging technologies [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 150], with significant economic consequences. The central impediment comes from the need to consider the relevant details of the hierarchical internal structure (spanning a multitude of length scales) that control the properties of interest to a specific application. Additionally, a diverse range of coupled physical phenomena occur at different timescales at each of the different length scales. Therefore, one is generally daunted by the enormous difficulty involved in tailoring the material structure to yield desired combinations of properties or performance characteristics.

Historically, and mainly because of the difficulties mentioned above, materials development efforts have relied largely on experimentation. Consequently, many of the efforts aimed at designing and developing new/improved materials have incurred significant cost and time. Recent advances in physics-based modeling of multiscale materials phenomena [151, 152, 153, 154, 155, 156, 120, 157, 23, 158] have raised the exciting possibility that the vast design space for experimentation can be constrained

to a significant degree by embracing in-silico simulations and explorations. In other words, there is a tremendous potential for significant reductions in cost and time incurred in materials development effort if one could judiciously utilize multiscale materials modeling and simulation tools in combination with a reduced number of experiments.

The central impediments associated with the effective utilization of physics-based multiscale materials models in the materials development include: path dependent microstructure evolutions that depend on initial conditions, non-unique parameter selection for coupling multiscale models, approximations in microstructure representation, material property dependence on extreme values of microstructure distributions, large optimization space, metastablity of microstructure during use, and uncertainty in data, models and model parameters [120, 69]. An important strategy in addressing these impediments involves the formulation and utilization of robust surrogate models (also called metamodels or emulators) for computationally efficient communication of critical information between well separated structure/length/time scales. Such low-dimensional, but sufficiently accurate, models present a computationally viable approach for exploring efficiently the extremely large materials design space.

In the context of hierarchical materials (with details of the material structure spanning multiple well-separated scales) surrogate models are needed to exchange high value information in both directions between the scales. Depending on the direction of information flow, the models can be classified as homogenization (information flowing from lower scales to higher scales) or localization (information flowing from higher scales to lower scales) relationships. It should be noted that localization linkages are significantly more difficult to establish compared to the homogenization linkages; indeed the latter are implicitly embedded within the former and can be recovered from them when needed.

5.2 Review of Homogenization and Localization Approaches

Theories for predicting the properties of composite materials go as far back as 1873, with Maxwell predicting an effective conductivity for a region of a materials with dilute inhomogeneities through a mean-field approximation [159, 160, 161]. The simplest, and most commonly used, homogenization methods for mechanical properties were developed by Voigt and Reuss [162, 163], and provide "elementary" bounds for the estimates of the effective properties. These calculations typically involve simple volume-averaging of the properties at the microscale. The bounds obtained in this approach also correspond to the correct effective values for highly specialized microstructures. For example, the upper bounds obtained in these approaches typically correspond to microstructures where the microscale constituents have uniform shape, and are continuous and perfectly aligned along a loading direction (e.g., unidirectional, straight, and continuous fibers).

Hill and Hashin introduced the concept of a Representative Volume Element (RVE) [94, 95] which can be defined as a statistically homogeneous subvolume where the length scale associated with the local perturbation in material properties is sufficiently small compared to the length scale of the subvolume (typically referred as "well-separated" length scales). With this definition, a mean-field approximation can be used to assign an effective property to a RVE. It can be shown that good estimates for a broad class of effective properties associated with an RVE can be expressed in the following generalized form [94]:

$$P_{eff} = \left\langle A(x)P(x) \right\rangle \tag{54}$$

where P_{eff} denotes the effective property, $\left\langle \right\rangle$ denotes ensemble average (also equal to volume average by virtue of the ergodic assumption), and $A(x)$ is a suitably defined tensor operator that depends on the spatial location x in the RVE. The central challenge of this theory lies in the computation of the spatially-varying tensor operator.

While the theory described above emerged in the context of mechanical properties, it has also been successfully applied to material properties such as thermoelectric, piezoelectric, diffusion, and conductivity for composite materials [20]. A large variety of approaches have been built on this foundational framework, and have been employed successfully in addressing practical problems of interest in composite material systems. Hill developed the self-consistent method which employs Eshelby's solution to ellipsoidal inclusions in an infinite medium to find an approximate estimate of the effective properties [164, 165]. An improved generalized self-consistent method emerged from the work of Hashin, Shtrikman, Christensen and Lo [95, 166, 167, 168, 169], which allows for more complex geometric shapes of the reinforcement phase. A good overall treatment of such approaches for homogenization theory or estimates can be found in the textbook by Qu and Cherkaoui [170] as well as the report by Bohm [160]. Further advanced theories of homogenization were established by Willis [171], and subsequently by Ponte-Castanada [172].

In a completely different approach, advanced composite theories were developed to specifically take into account the rich details of the material microstructure. These approaches utilized the formalism of n-point spatial correlations to quantify the details of the material microstructure together with the concept of Green's function to estimate the effective property of interest [99, 100, 101, 102, 103, 104, 105]. An overview of this more sophisticated approach for composite theories can be found in the book by Milton [20]. One of the earliest demonstration of this approach comes from Brown, who used a series expansions of a localization tensor to predict the electrical conductivity of a 2-phase material [96]. More rigorous applications of this approach can be found in the work of Torquato and co-workers [173, 174].

The main limitation of the approach described above is that the Green's functions needed to implement the method are only available for cases involving highly

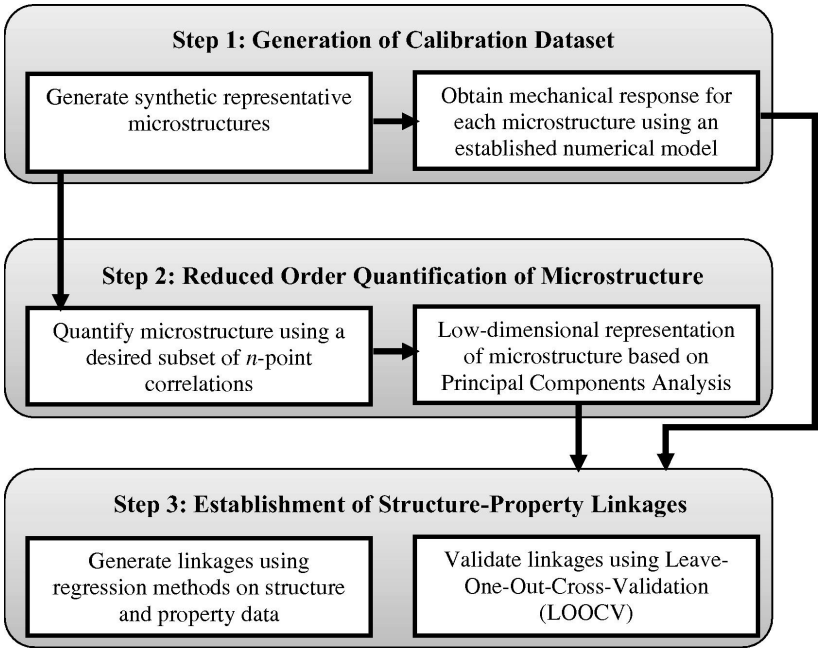


Figure 33: Generalized MKS homogenization workflow for structure-property linkages [2].

idealized and simplified physics (i.e., material constitutive laws). The recently formulated Materials Knowledge Systems framework (MKS) addresses this critical gap by advancing a data-driven approach [23, 24, 25, 27, 3].

5.3 Homogenization and Localization with MKS

MKS homogenization and localization linkages are created by merging concepts from the physics-based statistical continuum theories developed by Kroner [97, 98], machine learning [123, 124, 125] and digital signal processing [122]. A generalized workflow for establishing the homogenization linkage (e.g., structure-property linkage) is shown in Fig. 33. Broadly, this worflow includes a calibration step and a validation step. More specifically, this data-driven approach captures the pertinent microstructure features through n-point spatial correlations and employs dimensionality reduction techniques to create low-dimensional microstructure descriptors [109, 110, 113]. Linkages between effective properties and these low dimensional descriptors are then created using regression techniques [22, 2, 33].

The MKS localization linkages are expressed as a series, where each term involves convolutions of physics-capturing kernels (based on Green's functions) with hierarchical microstructure descriptors. These kernels (referred to as influence functions) capture and organize the governing physics as convolution operators that are independent of the spatial arrangements of the local states in the material microstructure. Therefore, in the MKS localization approach, these kernels are calibrated with results produced using numerical tools (e.g., finite element models).

It is emphasized here that once the influence kernels in the MKS linkages are calibrated and validated, they can be used to predict the local responses for new microstructures at very minimal computational expense. Therefore, this approach is of particular value when one needs to explore a very large number of potential microstructures. It should be noted that the design of multiscale material systems for optimized performance is expected to require a consideration of a very large number of potential microstructures. As a simple illustration, let us assume that the specification of an RVE at any selected material structure scale would require a minimum of 8000 (i.e. 20 X 20 X 20) spatial cells or voxels. Further assume that the specific material system being explored allows for placement of only ten distinct potential local states (could be based on differences in thermodynamic phases and/or chemical composition and/or defect densities and/or local orientation attributes such as the crystal lattice orientation). Even with such highly conservative estimates, the number of different RVEs that one can imagine producing in a comprehensive materials design exploration is 10^{8000}. While an exhaust search of the structure space is impractical, the subspaces explored during optimization are also extremely large. An efficient exploration of such large design spaces demands innovative new approaches.

Prior effort in MKS localization was largely focused on steady state structure-property localization linkages [25, 24, 27, 23], with the exception of one prior study exploring the time evolution of the microstructure field [26]. Previous studies have

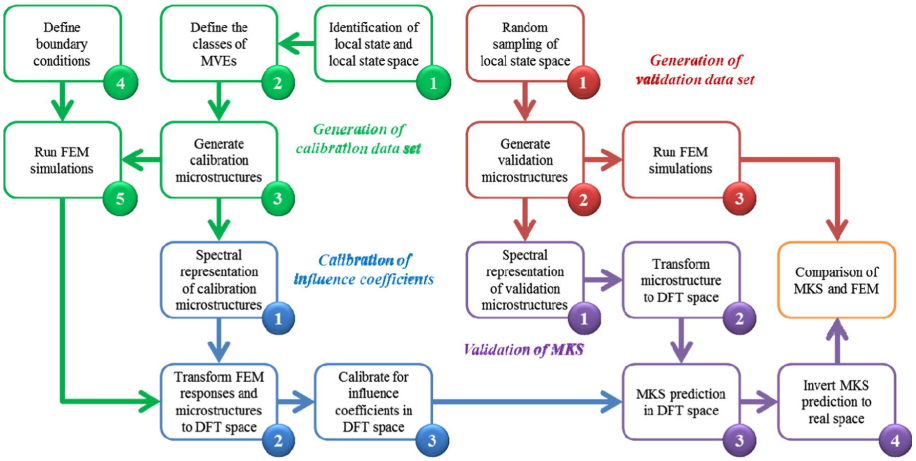


Figure 34: Generalized MKS Localization workflow for structure-property linkage at the meso-scale [3].

shown that if the local states are discrete, influence coefficients can be calibrated with a small number of simulations and then used to predict the local response of microstructures with any new spatial configuration of the local states [25, 24]. Similar to Green's functions, the MKS localization kernels depend on the boundary conditions and physical constants that govern the constitutive behavior of the local states present in the material system. It has been shown that the influence kernels can be suitably parametrized to include such dependencies (i.e., interpolation between sets of kernels can provide remarkably accurate predictions for new conditions [3]). A generalized MKS localization workflow, including both calibration and validation steps, can be found in Fig. 34. It is noted here that the workflows presented in Fig. 33 and 34 are highly generalized, and can be applied broadly to a range of material systems experiencing a range of multiscale materials phenomenon.

A schematic illustration of a multiscale simulation using the MKS framework is shown in Fig. 35. This chain of models passes homogenization information from the lower length scales to higher lengths. The thermodynamic model computes thermodynamic quantities that define the phase field model parameters. In turn the phase field model predicts the microstructure as a result of processing conditions.

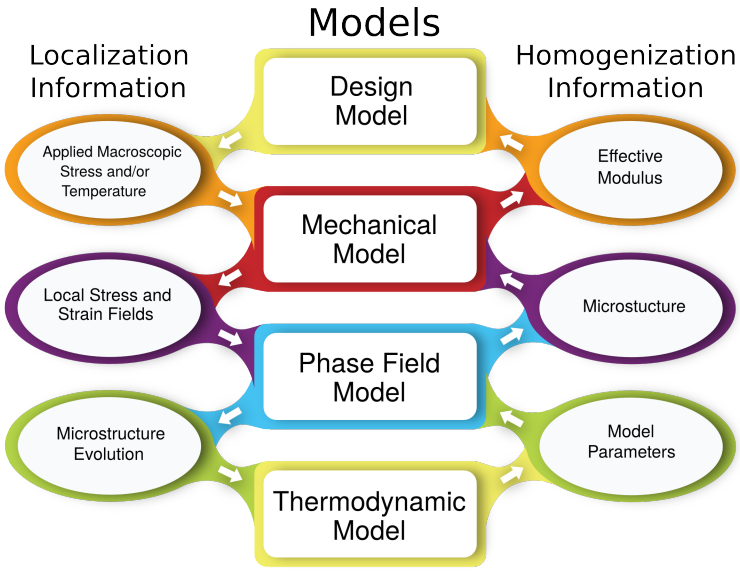


Figure 35: One instance of a multiscale simulation covering multiple length and time scales. The homogenization information is passed to models at higher length scales, while the localization information is passed to models at lower length scales. Sufficiently accurate and computationally cheap surrogate models created using the MKS framework serve as replacements for the models which use the microstructure (i.e., mechanical and phase field models) to speed up exploration for a new material.

The effective modulus of the microstructure is found using the mechanical model (e.g finite element method), which is used in the design model for a component. In the model chain localization information is also passed from the higher length scales to the lower length scales. Using a microstructure and the applied stress and/or temperature provided by the design model, the mechanical model computes local stress and strain fields. These local fields provide information about the free energy density to the phase field model for microstructure evolution which in turn invalidates thermodynamic calculations from the thermodynamic model. The current approaches for multiscale simulations based on numerical approaches such as finite element method or the phase-field method are not ideally suited for such bi-directional explorations due to their high computation costs. In the MKS framework, computationally cheap and sufficiently accurate surrogate models will serve as surrogates, and can greatly expedite this process.

Although most of the previous work in the MKS framework has focused on the mesoscale [25, 24, 27, 23, 22, 2, 33], the approach can indeed be extended to other length and time scales involved in multiscale materials phenomenon as long as the phenomena involved are well separated and the heterogeneity at each scale is statistically homogeneous or weakly stationary [21]. Although some preliminary work with MD simulations has been reported [121, 175], much additional work is still needed to further refine and demonstrate the details of such implementations.

The current effort is aimed at the extension and application of the MKS approach to include transient process-structure evolution localization linkages. In general, the process-structure evolution linkages are significantly harder to establish compared to the structure-property linkages, because of the need to explicitly account for the time evolution of the important field quantities (in addition to their spatial distributions), many of which demand continuous descriptions. Consequently, there currently exist only a few reports in literature describing efforts aimed at capturing the salient process-structure evolution metamodels for multiscale materials phenomena. One area that has received a lot of attention in prior literature is the evolution of crystallographic texture in deformation processing of polycrystalline metals [176, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61]. However, in this set of applications, all the attention is generally focused on capturing the salient details of the time evolution, while ignoring or grossly simplifying the spatial distribution of the important field variables involved. More recently, there have been a limited number of efforts aimed at mining low-dimensional process-structure evolution linkages from results accumulated in phase-field simulations [177, 178, 179, 158]. In these prior applications, gross simplifications were made by limiting the set of initial microstructures, the microstructure descriptors or the local states allowed in the microstructure.

In this paper, we extend the MKS localization framework to allow efficient capture of the process-structure evolution localization linkages. Indeed, this extension when

suitably combined with the existing MKS framework has the potential to facilitate a common, consistent, broadly applicable, framework for casting all of the relevant process-structure-property (PSP) linkages in a selected class of materials. A second major thrust of this paper is the derivation of the MKS framework using spectral representations for some of the main functions (kernels) involved in these linkages. The novel protocols described above are demonstrated in this paper through a specific case study involving the extraction of process-structure evolution linkages embedded in the simulation results produced by a selected phase-field model.

5.4 Generalized MKS Framework for Process-Structure Linkages

The development of the generalized MKS framework for process-structure linkages will be presented here using the Cahn-Hilliard model as an example. However, the generalized final expression formulated here is broadly applicable to various other microstructure evolution models. The Cahn-Hilliard description of microstructure evolution can be expressed as [180]

$$\frac{\partial \phi(x,t)}{\partial t} = D\nabla^2\Big(\big[\phi(x,t)(1-\phi(x,t)^2)\big] + \gamma\nabla^2\phi(x,t)\Big) \tag{55}$$

In Eq. (55), $\phi(x,t)$ is an order parameter used to represent the concentration field at location x and time t, $\sqrt{\gamma}$ represents the interface width, and D is the diffusivity. Note also that a double well potential with minima at -1 and 1 has been used in Eq. (55) for the free energy term [66, 67].

The theoretical framework of MKS is built on perturbation expansions. For the present case, we therefore start by expressing the concentration field $\phi(x,t)$ in terms of a reference quantity $\bar{\phi}$ (constant in both space and time) and a local perturbation $\phi'(x,t)$ from that reference as

$$\phi(x,t) = \bar{\phi} + \phi'(x,t) \tag{56}$$

Introducing Eq. (56) into Eq. (55) results in the following differential equation.

$$\begin{aligned}\frac{\partial \phi'}{\partial t} - D\nabla^2\phi' &= D\nabla^2\Big(\gamma\nabla^2\phi' - \left[3\phi'\bar{\phi}^2 + 3\phi'^2\bar{\phi} - \phi'^3\right]\Big) \\ &= D\nabla^2\Big(\gamma\nabla^2\phi' + \Psi(x,t)\Big)\end{aligned} \tag{57}$$

where

$$\psi(x,t) = \phi'^3 - 3\phi'\bar{\phi}^2 - 3\phi'^2\bar{\phi} \tag{58}$$

We can use a Green's function approach to find the solution to Eq. (57) where

$$\frac{\partial G(x-x',t-t')}{\partial t} - D\nabla_x^2 G(x-x',t-t') = \delta(x-x',t-t') \tag{59}$$

and with a suitable change of variables we have

$$\begin{aligned}\phi'(x,t) = &-\int_V G(r,t)\phi'(x-r,0)dr + \\ &\int_T\int_V G(r,\tau)D\nabla_r^2\Big(\gamma\nabla_r^2\phi'(x-r,t-\tau) + \psi(x-r,t-\tau)\Big)drd\tau\end{aligned} \tag{60}$$

In Eq. (60), $\phi'(x,0)$ is the initial value of the perturbed concentration. Assuming periodic boundary conditions, the operators ∇_r can be moved from concentration terms to the Green's functions.

$$\begin{aligned}\phi'(x,t) = &-\int_V G(r,t)\phi'(x-r,0)dr + \\ &\int_T\int_V D\gamma\nabla_r^4 G(r,\tau)\phi'(x-r,t-\tau) + D\nabla_r^2 G(r,\tau)\psi(x-r,t-\tau)drd\tau\end{aligned} \tag{61}$$

Recursive substitution of $\phi'(x,t)$ into equation (61) produces a series (called the weak contrast expansion) that can be used to compute the perturbed concentration field [98, 97, 174, 21].

$$\phi'(x,t) = -\int_V \Big[1 - \int_T\int_V \tilde{G}(r,r',t,\tau)dr'd\tau\Big]G(r,t)\phi'(x-r,0)dr + ... \tag{62}$$

with

$$\tilde{G}(r,r',t,\tau) = D\nabla_r^2\Big[\nabla_r^2 G(r',\tau)\gamma - 3\bar{\phi}^2 G(r',\tau)\Big] \tag{63}$$

The higher order terms in Eq. (62) (i.e., the terms denoted by ...) will be discussed later (see the description near Eq. (68)).

Simplified analytical solutions for Eqs. (62) and (63) are very difficult and demand highly sophisticated approaches to handle the convergences of the terms in the series [181, 102]. There have also been numerous approaches utilizing numerical iterative schemes to solve the same equations [182, 183, 184, 185, 186, 187]. The numerical approaches generally demand significant computational resources because of the highly nonlinear expressions embedded in Eqs. (62) and (63). More importantly, most conventional numerical approaches do not facilitate learning. In other words, when the equations are solved for one specific set of inputs there is no established formalism for transferring the knowledge gained in the process to the next application of the same set of equations for a different set of inputs. This is precisely where data science approaches, such as the MKS approach, bring many potential benefits. In the data science approach, we recognize that each term in the series is essentially a convolution, where the kernel is completely independent of the topological details of the material microstructure. Suitable algorithms are then designed and employed to efficiently learn these kernels from previously accumulated results. In many ways, the calibrated MKS localization linkages take full advantage of the known physics of the phenomena, and supplement only the mathematically intractable components with data science approaches, where they exhibit a decisive advantage.

The MKS kernels facilitate learning and transfer of knowledge to a new set of microstructure inputs. In order to accomplish this, Eqs. (62) and (63) need to be reformulated using the concepts of microstructure function and local states [107, 188]. The local state captures all the attributes (thermodynamic state variables) needed to identify the physical properties to be assigned to the spatiotemporal location of interest in the material internal structure. In the problem described here, either the concentration value or the perturbed concentration value (after selecting a reference concentration value) can serve as local state variables. The local state will be denoted as h. The set of all values that h can take is denoted as the local state space, H. The

main distinction between h and ϕ' (or ϕ) is that the later is a specific value assigned to a specific spatiotemporal location, while the former denotes any value that could have been assigned to the later.

The introduction of the concept of the local state now allows us to describe a microstructure function $m(h, x, t)$ as the probability density associated with finding the local state h at spatial position x at time t. The expectation value obtained using this probability density distribution (on the local state space) should be taken as the specific value assigned in the fully deterministic framework described earlier. In other words, for the fully deterministic case, one could write $m(h, x, t) = \delta(h - \phi'(x, t))$. For the more general case, the definitions introduced above lead to the following mathematical statements

$$\int_H m(h, x, t)dh = 1 \tag{64}$$

$$\phi'(x, t) = \int_H hm(h, x, t)dh \tag{65}$$

The introduction of the microstructure function as a probability density function brings the added benefit that it maps complex descriptions of local state (potentially could be a combination of several scalar and tensor thermodynamic state variables) into a continuous scalar-valued function that lends itself naturally to spectral representations [16, 17, 18, 19]. Extending the treatment above to the term containing the Green's function in Eq. (62) allows us to define an influence function or localization kernel as

$$\alpha(h, r, t) = -h\Big[1 - \int_T \int_V \tilde{G}(r, r', t, \tau)dr'd\tau\Big]G(r, t) \tag{66}$$

Recasting Eq. (62) with the terms defined in Eq. (65) and (66) takes the following form.

$$\phi'(x, t) = \int_V \int_H \alpha(h, r, t)m(h, x - r, 0)dhdr + ... \tag{67}$$

The derivation of Eq. (67) is a key step in the formulation of the MKS approach for the process-structure evolution localization linkages sought in this work. The

main benefit of the form of Eq. (67) lies in the fact that $\alpha(h, r, t)$ serves as a convolution kernel capturing all of the relevant physics in the problem, and operates on the initial microstructure function $m(h, x, 0)$. Even more importantly, when the series is expanded properly, the localization kernel is completely independent of the microstructure function $m(h, x, 0)$. Furthermore, Eq. (67) is the exact analog of the structure-property localization linkages established previously in the MKS framework [23, 24, 25, 27, 3]. Therefore, the extension presented here now makes it possible to explore the complete set of process-structure-property linkages in a consistent MKS framework in both space and time.

Because of the specific way in which Eq. (67) was derived, it is relatively easy to write the additional terms in the series expansion. For example, the second term in this expansion would be expressed as

$$\int_H \int_H \int_V \int_V \int_T \alpha(h, h', r, r', t, t') m(h, x - r, 0) m(h', x - r', 0) dh dh' dr dr' dt' \tag{68}$$

Therefore, another way to interpret the series expansion in Eq. (67) is to recognize that each term in the series captures the contribution arising from a specific arrangement of the local microstructure in the neighborhood of the spatial voxel of interest as a function of time.

The next step in the practical implementation of the MKS framework is to transform Eq. (67) into a discrete representation. In other words, the functions $\phi'(x, t)$, $\alpha(h, x, t)$ and $m(h, x, t)$ need to be discretized. Following notations and conventions employed in signal processing [189, 190], we will use round brackets to represent variables with continuous domains and square brackets to represent variables with discrete domains.

The discrete version of $\phi'(x, t)$ is denoted as $\psi[s, n]$; these are formally related to each other as

$$\frac{1}{\Delta x \Delta t} \int_s \int_n \phi'(x, t) dx dt = \psi[s, n] \tag{69}$$

In Eq. (69), s and n enumerate uniformly partitioned intervals that fully span the continuous domains of space V and time T, respectively, and Δx and Δt denote appropriate measures of the intervals. Therefore, the discrete version of $\phi'(x,t)$ essentially captures the averaged values within the uniformly subdivided intervals in space within the time step n.

5.5 Discretization Methods for Local State Space

Functions $\alpha(h,x,t)$ and $m(h,x,t)$ exhibit a dependence on the local state variable, in addition to the spatial and temporal variables. The spatial and temporal variables are discretized using the same method outlined in Eq. (69), but there are two potential strategies to deal with the discretization of these functions with respect to the local state variable. The simplest approach is to discretize the local state space H using triangle or hat basis functions $\Lambda(h-l)$ to divide the local state space into intervals. The hat basis functions are defined in Eq. (70).

$$\Lambda(h-l) = max\left(1 - \left|\frac{h(L-1)}{H} - \frac{Hl}{L-1}\right|, 0\right) \tag{70}$$

In Eq. (19), H and h maintain their definitions as the local state and local state space variable respectively, L is the total number of hat basis functions used to span the local state space and l enumerates the hat functions. The hat functions are placed along the local state space such that the maximum and minimum values of the local state space domain fall on the peak values of the hat functions associated with the largest and smallest values of l. An example of these hat functions with $L = 3$ and $h \in [0,1]$ is found in Fig. 36. Using this method to position the hat functions in the local state space ensures the sum of all hat functions contained within the local state space sum to 1, and that a summation of the hat functions times the microstructure function returns the original microstructure function.

$$\sum_{l=0}^{L-1} \Lambda(h-l) = 1; h \in H \tag{71}$$

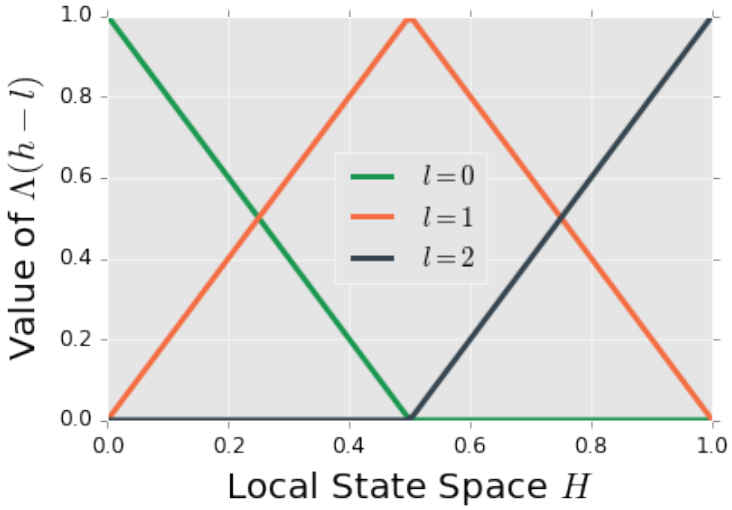


Figure 36: Three hat basis functions Λ used to discretize the local state space H. This method is referred to as Primitive basis functions.

$$\sum_{l=0}^{L-1} \Lambda(h-l)m(h,x,t) = m(h,x,t) \tag{72}$$

In previous work [26], this approach has been referred to as primitive binning using the Primitive basis functions [23, 24, 25, 27, 3] leading to

$$\frac{1}{\Delta x \Delta t} \int_H \int_s \int_n \Lambda(h-l)m(h,x,t)dxdtdh = m[l,s,n] \tag{73}$$

$$\frac{1}{\Delta x \Delta t} \int_H \int_s \int_n \Lambda(h-l)\alpha(h,x,t)dxdtdh = \alpha[l,s,n] \tag{74}$$

where l now enumerates the number of basis functions used to represent the local state variable h. This primitive binning approach results in the MKS formulation that is consistent with most of the prior studies [23, 24, 25, 27, 3], where it is expressed as

$$\psi[s,n] = \sum_{r=0}^{S-1} \sum_{l=0}^{L-1} \alpha[l,r,n]m[l,s-r,0] + ... \tag{75}$$

Alternatively, the functions on the local state space can be represented to adequate accuracy using highly efficient orthogonal basis functions. For example, it is well known that orthogonal functions developed through classical Sturm-Liouville theory can be used as basis functions in many applications. Using such basis orthogonal

functions, one can establish representations such as the ones shown below for an arbitrary function [191, 192, 193, 194]:

$$f(h) = \sum_{l=-\infty}^{\infty} c_l \xi_l(h) \tag{76}$$

$$\frac{1}{N_l} \int_a^b \xi_{l'}(h) \xi_l(h) w(h) dh = \delta_{ll'} \tag{77}$$

In Eq. (77) ξ_l is the l^{th} order orthogonal basis function, $w(h)$ is the weighting function, $\delta_{ll'}$ is the Kronecker delta and N_l is a normalization constant that depends on the order and type of the basis functions. The most important feature of Eq. (76) is that the set of coefficients c_l now provide a discrete representation of the function $f(h)$. This approach is particularly attractive when only a small number of c_l dominate the representation, but requires that the local state domain H is mapped into the interval over which the basis function is orthogonal and orthogonality relationship has a weighting function equal to one, i.e., $w(h) = 1$. Two potential orthogonal bases that meet these criteria are Legendre polynomials and Fourier series [195, 196, 31].

Applying this discretization approach to capture the h dependence in functions $m(h, x, t)$ and $\alpha(h, x, t)$ in Eq. (67), by selecting orthogonal basis functions with $w(h) = 1$ and mapping the local state space to the orthogonal domain leads to the following discretized versions:

$$\frac{1}{\Delta x \Delta t} \int_s \int_n m(h, x, t) dx dt = \sum_{l=0}^{L-1} m[l, s, n] \xi_l(h) \tag{78}$$

$$\frac{1}{\Delta x \Delta t} \int_s \int_n \alpha(h, x, t) dx dt = \sum_{l=0}^{L-1} \alpha[l, s, n] \xi_l(h) \tag{79}$$

The introduction of these discretized representations into Eq. (67) produces the exact same MKS formulation as shown previously in Eq. (80), but with a new interpretation of the index l.

$$\psi[s, n] = \sum_{r=0}^{S-1} \sum_{l=0}^{L-1} \alpha[l, r, n] m[l, s - r, 0] + ... \tag{80}$$

In the remainder of this paper, we first demonstrate the viability of the extended MKS framework presented above for process-structure evolution linkages. Furthermore, in conducting this case study, we will explore using the two approaches described above for the discretization of the functions on the local state space for multiscaling in time by extending the length of time step n to match the time domain of the simulation T through a specific case study.

5.6 Cahn-Hilliard Simulations and MKS Linkage Calibrations

5.6.1 Simulation and MKS Linkage Details

The simulation data as well as the MKS localization linkages used in this case study were generated using the Python library PyMKS [126]. A Cahn-Hilliard simulation is used to generated data for the calibration of the MKS linkages and serves as a reference to compare and validate their performance. The Cahn-Hilliard equation presented earlier in Eq. (55) was solved using the optimized semi-implicit spectral scheme with periodic boundary conditions described by Cheng and Rutenberg [62]. The parameter γ was set equal to 0.2, and the time step for the calibration dataset was set equal to 10^{-2} sec. Two spatial domains with sizes of 100 X 100 and 300 X 300 were used to examine how the different methods scale for larger simulations. The simulations and the MKS linkages were computed on a machine with 8 1.0 GHz processors and 8 GB of memory.

In the present study, we focus on capturing process-microstructure localization linkage in one large time step using the MKS framework. Note that this differs from the approach used in the earlier study [26] where the time derivative of the concentration was used as the output response field (the LHS of Eq. (80)) and it was shown that once this linkage is established for one time step, it can be recursively applied to march forward in time. In the present study we have used the concentration at the end of 500 time steps (or one large time step) as the output response field.

In this study, we explore both the discretization approaches described earlier for the functions on the local state space. The approaches result in two different MKS localization linkages that look alike in their mathematical forms. The linkage referred to as Legendre MKS linkage throughout the remainder of the paper uses the discretization method outlined in Eq. (78) and Eq. (79) with Legendre polynomials as the basis functions. The other linkage uses the discretization method outlined in Eq. (73) and Eq. (74) and is referred to as the Primitive MKS linkage throughout the remainder of this paper. Both models restrict the local state space domain to $h \in$ [-1, 1].

5.6.2 First Order Influence Coefficients and Discrete Fourier Transforms

Only the first term in Eq. (80) is used in this work. In prior work [23, 24, 25, 27, 3], it was shown that the first term is dominant for problems with low to moderate contrast, which in turn controls the degree of heterogeneity of the response field. In the present problem, this criterion is met within the 500 time steps of the simulation. This results in significant computational advantages as the calibration of the first-order MKS localization kernels can be done efficiently by taking advantage of discrete Fourier transforms and the convolution theorem [34, 35]. This transformation leads to

$$P[k,n] = \sum_{l} \beta[l,k,n] M[l,k,0] \tag{81}$$

In Eq. (81) $P[k,n]$, $\beta[l,k,n]$ and $M[l,k,0]$ are the discrete Fourier transforms of $\psi[s,n]$, $\alpha[l,s,n]$ and $m[l,s,n]$ for Eq. (80) respectively.

With the uncoupled spatial frequency representation shown in Eq. (81), the β terms can be calibrated easily using multiple linear regression techniques using the known values for P and M. The discretization used for the Primitive MKS linkage (Eq. (73) and Eq. (74) is subject to the constraint that the discretized microstructure

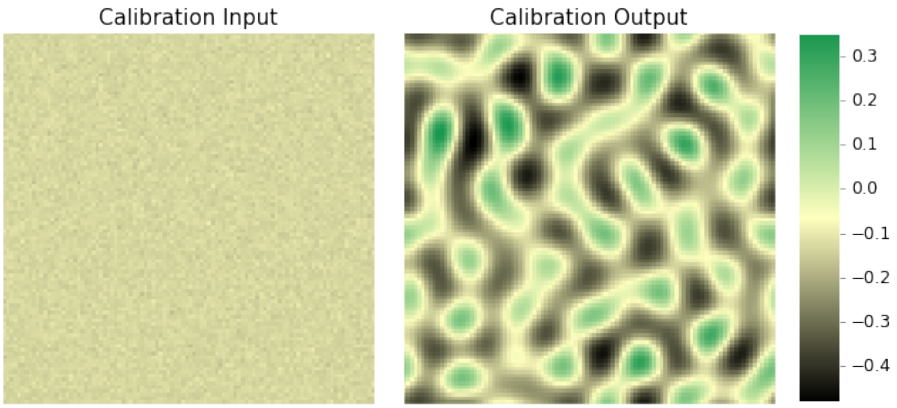


Figure 37: One instance of an initial microstructure (100 x 100) and its corresponding microstructure after 500 small time steps, which were used to calibrate the influence coefficients for MKS localization linkages using both the Legendre and the Primitive basis functions.

function sums to one at any instance in space and time,

$$\sum_{l} m[l, s, n] = 1 \tag{82}$$

therefore multiple linear regression with categorical variables as outlined in previous studies is used [23, 24, 25, 27]. The discretized microstructure function in the Legendre MKS linkage is not subject to the constraint shown in Eq. (82) and therefore standard multiple linear regression is used.

5.6.3 Calibration Data

The first step in the calibration of MKS localization kernels is the generation of a calibration dataset. For this purpose, Eq. (55) was numerically solved for 500 time steps for 500 randomly generated initial concentration fields with values sampled from a normal distributions. The mean values of the normal distribution were randomly selected between [-0.5, 0.5] and standard deviations of 10^{-2}. The concentration fields at the beginning and the end of the 500 small time steps constitute the input and output, respectively for the calibration of the MKS localization linkages. An example of these fields is shown in Fig. 37.

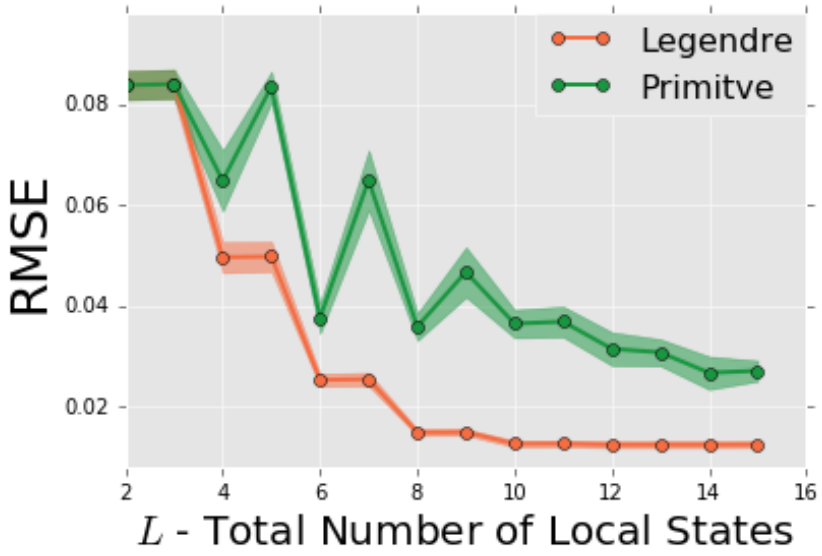


Figure 38: Root mean squared error (points) and standard deviation (line widths) values of the predicted concentration fields found using 10-fold cross-validation of MKS localization evolution-linkages using the Legendre and Primitive basis functions to represent the microstructure function and influence function.

5.6.4 Selection of Local States and Calibration of Influence Coefficients

One of the main MKS parameters for either of the discretization methods is the selection of the of total number of local states (or number of basis functions) L used to describe the local state space. In the previous study it was shown that an increase the in variable L can potentially lead to increased accuracy for small time steps at the cost of making the MKS linkage computationally more expensive [26]. Therefore, the minimum value of L that provides a sufficient level of accuracy is desired.

In order to explore the selection of L for both discretization methods with one large step, the calibration ensemble of data was randomly split into two sets. One set which will be referred to as the linkage selection dataset contains 320 (or 80%) of the microstructures and remaining 80 (or 20%) of the microstructures will be referred to as the linkage validation dataset. In the present work, the two MKS localization linkages were calibrated using the linkage selection dataset while varying L between 2 and 15. In order to avoid over fitting the linkage for a given value of L, 10-fold

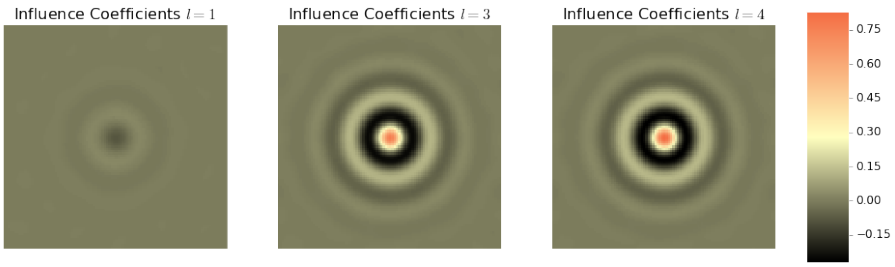


Figure 39: Significant influence coefficients for the Primitive basis for with L equal to 6. All other influence coefficients were less than 10^{-5}.

cross-validation was used. This method randomly partitions the linkage selection dataset into ten equally sized sub-datasets and calibrates the linkage ten times while systematically leaving out each of the sub-datasets once. This method results in 10 calibrations for each value of L amounting to a total of 280 calibrations between the two MKS localization linkages.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(\phi[i]_{simulation} - \phi[i]_{prediction})^2} \tag{83}$$

In order to evaluate the accuracy of the linkages while varying L, the root mean squared error (RMSE) was value computed over every voxel in the sub-datasets used for each of the calibrations with reference to the Cahn-Hilliard simulation (as shown in Eq. 83) and was averaged over the ten cross-validation scores. The averaged RMSE values and their standard deviations are shown in Fig. 38.

The two MKS linkages exhibit a downward overall trend with increasing L. Large and small oscillations were seen in the case of Primitive and Legendre MKS linkages respectively. In the Legendre MKS, it was also observed that the coefficients of the even polynomials in the series were all orders of magnitude smaller than the coefficients of odd polynomials. These observations suggest that the influence functions for the Cahn-Hilliard in Eq. (66) are odd functions. For L greater than 3, the Legendre MKS linkage consistently produced a lower RMSE values compared to the

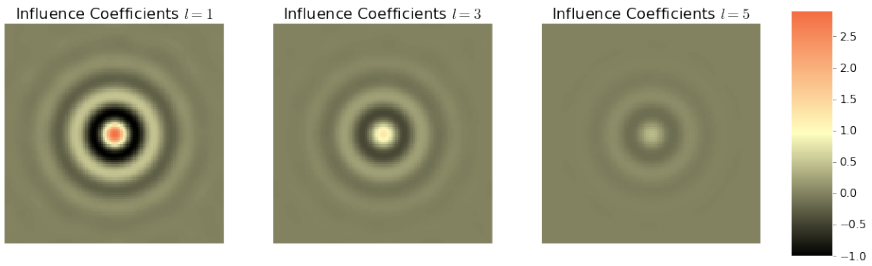


Figure 40: Significant influence coefficients for the Legendre basis with L equal to 6. All other fields had maximum values of less than 10^{-2}.

corresponding Primitive MKS linkage. It should also be noted that with the Primitive basis, the dominant kernels are the ones associated with the important regions in the local state space. With Legendre basis, we get a more organized descriptions of the kernels with the higher-order terms representing the less important contributions, in general, as one would expect for smooth decaying functions such as the influence functions. Figure 38 shows that $L = 6$ provides sufficiently accurate low-cost linkages that can be used to predict processing-structure evolution for the present case study.

The Primitive and Legendre MKS linkages were both calibrated using the entire calibration dataset with the value of L set equal to 6. These two linkages are used for the remainder of the case study. The discretized influence functions (referred to as influence coefficients) for both linkages can be found on Fig. 39 and Fig. 40.

5.7 Microstructure Evolution Linkages for Cahn-Hilliard Simulation

Both MKS linkages calibrated in this study were used to predict the microstructure evolution of the same set of 250 initial concentration fields. The initial microstructures, one instance shown in Fig. 41, generated from normal distributions with mean values randomly selected between [-0.1, 0.1] and standard deviations of 10^{-2} were used as inputs into the Cahn-Hilliard simulation and the two MKS linkages. The simulation numerically predicted each of the microstructures after 500 small time

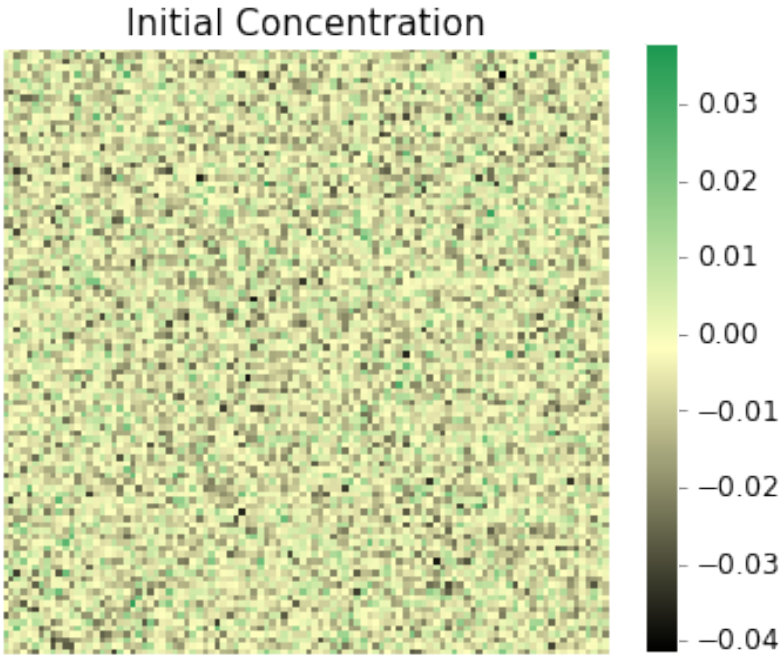


Figure 41: Initial microstructure (100 x 100) used as a common input for the Cahn-Hilliard simulations as well as the MKS localization linkages with Primitive and Legendre basis functions.

steps with an average run time of 1.33 seconds. The MKS linkages used the same initial inputs and the predicted microstructures equivalent to running the simulation for 500 small time steps, but with one large step. The average run time of the Primitive and Legendre MKS linkages were 3.82 x 10^{-3} seconds and 4.94 x 10^{-3} seconds with RMSE values of 5.68 x 10^{-2} and 3.36 x 10^{-2} respectively. An instance of the final predicted microstructures for the MKS linkages and the simulation are compared in Fig. 42.

One of the major advantages of using the MKS localization linkage is that the learning in the form of the MKS influence coefficients (discretized kernels) can be transferred to other initial microstructures that may be defined on larger spatial domains. In other words, the same influence coefficients that are calibrated on a small dataset can be used to predict the structure evolution for a much larger microstructures. This allows for the influence coefficients to be calibrated once and used to

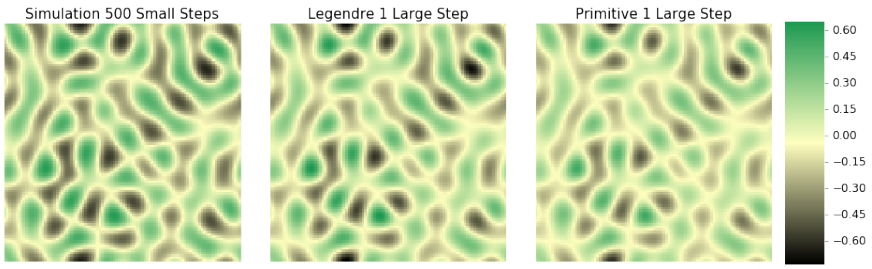


Figure 42: Predicted concentration fields by simulations using 500 small time steps (left) as well as the concentration fields predicted by the two MKS localization linkages using one large time step.

represent the processing-microstructure evolution for simulations with equal or larger domains sizes. Because of the decaying nature of the discretized influence functions in real space as shown in Fig. 39 and Fig. 40, the edges can be zero padded to expand their domain size to match the domain of new initial microstructure. The influence coefficients for both of the Primitive and the Legendre MKS linkages were scaled up from a domain of size 100 by 100 to 300 by 300 using the method outlined by Landi et al. [25].

A set of 250 initial microstructures were created using the same method as described above, but on a larger spatial domain of 300 x 300, is shown in Fig. 43. These microstructures were used as inputs to the same the Cahn-Hilliard simulations and the MKS localization linkages and the final microstructures were predicted. The average run time for the simulation using 500 time steps was 1.13 x 10^2 seconds. Primitive and Legendre MKS linkages had run times of 3.54 x 10^{-2} seconds and 7.00 x 10^{-2} seconds, with RMSE values 7.10 x 10^{-2} and 4.00 x 10^{-2} respectively. The predicted concentration fields are compared in Fig. 44.

As one should expect, the accuracy of the MKS linkages generally improves when a larger number of terms in the series are retained (see Fig. 38). In practice, the level of desired accuracy would be controlled by the application. However, it is important to note that the MKS approach allows the user to establish the desired accuracy level and make the necessary trade-offs between the accuracy and the effort involved.

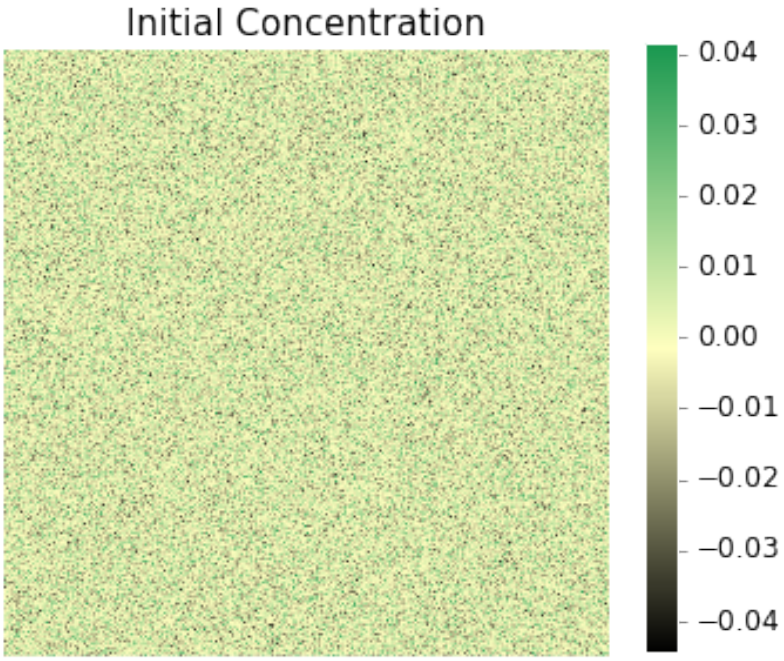


Figure 43: Initial large microstructure (300 x 300) used as a common input for the Cahn-Hilliard simulations and the MKS localization linkages.

5.8 Conclusion

A new generalized MKS localization framework with two different discretization methods for the local state variables has been developed for formulating computationally low-cost process-structure linkages which allow for temporal multiscaling. This framework is quite general and allows compact representation of the influence functions (or kernels) on the local state spaces. The overall framework was presented and demonstrated using a Cahn-Hilliard microstructure evolution as a prime example.

Although the computational cost of the Primitive MKS linkage was slightly lower than the Legendre MKS linkage, and the Legendre MKS linkage was more accurate and showed more smooth decay of error with increasing number of terms in the series. Both MKS localization linkages predicted the process-structure evolution for the concentration fields three orders of magnitude faster than the simulation with a small time step.

This case study suggests that MKS localization framework provides an alternate

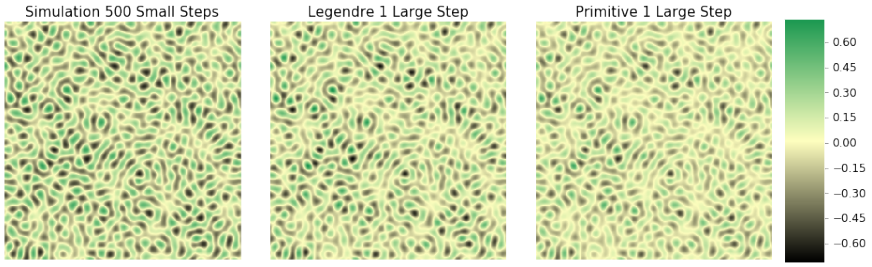


Figure 44: The concentration fields predicted by the numerical simulation with a 500 small time steps (left) as well as the concentration fields predicted by the MKS localization linkages with one large time step using scaled up influence coefficients from a domain size of 100 by 100 to 300 by 300 with Primitive and Legendre bases.

method to learn the underlying embedded physics in a numerical model. This form of expression of the underlying physics as Green's function based influence kernels (as opposed to expression in the form of differential equations) may provide certain computational advantages in rapid exploration of large spaces in process design to attain desired or specified microstructures. This is especially the case for problems where traditional numerical integration schemes have been difficult to optimize.

Overall, it was demonstrated that the MKS kernels extracted for the example studied were indeed insensitive to the details of the initial microstructure (in other words the same kernel can be applied to any initial microstructure in the selected material system) and could be trivially expanded for applications to larger domain sizes with comparable accuracy. The method described here has laid a strong foundation for future developments addressing a broad range of materials systems with richer microstructures and more complex governing physics.

REFERENCES

[1] CORE-Materials. CORE-Materials - A resource repository contains a large number of open educational resources (OERs) in Materials Science and Engineering. `https://www.flickr.com/people/core-materials/`, 2009. [Online; accessed 6-April-2016].

[2] Akash Gupta, Ahmet Cecen, Sharad Goyal, Amarendra K Singh, and Surya R Kalidindi. Structure–property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system. *Acta Materialia*, 91:239–254, 2015.

[3] Yuksel C Yabansu and Surya R Kalidindi. Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals. *Acta Materialia*, 94:26–35, 2015.

[4] Charles Ward. Materials genome initiative for global competitiveness. In *23rd Advanced Aerospace Materials and Processes (AeroMat) Conference and Exposition*. Asm, 2012.

[5] John Allison, Dan Backman, and Leo Christodoulou. Integrated computational materials engineering: a new paradigm for the global materials profession. *JOM*, 58(11):25–27, 2006.

[6] National Science and Technology Council Executive Office of the President. Materials genome initiative for global competitiveness, 2011.

[7] Materials Genome Initiative National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative. Materials genome initiative strategic plan, 2014.

[8] John Allison. Integrated computational materials engineering: A perspective on progress and future steps. *JOM Journal of the Minerals, Metals and Materials Society*, 63(4):15–18, 2011.

[9] Gregory B Olson. Designing a new material world. *Science*, 288(5468):993–998, 2000.

[10] National Research Council (US). Committee on Integrated Computational Materials Engineering. *Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security*. National Academies Press, 2008.

[11] Georg J Schmitz and Ulrich Prahl. *Integrative computational materials engineering: Concepts and applications of a modular simulation platform.* John Wiley & Sons, 2012.

[12] Lynne Robinson. Tms study charts a course to successful icme implementation, 2013.

[13] John E Allison. Integrated computational materials engineering (icme): A transformational discipline for the global materials profession. *Metals and Materials*, page 223.

[14] A Study Organized by The Minerals, Metals & Materials Society. Integrated computational materials engineering (icme): Implementing icme in the aerospace, automotive, and maritime industries, 2013.

[15] Alan Anderson. Report to the president on ensuring american leadership in advanced manufacturing. *Executive Office of the President*, 2011.

[16] Nicholas Young. *An introduction to Hilbert space.* Cambridge university press, 1988.

[17] Paul Richard Halmos, Paul Richard Halmos, Paul Richard Halmos, Paul Richard Halmos, and Hungary Mathematician. *Introduction to Hilbert space and the theory of spectral multiplicity.* Chelsea New York, 1957.

[18] Walter Rudin. *Real and complex analysis.* Tata McGraw-Hill Education, 1987.

[19] Gerald B Folland. *Fourier analysis and its applications*, volume 4. American Mathematical Soc., 1992.

[20] Graeme W Milton. *The theory of composites*, volume 6. Cambridge University Press, 2002.

[21] Brent L Adams, Surya Kalidindi, and David T Fullwood. *Microstructure-sensitive design for performance optimization.* Butterworth-Heinemann, 2013.

[22] Surya R Kalidindi, Stephen R Niezgoda, and Ayman A Salem. Microstructure informatics using higher-order statistics and efficient data-mining protocols. *Jom*, 63(4):34–41, 2011.

[23] Hamad F Al-Harbi, Giacomo Landi, and Surya R Kalidindi. Multi-scale modeling of the elastic response of a structural component made from a composite material using the materials knowledge system. *Modelling and Simulation in Materials Science and Engineering*, 20(5):055001, 2012.

[24] Surya R Kalidindi, Stephen R Niezgoda, Giacomo Landi, Shraddha Vachhani, and Tony Fast. A novel framework for building materials knowledge systems. *Computers, Materials, & Continua*, 17(2):103–125, 2010.

[25] Giacomo Landi, Stephen R Niezgoda, and Surya R Kalidindi. Multi-scale modeling of elastic response of three-dimensional voxel-based microstructure datasets using novel dft-based knowledge systems. *Acta Materialia*, 58(7):2716–2725, 2010.

[26] Tony Fast, Stephen R Niezgoda, and Surya R Kalidindi. A new framework for computationally efficient structure–structure evolution linkages to facilitate high-fidelity scale bridging in multi-scale materials models. *Acta Materialia*, 59(2):699–707, 2011.

[27] Yuksel C Yabansu, Dipen K Patel, and Surya R Kalidindi. Calibrated localization relationships for elastic response of polycrystalline aggregates. *Acta Materialia*, 81:151–160, 2014.

[28] SR Niezgoda, DT Fullwood, and SR Kalidindi. Delineation of the space of 2-point correlations in a composite material system. *Acta Materialia*, 56(18):5285–5292, 2008.

[29] David T Fullwood, Surya R Kalidindi, and Brent L Adams. Second-order microstructure sensitive design using 2-point spatial correlations. In *Electron backscatter diffraction in materials science*, pages 177–188. Springer, 2009.

[30] Stephen R Niezgoda, David M Turner, David T Fullwood, and Surya R Kalidindi. Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics. *Acta Materialia*, 58(13):4432–4445, 2010.

[31] George B Arfken. *Mathematical methods for physicists*. Academic press, 2013.

[32] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[33] A Çeçen, T Fast, EC Kumbur, and SR Kalidindi. A data-driven approach to establishing microstructure–property relationships in porous transport layers of polymer electrolyte fuel cells. *Journal of Power Sources*, 245:144–153, 2014.

[34] Ron Bracewell. The fourier transform and iis applications. *New York*, 1965.

[35] E Oran Brigham and EO Brigham. *The fast Fourier transform and its applications*, volume 1. Prentice Hall Englewood Cliffs, NJ, 1988.

[36] Tony Fast and Surya R Kalidindi. Formulation and calibration of higher-order elastic localization relationships using the mks approach. *Acta Materialia*, 59(11):4595–4605, 2011.

[37] Mohanbir Sawhney, Gianmario Verona, and Emanuela Prandelli. Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of interactive marketing*, 19(4):4–17, 2005.

[38] Aled M Edwards, Chas Bountra, David J Kerr, and Timothy M Willson. Open access chemical and clinical probes to support drug discovery. *Nature Chemical Biology*, 5(7):436–440, 2009.

[39] Marcia Bayne-Smith, Terry Mizrahi, and Martha Garcia. Interdisciplinary community collaboration: Perspectives of community practitioners on successful strategies. *Journal of Community Practice*, 16(3):249–269, 2008.

[40] Kevin Boudreau. Open platform strategies and innovation: Granting access vs. devolving control. *Management Science*, 56(10):1849–1872, 2010.

[41] Georges Aad, T Abajyan, B Abbott, J Abdallah, S Abdel Khalek, AA Abdelalim, O Abdinov, R Aben, B Abi, M Abolins, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.

[42] Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[43] Justin Cranshaw and Aniket Kittur. The polymath project: lessons from a successful online collaboration in mathematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1865–1874. ACM, 2011.

[44] Janis L Dickinson, Benjamin Zuckerberg, and David N Bonter. Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution and systematics*, 41:149–72, 2010.

[45] Wesley M Hochachka, Daniel Fink, Rebecca A Hutchinson, Daniel Sheldon, Weng-Keen Wong, and Steve Kelling. Data-intensive science applied to broad-scale citizen science. *Trends in ecology & evolution*, 27(2):130–137, 2012.

[46] Daniel Atkins. Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. 2003.

[47] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[48] Eric Jones, Travis Oliphant, and Pearu Peterson. {SciPy}: open source scientific tools for {Python}. 2014.

[49] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[50] Surya R Kalidindi, Hari K Duvvuru, and Marko Knezevic. Spectral calibration of crystal plasticity models. *Acta Materialia*, 54(7):1795–1804, 2006.

[51] Joshua B Shaffer, Marko Knezevic, and Surya R Kalidindi. Building texture evolution networks for deformation processing of polycrystalline fcc metals using spectral approaches: applications to process design for targeted performance. *International Journal of Plasticity*, 26(8):1183–1194, 2010.

[52] Marko Knezevic, Amanda Levinson, Ryan Harris, Raja K Mishra, Roger D Doherty, and Surya R Kalidindi. Deformation twinning in az31: influence on strain hardening and texture evolution. *Acta Materialia*, 58(19):6230–6242, 2010.

[53] Hamad F Al-Harbi, Marko Knezevic, and Surya R Kalidindi. Spectral approaches for the fast computation of yield surfaces and first-order plastic property closures for polycrystalline materials with cubic-triclinic textures. *Computers, Materials, & Continua*, 15(2):153–172, 2010.

[54] Hari K Duvvuru, Marko Knezevic, Raja K Mishra, and Surya R Kalidindi. Application of microstructure sensitive design to fcc polycrystals. In *Materials science forum*, volume 546, pages 675–680. Trans Tech Publ, 2007.

[55] DS Li, H Garmestani, and Scott Schoenfeld. Evolution of crystal orientation distribution coefficients during plastic deformation. *Scripta Materialia*, 49(9):867–872, 2003.

[56] DS Li, H Garmestani, and BL Adams. A texture evolution model in cubic-orthotropic polycrystalline system. *International journal of plasticity*, 21(8):1591–1617, 2005.

[57] DS Li, H Garmestani, and S Ahzi. Processing path optimization to achieve desired texture in polycrystalline materials. *Acta materialia*, 55(2):647–654, 2007.

[58] Dong Sheng Li, J Bouhattate, and Hamid Garmestani. Processing path model to describe texture evolution during mechanical processing. In *Materials Science Forum*, volume 495, pages 977–982. Trans Tech Publ, 2005.

[59] Adam Creuziger, Lin Hu, Thomas Gnäupel-Herold, and Anthony D Rollett. Crystallographic texture evolution in 1008 steel sheet during multi-axial tensile strain paths. *Integrating Materials and Manufacturing Innovation*, 3(1):1, 2014.

[60] Veera Sundararaghavan and Nicholas Zabaras. A multi-length scale sensitivity analysis for the control of texture-dependent properties in deformation processing. *International Journal of Plasticity*, 24(9):1581–1605, 2008.

[61] Veera Sundararaghavan and Nicholas Zabaras. Linear analysis of texture–property relationships using process-based representations of rodrigues space. *Acta materialia*, 55(5):1573–1587, 2007.

[62] Mowei Cheng and Andrew D. Rutenberg. Maximally fast coarsening algorithms. *Phys. Rev. E*, 72:055701, Nov 2005.

[63] Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.

[64] GR Strobl. A new method of evaluating slit-smeared small-angle x-ray scattering data. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 26(3):367–375, 1970.

[65] John W Cahn and John E Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of chemical physics*, 28(2):258–267, 1958.

[66] Long-Qing Chen. Phase-field models for microstructure evolution. *Annual review of materials research*, 32(1):113–140, 2002.

[67] Zhiqiang Bi and Robert F Sekerka. Phase-field model of solidification of a binary alloy. *Physica A: Statistical Mechanics and its Applications*, 261(1):95–106, 1998.

[68] David L McDowell and Surya R Kalidindi. The materials innovation ecosystem: A key enabler for the materials genome initiative. *MRS Bulletin*, 41(04):326–337, 2016.

[69] Surya R Kalidindi. Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials. *International Materials Reviews*, 60(3):150–168, 2015.

[70] Surya R Kalidindi. *Hierarchical Materials Informatics: Novel Analytics for Materials Data*. Elsevier, 2015.

[71] Talapady N Bhat, Laura M Bartolo, Ursula R Kattner, Carelyn E Campbell, and John T Elliott. Strategy for extensible, evolving terminology for the materials genome initiative efforts. *JOM*, 67(8):1866–1875, 2015.

[72] National Institute of Standards and Technology. Nist data gateway, 2001.

[73] NIST Materials Measurement Laboratory. Nist repositories dspace, 2015.

[74] National Institute of Standards and Technology. Nist data curation system, 2014.

[75] James E Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Christopher Wolverton. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *Jom*, 65(11):1501–1509, 2013.

[76] LLC MatWeb. Matweb - materials property data, 1996.

[77] Stefano Curtarolo, Wahyu Setyawan, Gus LW Hart, Michal Jahnatek, Roman V Chepulskii, Richard H Taylor, Shidong Wang, Junkai Xue, Kesong Yang, Ohad Levy, et al. Aflow: an automatic framework for high-throughput materials discovery. *Computational Materials Science*, 58:218–226, 2012.

[78] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[79] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *Apl Materials*, 1(1):011002, 2013.

[80] Kim Project. Openkim - the knowledgebase of interatomic models, 2010.

[81] Prism Project. Predictive integrated structural materials science (prisms), 2010.

[82] S Plimpton, A Thompson, and A Slepoy. Spparks kinetic monte carlo simulator, 2012.

[83] Derek Gaston, Chris Newman, Glen Hansen, and Damien Lebrun-Grandie. Moose: A parallel computational framework for coupled systems of nonlinear equations. *Nuclear Engineering and Design*, 239(10):1768–1778, 2009.

[84] Michael A Groeber and Michael A Jackson. Dream. 3d: a digital representation environment for the analysis of microstructure in 3d. *Integrating Materials and Manufacturing Innovation*, 3(1):1–17, 2014.

[85] Ramon C Littell. *Sas*. Wiley Online Library, 2006.

[86] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pages 57–61, 2010.

[87] Davide Albanese, Roberto Visintainer, Stefano Merler, Samantha Riccadonna, Giuseppe Jurman, and Cesare Furlanello. mlpy: Machine learning python. *arXiv preprint arXiv:1202.6548*, 2012.

[88] Ian J Goodfellow, David Warde-Farley, Pascal Lamblin, Vincent Dumoulin, Mehdi Mirza, Razvan Pascanu, James Bergstra, Frédéric Bastien, and Yoshua Bengio. Pylearn2: a machine learning research library. *arXiv preprint arXiv:1308.4214*, 2013.

[89] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.

[90] Andreas C Müller and Sven Behnke. Pystruct: learning structured prediction in python. *The Journal of Machine Learning Research*, 15(1):2055–2060, 2014.

[91] Janez Demšar, Blaž Zupan, Gregor Leban, and Tomaz Curk. *Orange: From experimental machine learning to interactive data mining*. Springer, 2004.

[92] Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[93] Stefan Van Der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

[94] R Hill. Elastic properties of reinforced solids: some theoretical principles. *Journal of the Mechanics and Physics of Solids*, 11(5):357–372, 1963.

[95] Zvi Hashin. Analysis of composite materialsa survey. *Journal of Applied Mechanics*, 50(3):481–505, 1983.

[96] William Fuller Brown Jr. Solid mixture permittivities. *The Journal of Chemical Physics*, 23(8):1514–1517, 1955.

[97] E Kröner. Statistical modelling. In *Modelling small deformations of polycrystals*, pages 229–291. Springer, 1986.

[98] E Kröner. Bounds for effective elastic moduli of disordered materials. *Journal of the Mechanics and Physics of Solids*, 25(2):137–155, 1977.

[99] Ekkehart Kröner. *Statistical continuum mechanics*. Springer, 1972.

[100] Pavel Etingof and Brent L Adams. Representations of polycrystalline microstructure by n-point correlation tensors. *Texture, Stress, and Microstructure*, 21(1):17–37, 1993.

[101] Brent L Adams and T Olson. The mesostructureproperties linkage in polycrystals. *Progress in Materials Science*, 43(1):1–87, 1998.

[102] David T Fullwood, Brent L Adams, and Surya R Kalidindi. A strong contrast homogenization formulation for multi-phase anisotropic materials. *Journal of the Mechanics and Physics of Solids*, 56(6):2287–2297, 2008.

[103] Salvatore Torquato. *Random heterogeneous materials: microstructure and macroscopic properties*, volume 16. Springer Science & Business Media, 2013.

[104] DS Li, Ghazal Saheli, M Khaleel, and Hamid Garmestani. Quantitative prediction of effective conductivity in anisotropic heterogeneous media using two-point correlation functions. *Computational Materials Science*, 38(1):45–50, 2006.

[105] Jacqueline Milhans, DS Li, M Khaleel, Xin Sun, and Hamid Garmestani. Prediction of the effective coefficient of thermal expansion of heterogeneous media using two-point correlation functions. *Journal of Power Sources*, 196(8):3846–3850, 2011.

[106] H Garmestani, S Lin, BL Adams, and S Ahzi. Statistical continuum theory for large plastic deformation of polycrystalline materials. *Journal of the Mechanics and Physics of Solids*, 49(3):589–607, 2001.

[107] Brent L Adams, Xiang Carl Gao, and Surya R Kalidindi. Finite approximations to the second-order properties closure in single phase polycrystals. *Acta Materialia*, 53(13):3563–3577, 2005.

[108] Massimiliano Binci, David Fullwood, and Surya R Kalidindi. A new spectral framework for establishing localization relationships for elastic behavior of composites and their calibration to finite-element models. *Acta Materialia*, 56(10):2272–2282, 2008.

[109] Stephen R Niezgoda, Anand K Kanjarla, and Surya R Kalidindi. Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integrating Materials and Manufacturing Innovation*, 2(1):1–27, 2013.

[110] Stephen R Niezgoda, Yuksel C Yabansu, and Surya R Kalidindi. Understanding and visualizing microstructure and microstructure variance as a stochastic process. *Acta Materialia*, 59(16):6387–6400, 2011.

[111] Siddiq M Qidwai, David M Turner, Stephen R Niezgoda, Alexis C Lewis, Andrew B Geltmacher, David J Rowenhorst, and Surya R Kalidindi. Estimating the response of polycrystalline materials using sets of weighted statistical volume elements. *Acta Materialia*, 60(13):5284–5299, 2012.

[112] David B Brough, Daniel Wheeler, James A Warren, and Surya R Kalidindi. Microstructure-based knowledge systems for capturing process-structure evolution linkages. *Current Opinion in Solid State and Materials Science*, 2016.

[113] Ahmet Cecen, Tony Fast, and Surya R Kalidindi. Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure. *Integrating Materials and Manufacturing Innovation*, 5(1):1–15, 2016.

[114] Fernando Prez, Brian E. Granger, and John D. Hunter. Python: An ecosystem for scientific computing. *Computing in Science & Engineering*, 13(2):13–21, 2011.

[115] The mit license (mit). `https://opensource.org/licenses/mit-license.php`. Accessed: 2016-05-18.

[116] Pytest. `http://pytest.org`, 2016.

[117] Robert Cimrman. Sfepy-write your own fe application. *arXiv preprint arXiv:1404.6391*, 2014.

[118] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 3, pages 1381–1384. IEEE, 1998.

[119] John D Hunter et al. Matplotlib: A 2d graphics environment. *Computing in science and engineering*, 9(3):90–95, 2007.

[120] Jitesh H Panchal, Surya R Kalidindi, and David L McDowell. Key computational modeling issues in integrated computational materials engineering. *Computer-Aided Design*, 45(1):4–25, 2013.

[121] Surya R Kalidindi, Joshua A Gomberg, Zachary T Trautt, and Chandler A Becker. Application of data science tools to quantify and distinguish between structures and models in molecular dynamics datasets. *Nanotechnology*, 26(34):344006, 2015.

[122] Vito Volterra. *Theory of functionals and of integral and integro-differential equations*. Courier Corporation, 2005.

[123] Daniel B Suits. Use of dummy variables in regression equations. *Journal of the American Statistical Association*, 52(280):548–551, 1957.

[124] Francis Galton. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, pages 246–263, 1886.

[125] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[126] Daniel Wheeler, David Brough, Tony Fast, Surya Kalidindi, and Andrew Reid. PyMKS: Materials Knowledge System in Python. 05 2014.

[127] Sebastian Mika, Bernhard Schölkopf, Alexander J Smola, Klaus-Robert Müller, Matthias Scholz, and Gunnar Rätsch. Kernel pca and de-noising in feature spaces. In *NIPS*, volume 4, page 7. Citeseer, 1998.

[128] Jeffrey D Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.

[129] Rebecca M Warner. *Spectral analysis of time-series data*. Guilford Press, 1998.

[130] Clive William John Granger, Michio Hatanaka, et al. Spectral analysis of economic time series. *Spectral analysis of economic time series.*, 1964.

[131] Kin-Pong Chan and Ada Wai-Chee Fu. Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 126–133. IEEE, 1999.

[132] Aslak Grinsted, John C Moore, and Svetlana Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics*, 11(5/6):561–566, 2004.

[133] Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge university press, 2006.

[134] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[135] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[136] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.

[137] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[138] Lawrence R Rabiner and Biing-Hwang Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

[139] Cl Masreliez and R Martin. Robust bayesian estimation for the linear model and robustifying the kalman filter. *IEEE transactions on Automatic Control*, 22(3):361–371, 1977.

[140] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

[141] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee, 2000.

[142] Fredrik Gustafsson and Gustaf Hendeby. Some relations between extended and unscented kalman filters. *IEEE Transactions on Signal Processing*, 60(2):545–555, 2012.

[143] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[144] Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.

[145] Alan Pankratz. *Forecasting with univariate Box-Jenkins models: Concepts and cases*, volume 224. John Wiley & Sons, 2009.

[146] Samuele Salti and Luigi Di Stefano. On-line support vector regression of the transition model for the kalman filter. *Image and Vision Computing*, 31(6):487–501, 2013.

[147] Surya R Kalidindi and Marc De Graef. Materials data science: current status and future outlook. *Annual Review of Materials Research*, 45:171–193, 2015.

[148] Peter AW Lewis and James G Stevens. Nonlinear modeling of time series using multivariate adaptive regression splines (mars). *Journal of the American Statistical Association*, 86(416):864–877, 1991.

[149] Jan G De Gooijer, Bonnie K Ray, and Horst Kräger. Forecasting exchange rates using tsmars. *Journal of International Money and Finance*, 17(3):513–534, 1998.

[150] G Spanos, J Allison, B Cowles, J Deloach, and T Pollock. Integrated computational materials engineering (icme): Implementing icme in the aerospace, automotive, and maritime industries, miner. *Met. Mater. Soc.(TMS)*, 2013.

[151] Markus J Buehler, Alexander Hartmaier, and Huajian Gao. Hierarchical multiscale modelling of plasticity of submicron thin metal films. *Modelling And Simulation In Materials Science And Engineering*, 12(4):S391, 2004.

[152] S Groh, EB Marin, MF Horstemeyer, and HM Zbib. Multiscale modeling of the plasticity in an aluminum single crystal. *International Journal of Plasticity*, 25(8):1456–1473, 2009.

[153] Darby J Luscher, David L McDowell, and Curt A Bronkhorst. A second gradient theoretical framework for hierarchical multiscale modeling of materials. *International Journal of Plasticity*, 26(8):1248–1275, 2010.

[154] J Tinsley Oden, Kumar Vemaganti, and Nicolas Moës. Hierarchical modeling of heterogeneous solids. *Computer Methods in Applied Mechanics and Engineering*, 172(1):3–25, 1999.

[155] David L McDowell. A perspective on trends in multiscale plasticity. *International Journal of Plasticity*, 26(9):1280–1309, 2010.

[156] Gregory B Olson. Computational design of hierarchically structured materials. *Science*, 277(5330):1237–1242, 1997.

[157] Mahesh Shenoy, Yustianto Tjiptowidjojo, and David McDowell. Microstructure-sensitive modeling of polycrystalline in 100. *International Journal of Plasticity*, 24(10):1694–1730, 2008.

[158] The Minerals Metals & Materials Society (TMS). *Modeling Across Scales: A Roadmapping Study for Connecting Materials Models and Simulations Across Length and Time Scales*. TMS, Warrendale, PA, 2015.

[159] James Maxwell et al. A treatise on electricity and magnetism. 1873.

[160] Helmut J Böhm. A short introduction to basic aspects of continuum micromechanics. Technical report, CDLFMD Report, Technical Report 020624, 31998, 1998.

[161] Igor Sevostianov and Albert Giraud. Generalization of maxwell homogenization scheme for elastic material containing inhomogeneities of diverse shape. *International Journal of Engineering Science*, 64:23–36, 2013.

[162] Woldemar Voigt. *Lehrbuch der kristallphysik:(mit ausschluss der kristalloptik)*, volume 34. BG Teubner, 1910.

[163] A Reuss. Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 9(1):49–58, 1929.

[164] R1 Hill. A self-consistent mechanics of composite materials. *Journal of the Mechanics and Physics of Solids*, 13(4):213–222, 1965.

[165] JOHN D Eshelby. The determination of the elastic field of an ellipsoidal inclusion, and related problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 241, pages 376–396. The Royal Society, 1957.

[166] Z Hashin and S Shtrikman. Note on a variational approach to the theory of composite elastic materials. *Journal of the Franklin Institute*, 271(4):336–341, 1961.

[167] Michel Bornert, Thierry Bretheau, and Pierre Gilormini. *Homogénéisation en mécanique des matériaux: comportements non linéaires et problèmes ouverts*. Hermes Science Publications, 2001.

[168] RM Christensen and KH Lo. Solutions for effective shear properties in three phase sphere and cylinder models. *Journal of the Mechanics and Physics of Solids*, 27(4):315–330, 1979.

[169] Eveline Herve and Andre Zaoui. N-layered inclusion-based micromechanical modelling. *International Journal of Engineering Science*, 31(1):1–10, 1993.

[170] Jianmin Qu and Mohammed Cherkaoui. *Frontmatter*. Wiley Online Library, 2006.

[171] JR Willis. Properties of composites. *Advances in applied mechanics*, 21:1, 1982.

[172] P Ponte Castañeda and JOHN R Willis. The effect of spatial distribution on the effective behavior of composite materials and cracked media. *Journal of the Mechanics and Physics of Solids*, 43(12):1919–1951, 1995.

[173] S Torquato. Effective electrical conductivity of two-phase disordered composite media. *Journal of Applied Physics*, 58(10):3790–3797, 1985.

[174] S Torquato. Effective stiffness tensor of composite mediai. exact series expansions. *Journal of the Mechanics and Physics of Solids*, 45(9):1421–1448, 1997.

[175] Xin Dong, David L McDowell, Surya R Kalidindi, and Karl I Jacob. Dependence of mechanical properties on crystal orientation of semi-crystalline polyethylene structures. *Polymer*, 55(16):4248–4257, 2014.

[176] HJ Bunge and C Esling. Texture development by plastic deformation. *Scripta metallurgica*, 18(3):191–195, 1984.

[177] Ken Sekimoto. Evolution of the domain structure during the nucleation-and-growth process with non-conserved order parameter. *Physica A: Statistical Mechanics and its Applications*, 135(2):328–346, 1986.

[178] Sai Kiranmayee Samudrala, Jaroslaw Zola, Srinivas Aluru, and Baskar Ganapathysubramanian. Parallel framework for dimensionality reduction of large-scale datasets. *Scientific Programming*, 2015:1, 2015.

[179] Kunok Chang, Carl E Krill III, Qiang Du, and Long-Qing Chen. Evaluating microstructural parameters of three-dimensional grains generated by phase-field simulation or other voxel-based techniques. *Modelling and Simulation in Materials Science and Engineering*, 20(7):075009, 2012.

[180] Mowei Cheng and James A Warren. Controlling the accuracy of unconditionally stable algorithms in the cahn-hilliard equation. *Physical Review E*, 75(1):017702, 2007.

[181] Surya R Kalidindi, Massimiliano Binci, David Fullwood, and Brent L Adams. Elastic properties closures using second-order homogenization theories: case studies in composites of two isotropic constituents. *Acta materialia*, 54(11):3117–3126, 2006.

[182] Yves-Patrick Pellegrini. Self-consistent effective-medium approximation for strongly nonlinear media. *Physical Review B*, 64(13):134211, 2001.

[183] Yves-Patrick Pellegrini and Marc Barthélémy. Self-consistent effective-medium approximations with path integrals. *Physical Review E*, 61(4):3547, 2000.

[184] V Monchiet and G Bonnet. A polarization-based fft iterative scheme for computing the effective properties of elastic composites with arbitrary contrast. *International Journal for Numerical Methods in Engineering*, 89(11):1419–1436, 2012.

[185] Vincent Monchiet and Guy Bonnet. Numerical homogenization of nonlinear composites with a polarization-based fft iterative scheme. *Computational Materials Science*, 79:276–283, 2013.

[186] Hervé Moulinec and Fabrice Silva. Comparison of three accelerated fft-based schemes for computing the mechanical response of composite materials. *International Journal for Numerical Methods in Engineering*, 97(13):960–985, 2014.

[187] François Willot, Bassam Abdallah, and Yves-Patrick Pellegrini. Fourier-based schemes with modified green operator for computing the electrical response of heterogeneous media with accurate local fields. *International Journal for Numerical Methods in Engineering*, 98(7):518–533, 2014.

[188] DT Fullwood, BL Adams, and SR Kalidindi. Generalized pareto front methods applied to second-order material property closures. *Computational Materials Science*, 38(4):788–799, 2007.

[189] Alan V Oppenheim, Ronald W Schafer, John R Buck, et al. *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs, 1989.

[190] Sanjit Kumar Mitra and Yonghong Kuo. *Digital signal processing: a computer-based approach*, volume 2. McGraw-Hill New York, 2006.

[191] George B Arfken, Hans J Weber, and Donald Spector. Mathematical methods for physicists. *American Journal of Physics*, 67(2):165–169, 1999.

[192] Anton Zettl. *Sturm-Liouville Theory*, volume 121. American Mathematical Soc., 2010.

[193] Joachim Weidmann. *Spectral theory of ordinary differential operators*. Springer, 1987.

[194] Mohammed Al-Gwaiz. *Sturm-Liouville theory and its applications*. Springer, 2007.

[195] Raymond M Redheffer and Robert M Young. Completeness and basis properties of complex exponentials. *Transactions of the American Mathematical Society*, 277(1):93–111, 1983.

[196] Earl David Rainville. *Special functions*. New York, 1960.