Chapter 3

Materials Informatics and Data System for Polymer Nanocomposites Analysis and Design

Wei Chen[*,‖], Linda Schadler[†], Cate Brinson[‡], Yixing Wang[*], Yichi Zhang[*],
Aditya Prasad[¶], Xiaolin Li[§] and Akshay Iyer[*]

[*]*Department of Mechanical Engineering, Northwestern University,*
2145, Sheridan Rd., Evanston, IL 60208–3111, USA
[†]*College of Engineering and Mathematical Sciences, University of Vermont,*
Burlington, VT 05405, USA
[‡]*Department of Mechanical Engineering and Materials Science,*
Duke University, Durham, NC 27708, USA
[§]*Theoretical & Applied Mechanics, Northwestern University,*
2145, Sheridan Rd., Evanston, IL 60208–3111, USA
[¶]*Department of Materials Science and Engineering, Rensselaer Polytechnic*
Institute, 110 8th St, Troy, NY 12180, USA
[‖]*weichen@northwestern.edu*

The application of Materials Informatics to polymer nanocomposites would result in faster development and commercial implementation of these promising materials, particularly in applications requiring a unique combination of properties. This chapter focuses on a new data resource for nanocomposites — NanoMine — and the tools, models, and algorithms that support data-driven materials design. The chapter begins with a brief introduction to NanoMine, including the data structure and tools available. Critical to the ability to design nanocomposites, however, is developing robust structure–property–processing (s–p–p) relationships. Central to this development is the choice of appropriate microstructure characterization and reconstruction (MCR) techniques that capture a complex morphology and ultimately build statistically equivalent reconstructed composites for accurate modeling of properties. A wide range of MCR techniques is reviewed followed by an introduction of feature selection and feature extraction techniques to identify the most significant microstructure features for dimension reduction. This is then followed by examples of using a descriptor-based representation to create processing–structure (p–s) and structure–property (s–p) relationships for use in design. To overcome the difficulty in modeling the interphase region surrounding nanofillers, an adaptive sampling approach is presented to inversely determine the interphase properties based on both FEM simulations and physical experiment data of bulk properties. Finally, a case study for nanodielectrics in a capacitor is introduced to demonstrate the integration of the p–s and s–p relationships to develop optimized materials for achieving multiple desired properties.

1. Introduction

The past 5 years have seen substantial growth in access to digital materials data with the goal to accelerate materials design under the national Materials Genome Initiative (MGI).[1–4] New materials informatics techniques[4–10] are being developed to centralize materials data and include information across a significant range of length and time scales in order to improve use of data mining, statistics, image processing and visualization, and predictive analytics over the lifecycle development of materials systems. However, unlike the metallic and inorganic alloy fields, polymers and their nanocomposites are less developed in both database system and data-driven analytical protocols. The complexity and high dimensionality of the polymer and polymer nanocomposite data space, including details on processing conditions, nanoscale filler dispersion, as well as properties, make it challenging to implement a universal standard that could archive all possible nanocomposite data and facilitate the intentional design of polymer nanocomposites. Additionally, small changes in processing conditions or surface chemistry can result in drastic changes in filler–matrix interaction and filler microstructure, which can cause significant changes in composite properties.[11] These factors all hinder the establishment of a comprehensive methodology to fully incorporate processing, structure, and property (p–s–p) information for nanocomposite materials into the design process. Instead, the design and development of new nanocomposite materials remains largely dependent on Edisonian, trial-and-error iterations. To improve our ability to design nanocomposites, it is essential to gain a deeper mechanistic understanding of, and the ability to map and quickly search, the p–s–p space for new polymer nanocomposite systems.

To meet these challenges, this chapter presents a data-centric approach to accelerate the development of next-generation nanostructured polymers with unprecedented and predictable combinations of properties. The proposed approach integrates physics-based models, empirical data, machine learning approaches, and a robust interphase model built using curated and custom-generated data, within a novel microstructural analysis and optimal design framework. The implementation of this approach is further enhanced by creating an open nanocomposite data resource ("NanoMine"[a]), an integral part of the national effort under the MGI and the Integrated Computational Materials Engineering (ICME) initiative.[12] However, a data resource itself is not sufficient for innovative material design. The data resource must be coupled with newly developed tools, models, and algorithms for data-driven material design. Critically, new models for interphase properties, both physics-based and from machine learning, are needed to create meaningful p–s–p work flows.

In this chapter, we present the architecture of the NanoMine data system and the backbone behind it, an integrated framework for microstructural analysis, and optimal material design. Microstructure analysis plays a key role in assessing p–s–p relationships and in the design of micro- and nanostructured materials systems like polymer nanocomposites. Central to microstructural analysis is the method of MCR, which consists of statistical methods to quantitatively represent the microstructure and its possible inherent randomness (aka characterization) and build ensembles of statistically equivalent microstructures[13] (aka reconstruction). MCR allows one to systematically go beyond the limits where empirical data are available and build forward and inverse p–s–p links through simulation-based analysis and design. While the nanocomposite materials design problem can be formulated as an optimization problem through parametric-based microstructure representation, two challenging research questions remain: (1) how to quantitatively represent a heterogeneous microstructure system using a small set of physically meaningful variables ("microstructure representation"), and (2) how to effectively explore the vast, high-dimensional design space to search for optimal material designs that can be readily synthesized through processing ("design synthesis").

As shown in Figure 1, our proposed microstructural analysis and optimal materials design framework is a multi-phase process in which image analysis is first utilized to analyze physical samples made from existing processes. After a digital representation of filler and matrix materials is obtained

---

[a] NanoMine can be accessed at http://nanomine.org. The site is still under development; the tools will be continuously updated and new capabilities will be added.

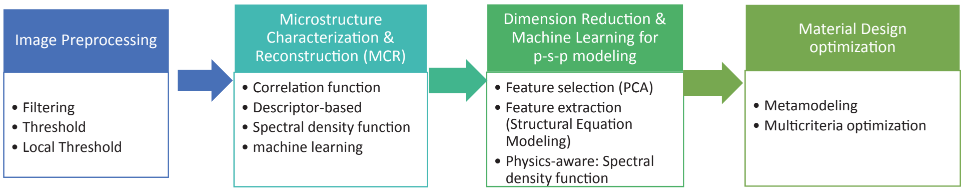


Figure 1: Microstructural analysis and optimal materials design framework.

through "Image Preprocessing", a range of techniques can be considered for MCR. We have developed and implemented a suite of MCR techniques in NanoMine, such as correlation functions,[14] descriptor-based methods,[15,16] and spectral density functions (SDFs). As indicated by details provided in Section 3, there are pros and cons associated with each technique; some of the techniques are either too high-dimensional, prohibitive for 3D digital reconstruction, or not applicable to arbitrary shaped nanofillers and local aggregation. For physically meaningful p–s–p mappings and quick exploration of microstructure designs that include processing feasibility, there is a need for "Dimension Reduction and Machine Learning" to identify the reduced-order representation of microstructures. In this chapter, we will present the use of machine learning[17] and structure equation modeling (SEM) techniques[18] to determine the key microstructure descriptors and processing descriptors in studying the processing–structure and structure–property relationships. In addition, physics-aware dimension-reduction methods, such as the SDF-based approach, are presented as powerful techniques for representing general material systems with high dimensionality and complex, irregular shapes of microstructures. In the last stage of "Material Design Optimization", either the descriptor-based or SDF-based microstructure representation enables efficient reconstructions and allows the use of a parametric optimization approach to search for the optimal microstructure design.[16] In this stage, high-throughput simulation data is used to construct surrogate metamodels for rapid design evaluations, and multi-criteria optimization is utilized to generate a set of materials design solutions for achieving multiple properties.

In the remaining part of this chapter, we first describe the major components of the NanoMine data system and a robust ontology for polymer nanocomposites supporting organization, search, and visualization services of the material data (Section 2). This is followed by an introduction of multiple MCR techniques (Section 3). Dimension reduction techniques for managing the complexity of microstructure representation are presented in Section 4. In Section 5, we provide details of using data mining techniques for constructing descriptor-based p–s relationships based on experimental data. Finite-element-based s–p prediction is presented in Section 6 where a combined physics-based and data-assisted modeling approach is utilized. While the overall model is physics-based, an adaptive optimization approach is used to calibrate the interphase model based on the collected physical data. Finally, in Section 7, a capacitor design problem is used as an example to demonstrate the full integration of p–s and s–p models presented in earlier sections for design of nanodielectric materials using multi-criteria optimization. A model system not typically used for capacitors, but for which we have significant data, polymethylmethacrylate (PMMA)-based nanocomposite with silica nanoparticles, surface modified with a monofunctional chlorosilane, is used as the candidate material system for the case study.

2. NanoMine Data System and Data Resources

During past decades, extensive research efforts have focused on property enhancement of nano-reinforced polymeric materials using both simulation and experimental methods.[19–21] Results from these studies have generated a tremendous amount of data in different forms such as images, plots, and text for a wide range of polymer, particle, and chemistry combinations. However, most of the research data reported in the literature lack a unified data format, terminology and uncertainty measures, and are incomplete. Conventional keyword-based web search engines cannot provide sufficiently detailed and annotated search results for effective material design or even simple exploratory query and comparison of facts. Consequently, it is hard to perform a comprehensive access or search of data according to user-specified criteria, making the design of new functional materials extremely inefficient.

Using the Material Data Curator System (MDCS) developed at the National Institute of Standards and Technology (NIST) and with sponsorship from the National Science Foundation, we have developed a prototype system for nanocomposite material data curation, exploration, and analysis, termed “NanoMine” (www.nanomine.org). NanoMine currently consists of a growing nanocomposite *database*, a collection of module *tools* for statistical learning, MCR, and *simulation software* to model bulk nanopolymer composite material response (see Figure 2). The underlying principle of NanoMine is to create a living, open-source data resource for nanocomposites that provides data archiving and exchange, statistical analysis, and physics-based modeling for property prediction and materials design.

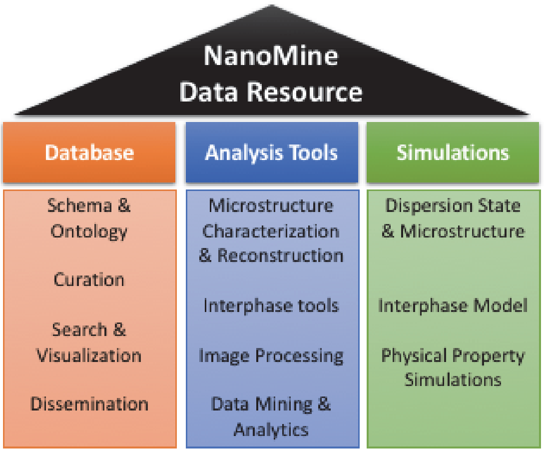


Figure 2: Major features of NanoMine and the key features in each component.

The current NanoMine database contains more than 1,200 samples with extensive information on p–s–p domains manually curated from over 150 papers as well as unpublished lab-generated data. It contains three simulation tools for studying the electrical and mechanical response of composite materials that include explicit representation of the interphase. Additionally, it has four statistical learning and analysis modules including downloadable packages that can be used to pre-process and analyze structure and property data. Continuous efforts have been made to expand the volume of the database and include state-of-the-art microstructure analysis and design tools for the community.

2.1. *NanoMine database and data schema*

The NIST MDCS is an open-source platform providing solutions to collect, share, and transfer material data. The system provides basic functions for data curation (data entry using a web-based system) and data exploration (search data by user specified criterion). MDCS is inherently a No-SQL database that organizes the data using a user-defined data structure or schema.

A well-defined data structure, or schema, is crucial to effectively collect and archive materials data, enable efficient data retrieval, and facilitate data exchange.[22] In order to provide a standard schema to archive the nanocomposite data, the terminology in the schema should be unified and the data types for storing the entities should be well-defined and self-explanatory.[23] In order to develop the customized template to archive

Table 1: Summary of parameters in NanoMine data schema.

Section	Description	Examples
Data source	Metadata of the source of literature.	DOI, author list, journal publication year.
Materials	Characteristics of constituent materials, polymer, particle, and functional groups.	Chemical structure, MW, density, volume fraction, surface treatment, graft density.
Processing	Extraction from Experiment section. Step-by-step description of synthesis/characterization conditions.	Solution processing, melt mixing, polymerization.
Characterization	Measurement equipment and parameters.	SEM/TEM, DMA, DSC, FTIR.
Properties	Function data, value, and observation of properties.	Modulus, dielectric constant, conductivity.
Microstructure	Nanophase dispersion capture in grayscale images. Quantified in morphological descriptors.	MAT file containing grayscale image matrix, descriptors (e.g., filler nearest center distance, equivalent radius).

the raw p–s–p parameters from data sources, 30 representative papers on polymer nanocomposites published within the past decade were investigated to find the most commonly recurring parameters and terminologies. Based on the literature survey, a prototype data template was developed and served as an initial structure to manage all the key parameters associated with p–s–p along with the metadata. The NanoMine data schema is continuously updated to incorporate a wider range of parameters such as additional processing methods or new properties.

As summarized in Table 1, the current NanoMine schema contains the following six major sections:

(1) **Data resource:** The metadata of the source of the literature guided by Dublin core standards which includes the DOI of the cited source, the authors, title, keyword, time, and source of the publication.
(2) **Materials:** Material constituent information, including the filler particle, polymer matrix, and surface treatments. The characteristics of pure matrix and filler such as the polymer chemical structure and molecular weight and the particle density can be entered along with compositions (volume/weight fraction).

(3) **Processing:** Sequential description of chemical syntheses and experimental procedures. The current template provides three major categories: solution processing, melt mixing, and *in situ* polymerization. For each processing step, detailed information such as temperature, pressure, and time can be entered.
(4) **Characterization:** Information on material characterization equipment, methods, and condition used. This information includes details on common microscopic imaging (SEM, TEM), thermal mechanical and electrical measurement, as well as nanoscale spectroscopy.
(5) **Properties:** Measured data of material properties. Properties include mechanical, electrical, thermal, and volumetric properties. The property data could be in the format of a scalar, or in higher dimension such as in 2D spectroscopy or 3D maps.
(6) **Microstructure:** Raw microscopic grayscale images capturing the nanophase dispersion state. Geometric descriptors can also be included to describe the statistical characteristics of the microstructure.

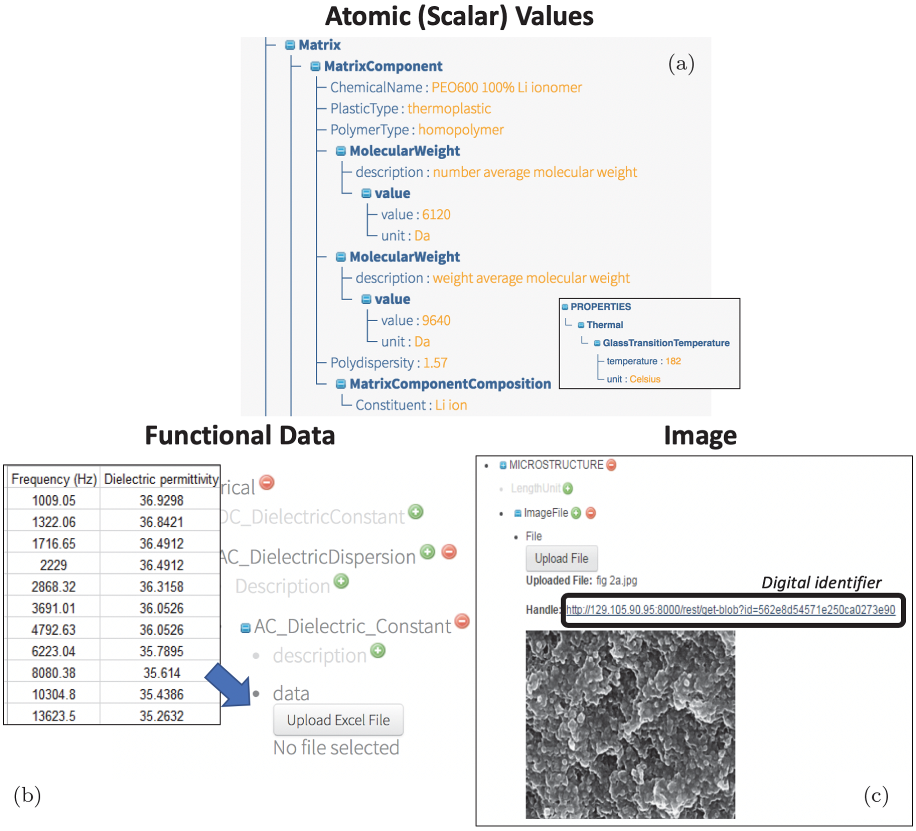
The three major data types used to store each entity are shown in Figure 3.

Using the schema, the non-relational data structure is well-defined and the raw XML document containing the nanocomposite data can be filled in readily, with multiple data formats and dimensionality. The current curation process focuses on nanocomposites with surface-treated spherical inorganic fillers, where many analysis and simulation tools have already been developed internally and the publications contained microscopic images with explicit dispersion information, well-documented processing and characterization methods, as well as clearly plotted property data.

Based on the NanoMine data schema, a robust ontology for polymer nanocomposites has been developed to support organization, search, and visualization services of the material data.[24] This ontology also formalizes relationships inherent in our XML schema and can act as a translator to accept multiple XML formats, enhancing the ability to share across different data resources. On top of the NanoMine ontology, we are building a search and visualization dashboard to allow users to browse and look up the data by using a simple query as shown in Figure 4.

2.2. *Analysis tools in NanoMine*

Apart from the database, NanoMine also provides functionalities for quantitative investigation of the curated data to assist in property prediction and material design. NanoMine aims to provide a practical suite of toolkits



- **Matrix**
  - **MatrixComponent**
    - ChemicalName : PEO600 100% Li ionomer
    - PlasticType : thermoplastic
    - PolymerType : homopolymer
    - **MolecularWeight**
      - description : number average molecular weight
      - **value**
        - value : 6120
        - unit : Da
    - **MolecularWeight**
      - description : weight average molecular weight
      - **value**
        - value : 9640
        - unit : Da
    - Polydispersity : 1.57
    - **MatrixComponentComposition**
      - Constituent : Li ion

- **PROPERTIES**
  - **Thermal**
    - **GlassTransitionTemperature**
      - temperature : 182
      - unit : Celsius

Frequency (Hz)	Dielectric permittivity
1009.05	36.9298
1322.06	36.8421
1716.65	36.4912
2229	36.4912
2868.32	36.3158
3691.01	36.0526
4792.63	36.0526
6223.04	35.7895
8080.38	35.614
10304.8	35.4386
13623.5	35.2632

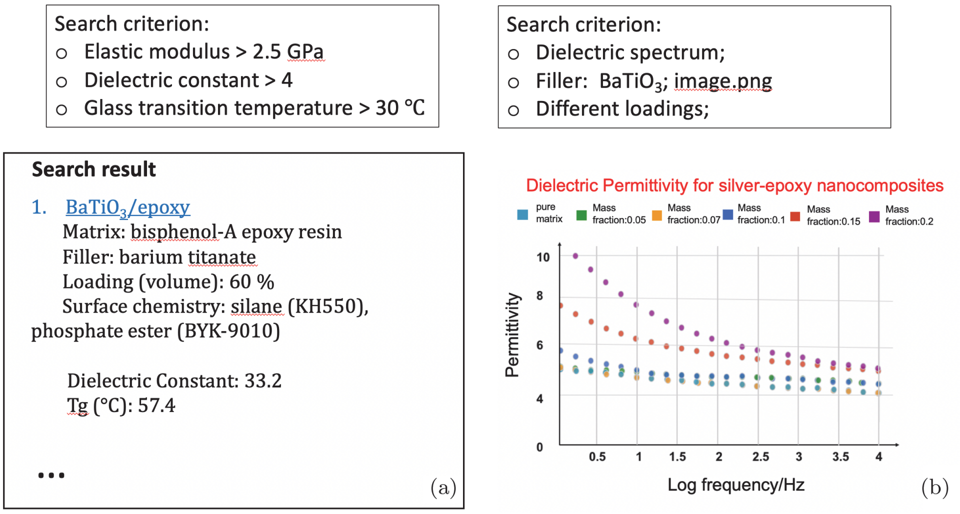
Uploaded File: fig 2a.jpg

Handle: http://129.105.90.95:8000/rest/get-blob?id=562e8d54571e250ca0273e90

Figure 3: NanoMine data types: (a) atomic (scalar) values store single-field numerical quantiles and strings; (b) functional data stores multi-dimensional data, most commonly property plots in tabular forms; (c) microstructure images are extracted and stored as 8-bit grayscale image files.

customized for the curated data. The current tools are open-source web-based modules, and the source code of those tools will be ready for download allowing for user customization. In the beta site, at the current stage, we have implemented four tools taking advantage of our most recent developments in microstructure analysis,[16,25] interphase property calculation,[26] and Prony series analysis.[27] Details of these methods and tools will be presented in Sections 3–5.

The analysis tools have a user-friendly web interface and other features. Figure 5 shows a screenshot of the microstructure reconstruction webtool. The landing page provides detailed instructions guiding the user through the process step by step. References of the algorithm are also included to enable the user to learn the algorithm behind the method. The image



Search criterion:
- ○ Elastic modulus > 2.5 GPa
- ○ Dielectric constant > 4
- ○ Glass transition temperature > 30 °C

Search result

1. BaTiO$_3$/epoxy
   Matrix: bisphenol-A epoxy resin
   Filler: barium titanate
   Loading (volume): 60 %
   Surface chemistry: silane (KH550), phosphate ester (BYK-9010)

   Dielectric Constant: 33.2
   Tg (°C): 57.4

...

(a)

Search criterion:
- ○ Dielectric spectrum;
- ○ Filler: BaTiO$_3$; image.png
- ○ Different loadings;

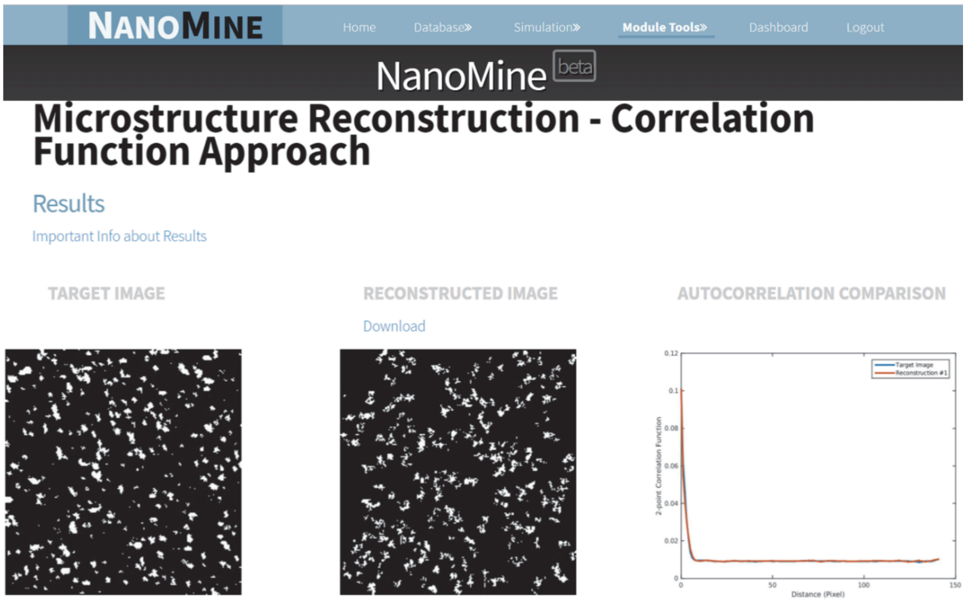
(b)

Figure 4: Examples of searching and visualization of NanoMine data shown in a list (only one entry shown for clarity and space constraints) (a) or a plot (b).

input format, correlation function for characterization/reconstruction, and number of reconstructed images can be chosen by the user. All computations are performed on the NanoMine web server, and an email is sent to the user after their request has been completed. Each user request is associated with a unique "Job ID" which ensures data privacy and can be used to retrieve, and download, a result at any time. Figure 5 shows a screenshot from NanoMine depicting the results of a reconstruction request using the two-point autocorrelation tool.

2.3. *Simulations tools in NanoMine*

NanoMine also includes a set of physics-based continuum models and simulation tools for predicting the macroscopic material response. FEA models have been developed to predict the electrical and viscoelastic response of the nanocomposite with explicit input of microstructure and detailed representations of the interphase. The system currently has implemented two FEA models as web-tools to simulate the viscoelastic[28] and dielectrical properties.[29,30] The available data in the database can be used as input to the FEA model, which then predicts the composite properties. Figure 6 shows the workflow of the simulations tools integrating with the curated data and analysis tools introduced before. The material constituents, including molecular structure, can be obtained from the curated data. The material



Microstructure Reconstruction - Correlation Function Approach

Results

Important Info about Results

Figure 5: Screenshot from the NanoMine platform depicting the results of a reconstruction result from source images.

molecular structures are used to derive the energetic terms that represent the surface energy and filler–matrix interaction.[31] Those mixing energy parameters are then applied to predict the interphase properties and the representative microstructure[30] if micrographs are not available. Taking the input of interphase properties and microstructure dispersion, a 3D FEA model is built with commercial software (COMSOL/Abaqus) using an API and subroutine to calculate the composite dielectric spectra or viscoelastic response. The simulation typically takes 30 minutes for a representative nanodielectric system of 50 clusters in a representative volume element (RVE). Similar to the analysis tools, users are assigned unique Job IDs upon submission of a task. The Job ID can be used to check the status of the job and retrieve the results.

2.4. *Materials design in NanoMine*

The workflow of material design using NanoMine is outlined in Figure 7. Using the platform, the user is able to conduct both property simulation (from left to right) given specific material constituents and processing combinations, as well as material design (from right to left) in order to obtain desired nanocomposite properties. Suppose the user would like to predict

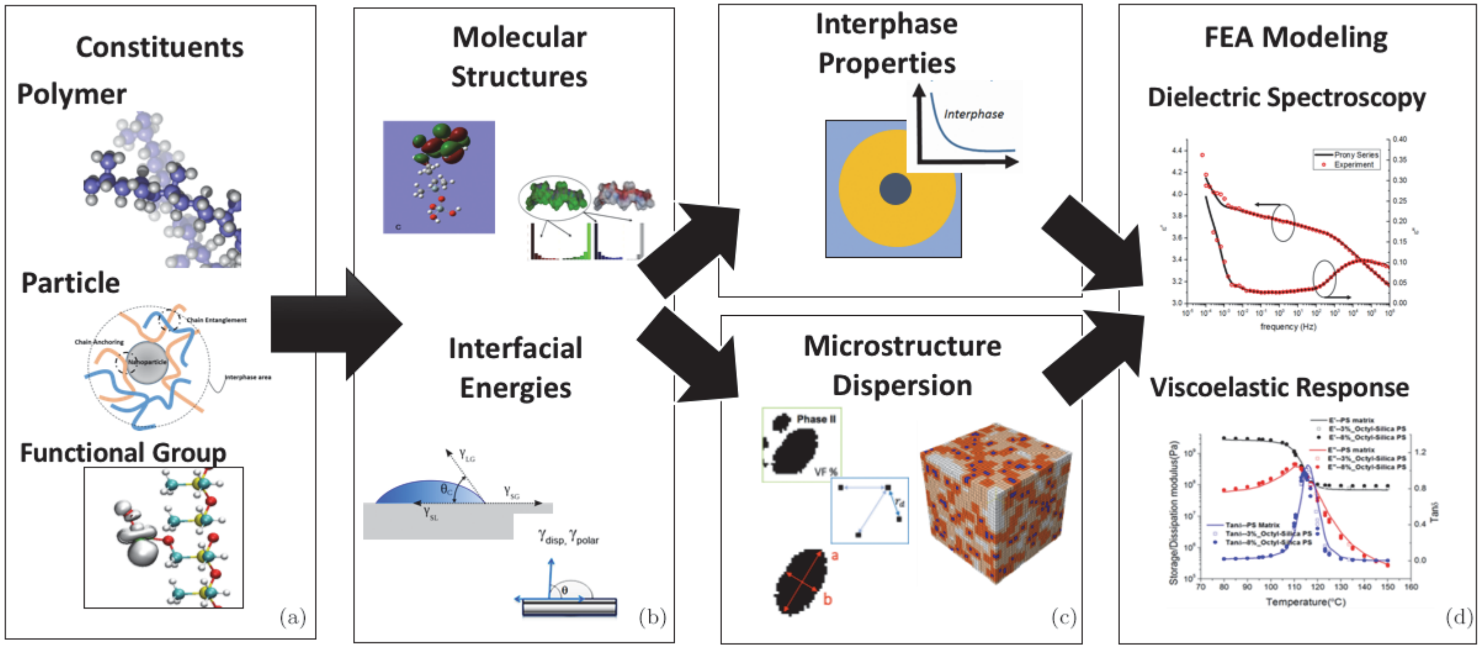


Figure 6: Simulation workflow integrating NanoMine data as material selection to macroscopic property prediction.

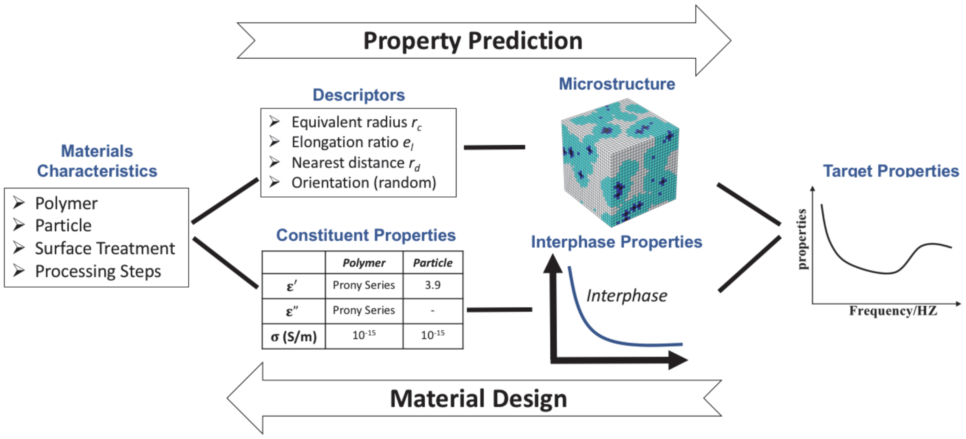


Figure 7: Application of NanoMine on property prediction and material design. Left to right: material property prediction from material selection to bulk composite properties. Right to left: design process starting from target properties to necessary material characteristics.

dielectric permittivity for a PMMA-based nanosilica filled nanocomposite. The first step is to query the database for existing property data for PMMA and nanosilica as material input to the FEA and to calculate the relevant surface energies using the embedded heuristic approach. Then analysis modules are applied to reconstruct the microstructure and predict the interphase properties using statistical correlations among surface energies, processing parameters, structural descriptors, and interphase property descriptors. The FEA model is then carried out to compute the dielectric spectra for this specific sample. This whole process is reversible: the user wants to find a specific material combination and processing steps that could lead to a desired property. If such data have already appeared in the literature and been curated into the database, the solution is simple: just query the database for such a sample and find the material constituents and associated processing steps. However, in most the cases, the target property is not in the database and a user usually wishes to have multiple targets (such as high dielectric permittivity and elastic modulus). NanoMine is being further developed to tackle precisely this challenge. Starting from the existing data, the analysis and simulation modules are coupled with a design optimization algorithm and metamodeling to explore different combinations of constituents, microstructure, and interphase descriptors to attain this goal. A comprehensive predictive framework will be built on top of the entire set of p–s–p data applying a combination of data mining models, optimization methods, and numerical simulations. Details of these techniques are provided in the upcoming sections.

3. Microstructure Characterization and Reconstruction for Polymer Nanocomposites

This section provides details of the MCR techniques for polymer nanocomposites. As a material's morphology significantly influences its properties,[32,33] an essential task in the creation of *process–structure–property* (*p–s–p*) linkages is analyzing the morphology of microstructure(s). The analysis is quantitative, and its outcome is a deep understanding of how processing conditions influence the formation of microstructure and how the microstructure in turn affects the properties. Material morphology can be recorded (in 2 or 3 dimensions) through several imaging techniques — scanning electron microscopy (SEM), transmission electron microscopy (TEM), atomic force microscopy (AFM), computer tomography (CT), etc., each providing morphological information in a distinct perspective and chosen based on the type of information one seeks. In their raw form, these images are stored as an array of pixel intensity values, often containing noise (random variation of brightness/color), and are not of much use. To analyze the image and extract useful morphological features, the following three-step strategy is recommended:

(i) **Image binarization:** Binarization is the process of converting a grayscale image to a black and white image (assuming there are only two phases — filler and matrix) by removing noise and consequently simplifying the analysis. It is accomplished by determining a threshold for pixel intensity of filler and matrix components. As its name suggests, binarized images are essentially an array composed of 0's (displayed in black) and 1's (white). Binarization algorithms are classified as global (single intensity threshold used for entire image) and adaptive thresholding (intensity thresholds varies in different regions of the image). The most widely used image binarization methods are Otsu's method[34] (global) and the Niblack algorithm[35] (adaptive).

(ii) **Microstructure characterization:** Several methods have been developed that can convert multi-dimensional microstructure morphology recorded in images into a set of functions (aka features/descriptors/predictors) that encode significant morphological details, i.e., *characterize* the microstructure. As in the case of imaging techniques, users can choose the characterization method according to the application.

A common observation in all characterization methods is their statistical nature, since most material systems are inherently associated with randomness.

(iii) **Microstructure reconstruction:** After characterization, one can *reconstruct* a statistically equivalent microstructure(s)[13] which embodies a prescribed set of features (obtained by image characterization or provided by user) and can be used as a RVE for simulating material behavior via finite element analysis (which creates the structure–property linkage) or serve as a training dataset for machine learning algorithms.[36,37] Reconstruction is often reduced to an optimization process solved by heuristic methods (due to nonlinearity and high dimensionality), leading to an ensemble of statistically equivalent reconstructions.

The choice of characterization method determines the type of reconstruction method applicable, i.e., the reconstruction algorithm must be *consistent* with the characterization method and vice-versa. It must be noted that although all characterization and reconstruction methods are applicable to any microstructure in general, certain methods are more suited to a material system as compared to others. Hence, users must consider properties of interest, their corresponding length scales, and the availability of computational resources to select an appropriate characterization/reconstruction method. The rest of this section provides an overview of three MCR techniques currently included in NanoMine: correlation functions, spectral density functions, and physical descriptor-based approaches. A brief introduction of emerging machine learning-based MCR techniques is provided at the end of this section. A detailed discussion of these methods and other associated techniques is available in the review article by Bostanabad *et al.*[38]

3.1. *Correlation functions for MCR*

Correlation functions (especially spatial correlation functions), some of the most widely used MCR methods, contain valuable information about the relative positioning of phases in an image. They do so by operating on binarized microstructure images and evaluating the probability distribution of a group of randomly chosen pixels obeying certain “rules”. Although correlation functions can be defined for a group of $\mathbf{n}$ pixels (aka n-point correlation function), it has been noticed that two-point (pixel) correlation functions embody significant details[39–41] and will be used in the following discussion for sake of simplicity. Mathematically, a two-point correlation function $\varphi_{\mathbf{2}}^{\boldsymbol{i}}$ for phase $\boldsymbol{i}$ can be expressed as

$$\varphi_2^i(r_1, r_2) = \langle \mathcal{M}_{r_1} \mathcal{M}_{r_2} \rangle, \tag{1}$$

where $\langle \cdot \rangle$ denotes the expectation operator and $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ represent location vector of points and $\boldsymbol{\mathcal{M}}$ is a microstructure indicator functions such that:

$$\mathcal{M} = \begin{cases} 1 & \text{if } r_1 \in \text{phase } i \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

For an isotropic and stationary microstructure, $\boldsymbol{\varphi_2^i}$ will only depend on the distance $\boldsymbol{r}$ between the two points (making computation efficient[42,43]). As a result

$$\varphi_2^i(r_1, r_2) = \varphi_2^i(|\Delta r_{12}|) = \varphi_2^i(r). \tag{3}$$

Based on the type of "rule" imposed on the two points, the following *three* two-point correlation functions are defined (refer Figure 8):

(i) **Autocorrelation function** ($\boldsymbol{S_2^i(r)}$)[39,40]: The probability that two randomly chosen points are occupied by phase $\boldsymbol{i}$ and is a measure of dispersion. Since many experimental characterization techniques provide structural information in the form of $S_2^i(r)$, it is one of the most widely used correlation functions.

(ii) **Lineal path correlation ($\boldsymbol{L_2^i(r)}$)**[44]: The probability of finding a line segment connecting two randomly chosen points entirely in phase i. $L_2^i(r)$ is a measure of cluster geometry and connectivity of phase $\boldsymbol{i}$ but underestimates these features due to the constraint of measuring along straight lines.

(iii) **Cluster correlation ($\boldsymbol{C_2^i(r)}$)**[32,45,46]: The probability that two randomly chosen points, occupied by phase i, are contained in the same cluster. This correlation contains important information about topological connectivity and is a superior structural signature as compared to $S_2^i(r)$ and $L_2^i(r)$.

For $r = 0$, all three correlation functions reduce to the volume fraction of phase i. Figure 8(d) plots the above three correlation functions as a function of radial distance r (assuming isotropy) for a silica–PMMA sample (Figures 8(b) and 8(c)) with silica volume fraction of 0.85%. The sample was obtained by mixing PMMA and silica (with a chloro-ended silane surface modification) in a twin screw extruder at 200 RPM with 200 kJ/kg energy input. The three correlations functions defined above capture morphological details from different perspectives and are very advantageous when used concurrently.[47,48] It is possible to define additional

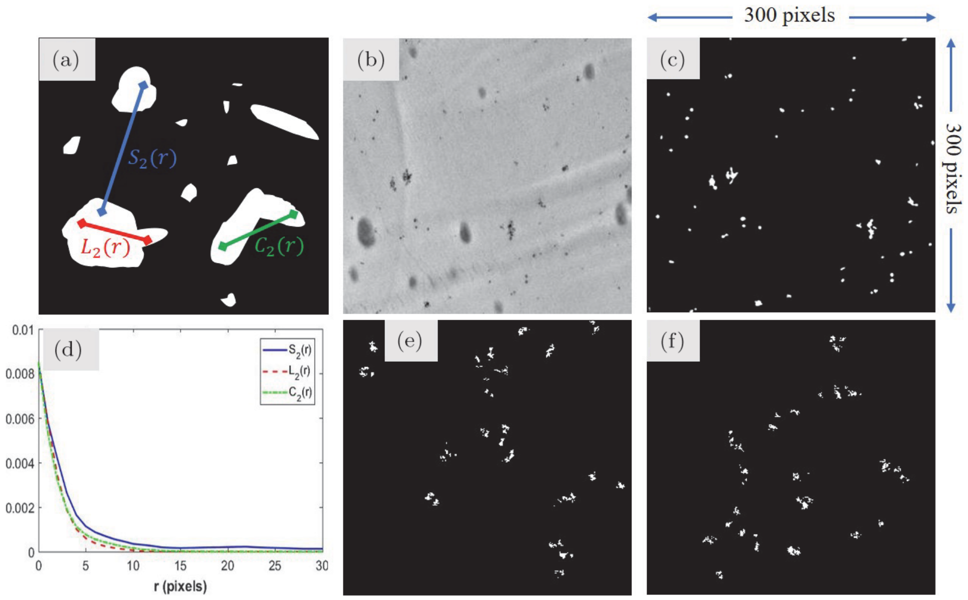


Figure 8: MCR using correlation functions: (a) a schematic representation of two-point correlation functions for white phase; (b) a TEM image of silica-PMMA sample (silica volume fraction ~0.008, 1 pixel = 3.246 nm); (c) image (b) binarized using Niblack algorithm — silica nanoparticles are white and PMMA is black; (d) plot of auto, lineal path, and cluster correlation function for silica; (e) and (f) two reconstructions of image (b) using Yeong–Torquato algorithm performed in NanoMine.

two-point correlation functions as well as higher-order functions such as three-point and four-point correlation functions;[32] however, they require significantly more computational resource.

After characterization, reconstruction of statistically equivalent microstructures can be cast as an optimization problem and tackled in the following way:

(i) Start with a trial microstructure image which has the same volume fraction as the original image. A convenient initial microstructure used widely is a randomly generated white noise image.

(ii) Define an energy (cost) function which measures the difference between the chosen correlation function(s) of the original and trial image. Mathematically, the energy function can be expressed as

$$\mathfrak{I} = \sum_{i=1}^{n} \sum_{r=0}^{m} \alpha_i [\hat{\varphi}_{0,i}(r) - \varphi_{t,i}(r)]^2, \tag{4}$$

where n is total number of correlation functions considered, m is the maximum spatial distance over which the functions are compared, α_i are weights to quantify the importance of each function, and $\hat{\varphi}_{O,i}(r)$ and $\varphi_{t,i}(r)$ are the ith correlation functions of original and trial microstructure images, respectively.

(iii) Adjust the trial microstructure image by swapping pixels belonging to different phases to minimize the energy function.

The energy function ($\mathfrak{J}$)

$$\mathfrak{J} = \sum_{i=1}^{n}\sum_{r=0}^{m} \alpha_i[\hat{\varphi}_{O,i}(r) - \varphi_{t,i}(r)]^2 \tag{5}$$

is almost always highly nonlinear with several local minimums and therefore requires a heuristic optimization method such as simulated annealing (SA)[45,47–49] or a genetic algorithm.[50–52] Yeong and Torquato[47,48] generalized the SA-based method (YT method) for stochastic reconstruction of 2D (images) and 3D (volumes) for any random microstructure. Their method, also centered on swapping two arbitrarily selected pixels of different phases, employs the Metropolis algorithm as the acceptance criterion — the probability of acceptance (P) for a swap is

$$P(\mathfrak{J}_{\text{old}} \to \mathfrak{J}_{\text{new}} = \begin{cases} 1, & \Delta\mathfrak{J} < 0 \\ \exp\left(\frac{-\Delta\mathfrak{J}}{T}\right), & \Delta\mathfrak{J} \geq 0, \end{cases} \tag{6}$$

where $\Delta\mathfrak{J} = \mathfrak{J}_{\text{new}} - \mathfrak{J}_{\text{old}}$ and T is temperature which decreases with each iteration (like annealing of metals and hence the name simulated annealing) and controls the acceptance probability. Figures 8(e) and 8(f) show two statistically equivalent reconstructions of silica–PMMA sample (Figure 8(c)) using the YT method with the autocorrelation ($S_2^i(r)$) of silica nanoparticles used to construct the energy function. The YT method, in its original form, is computationally expensive (prohibiting 3D reconstruction). In addition, the parameters in the correlation functions have no clear physical meaning nor can they be mapped to the processing conditions easily. Hence, the correlation function representation is not effective or computationally feasible for materials design.

3.2. *Physical descriptors-based MCR*

Physical descriptors (aka features/predictors) provide a meaningful and convenient approach for direct elucidation of p–s–p relationships. Descriptors are important structural parameters that are highly related to material properties and provide a reduced dimensional representation of the microstructure.[16] Since several key descriptors (often correlated) exist for each property, the aim of this approach is to find a small set of uncorrelated descriptors that sufficiently characterizes the material system of interest by eliminating (or at least minimizing) the sources of uncertainty in constructing p–s–p relationships.

Extracting descriptors from a microstructure image involves application of image segmentation techniques[53] to identify clusters of filler material followed by analysis of individual clusters. Since there are multiple descriptors available, the analysis leads to either: (i) extraction of a finite set of preselected descriptors (based on experience or design of experiments[54–56]) or (ii) extracting many descriptors yet selecting only a subset of uncorrelated and informative descriptors for building predictive models. Feature selection has been studied extensively[57] and implemented successfully in the creation of insightful p–s–p mappings. Xu *et al.*[17] studied the damping performance of polymer nanocomposites by using a four-step descriptor selection method using pairwise correlation analysis and machine learning-based RReliefF variable ranking.[58] The RReliefF algorithm and the descriptor-based approach are used in Section 5 to establish a relationship between dispersion of nanocomposites and processing conditions. Techniques for reducing the dimension of descriptors will be introduced in Section 4.

Descriptors can be categorized in the following ways:

(i) **Information scale:** Descriptors are hierarchical in terms of the information scales they represent. For example, volume fraction represents the highest scale of information for a microstructure in terms of composition, followed by nearest neighbor distance (which quantifies the relative distribution of filler clusters), while aspect ratio is at the lowest scale since it is associated with an individual cluster. A crude yet convenient definition is that higher scale descriptors are assigned to microstructure while lower scale descriptors are associated with an individual cluster.

(ii) **Nature of descriptor:** A descriptor can be *deterministic* such as volume fraction or *statistical* like aspect ratio/nearest neighbor distance. A single value is sufficient to quantify a deterministic descriptor, while a statistical descriptor requires a cumulative distribution function for representation.

Table 2 lists key descriptors extracted from Figure 8(c) and used for reconstructing a 3D RVE shown in Figure 8(a). These descriptors can be obtained using region analysis algorithms.[59]

Table 2: Key descriptors extracted from silica — PMMA sample shown in Figure 8(c).

		Value	
Description	Type	Mean	Std. Deviation
Volume fraction	Composition (Deterministic)	0.008	—
Cluster's nearest center distance	Dispersion (Statistical)	64.114 nm	43.005 nm
Number of clusters	Dispersion (Deterministic)	53	—
Aspect ratio	Geometry (Statistical)	1.468	0.527

Reconstruction follows a hierarchical procedure,[15–17,31,60,61] with descriptors at higher scales being considered first. Each descriptor may require a different method for reconstruction but usually involves optimization. Xu *et al.*[15,16] developed a four-step strategy for 3D reconstruction of isotropic polymer nanocomposites with ellipsoidal filler clusters. Given a 2D grayscale or binarized microstructure image, the procedure involves: (a) extracting descriptors (characterization) using image processing techniques and estimating their values for a 3D volume; (b) matching dispersion descriptor (nearest neighbor distance) using SA algorithm (discussed in Section 3.1) by moving cluster centroids; (c) constructing individual filler clusters using geometric descriptors obtained in step (a) and placing them at locations derived in step (b) (assuming random orientation); and (d) adjusting aggregate locations to eliminate overlap (if necessary) and compensating so that the volume fraction matches the original 2D image. Figure 9 shows the application of the above algorithm to the silica–PMMA sample shown in Figure 8(c) using descriptors in Table 2. The scale of reconstructed volume (Figure 9(a)) is the same as that of the original microstructure; i.e., edge length of a voxel is equal to edge length of a pixel. Figure 9(c) shows good agreement between the $S_2(r)$ for the target image and mean of 2D slices taken from the reconstructed volume (Figure 9(a)), thus validating the accuracy of the reconstruction technique. The significance of these descriptors (and some others) in developing p–s–p mappings will be illustrated in the subsequent sections. The reconstructed statistically equivalent 3D microstructures are RVEs and can be utilized for property prediction using, for example, finite element analysis.

Unlike correlation functions, which characterize microstructure from a probabilistic perspective and cannot be easily related to morphological features, descriptors offer a straightforward approach to material design. Often, descriptors are quantities that can be controlled by adjusting processing

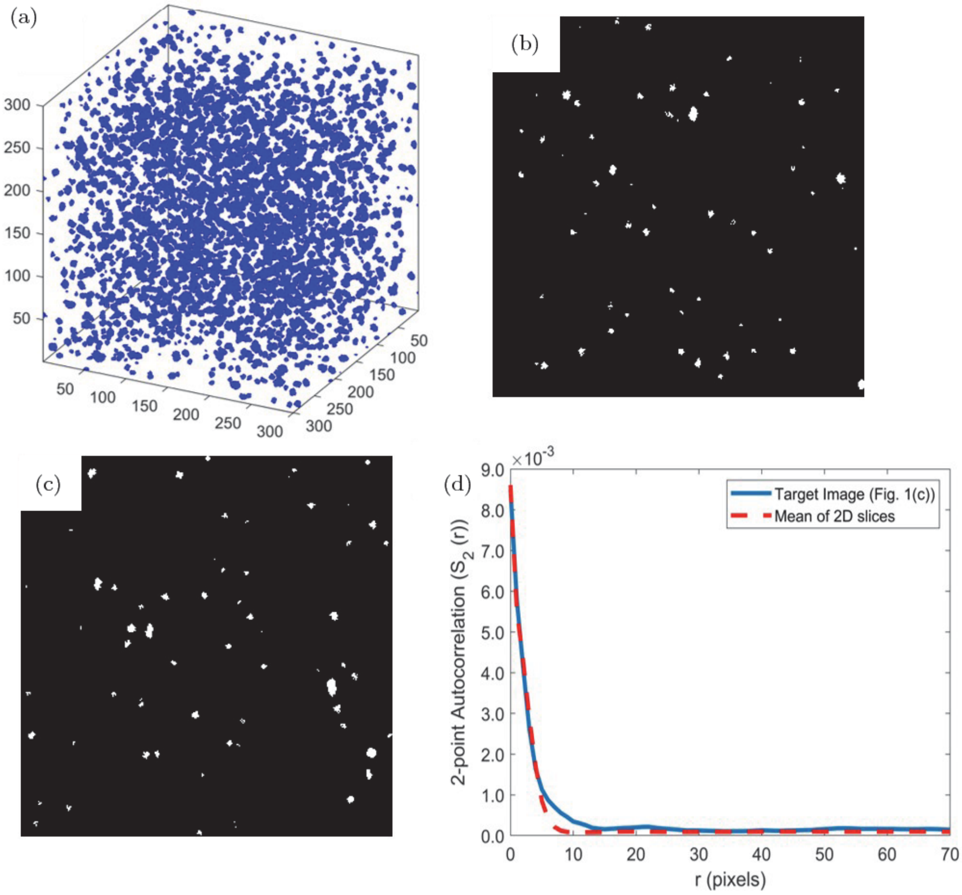


Figure 9: Microstructure reconstruction using descriptors in Table 1: (a) a 3D RVE ($300 \times 300 \times 300$ voxels) representing microstructure in Figure 1(c) generated using the algorithm developed Xu *et al.*[15] blue phase represents silica clusters while the PMMA matrix occupies the rest of the RVE; (b) and (c) two 2D slices taken from RVE shown in (a); (d) Comparison of $S_2(r)$ of target microstructure (Figure 1(c)) and mean of 2D slices taken from 3D RVE.

conditions and they also have a different impact on properties. Several investigations in property enhancement of composites have found descriptors such as volume fraction, size, shape and dispersion to play an important role.[15–17,31,61–67] For example, Karasek *et al.* observed that large variation in the size of filler aggregates leads to enhanced electrical conductivity of carbon black (filler)–rubber (matrix) composites.[66] Finally, the descriptor-based approach allows the use of parametric optimization algorithms to search the optimal microstructure design that meets the targeted properties.

3.3. *Spectral density function for MCR*

The SDF (aka, Fourier power spectrum) is a low-dimensional representation of microstructure in the frequency domain where different frequencies represent real space features at different length scales. It can be evaluated simply as the squared magnitude of the Fourier transform (FT) of a binary microstructure image $\mathcal{M}$:

$$\rho(k) = |\mathcal{F}[\mathcal{M}]|^2, \tag{7}$$

where $\mathcal{F}[.]$ denotes the Fourier transform operator and k is the frequency vector. Figure 10 depicts three isotropic, quasi-random channel-type nanostructures with ring-shaped SDF. Channel-type nanostructures originate from bottom-up processes such as phase separation[68] or thin film wrinkling.[69] Figure 10(a) contains a single dominant frequency, i.e., a single ring, and manifests in channels with uniform width and connectivity. The channel width is inversely proportional to ring radius. Figures 10(b) and 10(c) have additional rings at lower frequencies leading to wider channels with variations in channel width and increased disorder in nanostructure. Note that the type of nanostructure (and the form of SDF) is dependent on fabrication methods and materials used.

For isotropic microstructure, radial averaging can be used to convert vector k to a scalar (like radial averaging of position vector for correlation functions). According to the Winner–Khinchin Theorem,[70] the inverse FT of SDF is the two-point autocorrelation function. Previous research suggests that SDF is sufficient to represent some complex heterogeneous microstructures with irregular geometries. Studies have also shown that SDF is a physics-aware MCR technique that can map the SDF parameters to

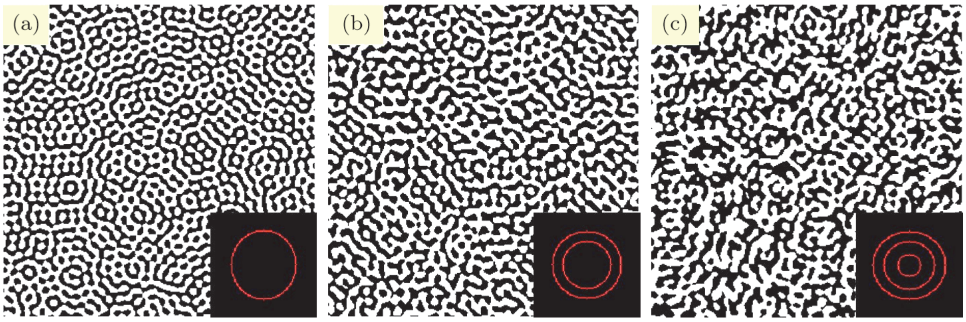


Figure 10: Three quasi-random nanostructures and their corresponding SDF (shown in insets).

properties which largely depend on the spatial correlations of microstructures, for example optical properties.[71] In our recent work, we developed an SDF-based approach to bridge the gap between structure–performance for organic photovoltaic cells[72] and p–s–p relationships in design of light-trapping nanostructures made for thin-film solar cells.[73,74] Research shows that SDF provides sufficient representation of quasi-random microstructures made from bottom-up manufacturing processes such as nanoparticle self-assembly and nanowrinkling.[75–77]

The SDF approach significantly reduces the dimensionality of microstructure representation, as a few parameters (2–5 normally) are sufficient to represent a unique microstructure morphology. Considering our design case study using PMMA–silica nanocomposite as an example, from the samples collected it is found that the gamma distribution, governed by two parameters k and θ, is sufficient to model the nanocomposite. As illustrated in Figure 11(a), a narrow SDF represents particles that are clustered and unevenly distributed in real space. In contrast, with a broader SDF, structure Figure 11(c) has a well-dispersed pattern.

All examples in Figure 11 are reconstructions that can be generated in less than a second using the Gaussian random field (GRF) method,[78] and they accurately reproduce the irregular shapes and local aggregation in real material systems as a result of the stochasticity and the embedded local correlations in GRF. In addition, a disk packing algorithm has been developed for microstructures with disk-shaped filler clusters.[79]

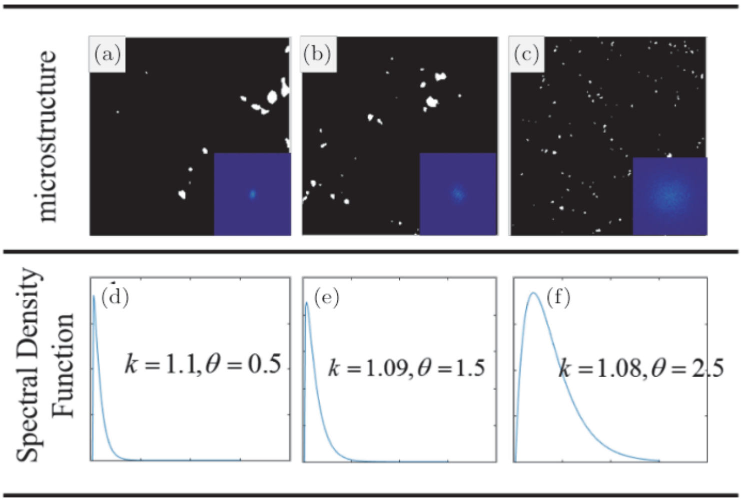


Figure 11: SDF approach for unevenly distributed fillers and local aggregations.

In the design case study shown in Section 7, the SDF representation is used to rapidly generate samples of microstructures and create metamodels of s–p relationships for a wide range of microstructural designs.

3.4. *Machine learning-based MCR techniques*

Though not implemented yet in NanoMine, a few state-of-the-art machine learning methods for MCR are briefly introduced here. Their ability to model highly nonlinear functions with nominal user input, coupled with speed and flexibility make them an attractive proposition. These methods can be categorized as follows:

(a) **Supervised learning**[80,81]: Using a single 2D/3D microstructure sample as a training dataset, a decision tree is used to learn the conditional probability distribution of any individual pixel's value given its surrounding pixel phases. Reconstruction consists of taking a realization from the learned model, and rendering the method cost-effective and applicable to a wide range of material systems.

(b) **Instance-based learning**[37,82]: This approach uses a large database of 2D/3D microstructures to search for an instance that is most similar (using predefined similarity metric) to the microstructure under consideration. Thus, the reconstruction procedure is essentially a rigorous search through the database.

(c) **Deep learning**[83–85]: This technique employs hierarchies of stacked neural network layers that encode implicit features into hyper-dimensional spaces through linear multiplications and nonlinear transformations.

The above techniques have shown success in characterization and reconstruction of polymer nanocomposite systems. However, the potential of these machine learning techniques for materials design is still being explored. One challenge is the lack of explicit description of microstructure design variables and their physical relationships with processing conditions.

4. Dimension Reduction and Machine Learning of P–S–P Relationships

To manage the dimensionality of microstructure design representation, data mining and supervised learning techniques have been employed for discovering important microstructure features and determining microstructure design variables. Recent work has achieved microstructure dimensionality reduction via manifold learning[37] and kernel principal components.[86]

However, unsupervised learning methods that rely on image data only do not capture the impact of microstructure on material properties, so that the reduced parameter set may not be suitable for the purpose of material design. Limited efforts have been made towards modeling the microstructure–property relationship using statistical learning and further reducing the high dimensionality of microstructure features obtained from microscopic images.

There are two common types of dimension reduction techniques in machine learning: *feature selection* and *feature extraction*. “Feature selection” reduces the number of variables in a system by selecting a subset of relevant features, while “feature extraction” projects the original high-dimensional feature space into a reduced space. Both selection and extraction can be either supervised or unsupervised. The projection incurred in extraction methods usually refers to a linear or nonlinear transformation of the original variables. Extraction methods are not suitable when the physical meaning of the original variables needs to be preserved. Feature selection, on the other hand, chooses a subset of more informative features from the original set and fits many design requirements. A supervised ranking method based on the Relief algorithm[17] has been developed in our earlier work to rank the importance of microstructure descriptors. However, the method does not address the redundancy among descriptors and it is also quite subjective in determining how many descriptors to keep from a ranked list.

We present here an structural equation modeling-based approach[18] that combines feature selection and feature extraction techniques for uncovering latent microstructure features. SEM is a multi-variate data analysis method often used in social science for problems with latent layers and path structures.[87] We view microstructures as observations of underlying structural characteristics and apply a measurement model in SEM to represent the relationship as shown in Figure 12. By introducing latent layers (structural features) in mapping input and output relationships, we are able to identify the relationships and dependencies among different microstructure descriptors. The whole procedure consists of two main parts. First, exploratory factor analysis (EFA)[88] is used to reduce the number of descriptors as a part of feature selection. Second, for feature extraction, the original descriptors are grouped under a few latent factors, and each latent factor is linked to a set of descriptors (called indicators). The extracted latent factors can be considered categories of microstructure features, and the grouped structure reflects the correlation patterns of descriptors. Depending on data availability, responses in an SEM structure may include

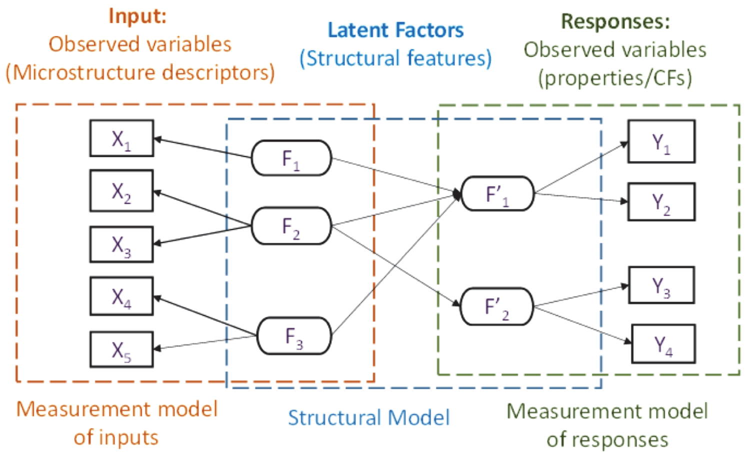


Figure 12: SEM-based learning (X and Y represent observed inputs and outputs, F and F' indicate latent factors associated with X and Y, respectively).

a microstructure correlation function (CF) or material properties, which are used to identify the underlying descriptor–CF or descriptor–property relationship. With the identified structure, the partial least square (PLS) technique[89] is employed to estimate the coefficients in SEM modeling; the coefficients represent the influences of descriptors.

By building an SEM model, we are able to deal with high correlations among all candidate descriptors, gain more insight into their relationships, and identify latent factors (e.g., under categories of "composition", "geometry", and "dispersion") for categorizing microstructure features. In Ref. [18], for epoxy-silica microstructures, four descriptors, *volume fraction, cluster size, nearest neighbor distance*, and *cluster roundness*, are found to be significant through the SEM analysis using CF as the response (image data only). The sufficiency of these descriptors is validated through confirmation of the correlation function of the reconstructed images using the reduced descriptors versus the original one.

Once the reduced dimensionality is determined, machine learning tools are widely used for establishing metamodels of s–p relationships or p–s relationships using available data. Figure 13 illustrates the metamodels created between microstructure key descriptors and permittivity for an epoxy-silica system[18] based on finite element simulations. As shown, smaller cluster size and larger volume fraction of silica lead to a better dielectric performance: higher energy storage capability (high ε') and smaller dielectric loss (small $\tan\delta$) of the epoxy-silica system. This observation is consistent with the findings in the literature that systems with small particles have high

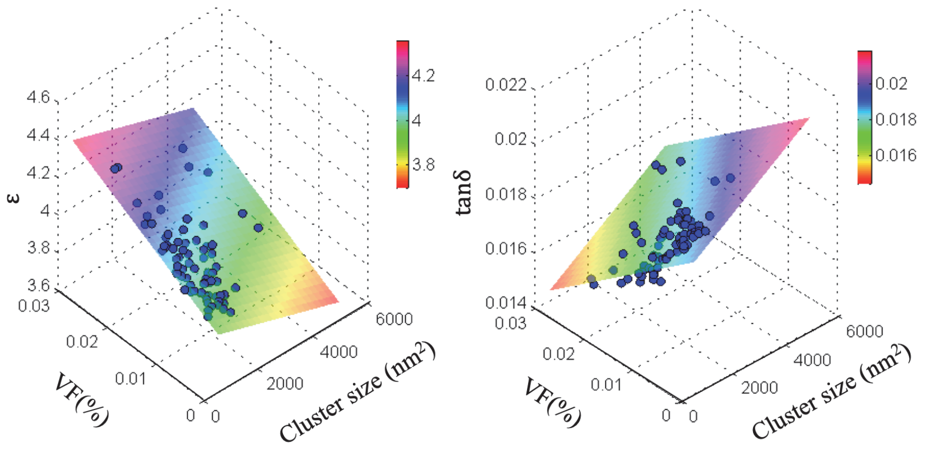


Figure 13: Structure–property relationship. Linear relationship between dielectric properties and key descriptors, only volume fraction and cluster size are considered for visualization.

surface area-to-volume ratio, which is critical in determining the properties of nanofilled materials.[90] In general cases, the s–p relationship can be more complicated so we may need to fit nonlinear models, such as Gaussian process (GP) models.

5. Descriptor-based Processing–Structure Modeling of PMMA–Silica System

To date, polymer nanocomposites have demonstrated outstanding properties and played a significant role in scientific discoveries.[21,63,91–96] However, their commercial use is very limited due to the difficulty of processing such materials on a commercial scale with control over the morphology. More specifically, because a quantitative p–s–p relationship is lacking, to obtain the optimized material property,[61,97–99] one has to tailor nanoparticle dispersion by controlling the processing conditions in a trial-and-error manner.[100,101] It is crucial to develop a quantitative modeling approach which can incorporate the particle/surface chemistry and processing required to achieve a specific nanofiller dispersion. Such a quantitative model will provide more in-depth understanding of nanocomposites and greatly accelerate the design and optimization of advanced materials. In this section, we present a descriptor-based approach for creating p–s relationships under non-equilibrium processing conditions. Three separate steps are involved: processing quantification, microstructure characterization, and processing–structure relationship development.

AQ: Can we change "processing-structure" to "p-s". Please confirm.

5.1. *Quantification of processing conditions*

Processing is usually complex. Typically, for each individual processing technique, there are many processing conditions and settings to be tweaked to achieve optimized material properties. Since some of the processing parameters are correlated and some do not have significant impact in microstructure dispersion, it is of great importance to select the proper set of parameters to represent the most significant aspects of processing that contributes to the dispersion. To this end, we extend the representative parameters for equilibrium conditions to the ones for non-equilibrium systems by adding a processing energy descriptor.

Prior work[100–102] in developing processing–structure relationships for nanocomposites primarily considered two different aspects. (1) The first aspect is how the surface chemistry of nanoparticles controls filler dispersion. For instance, Natarajan *et al.*[102] developed a quantitative relationship between interfacial energy and dispersion. The filler is found to be well-dispersed when the filler and the polymer matrix are thermodynamically compatible. Agglomeration increases when the work of adhesion between the fillers exceeds the work of adhesion between the filler and the polymer. Villmow *et al.*[100] found that the surface energy also determines the mobility of the interphase, which significantly contributes to material properties such as glass transition temperature. (2) In contrast, the second aspect of research studies how processing energy (e.g., specific processing energy input) affects the filler dispersion. For instance, Kasaliwal *et al.*[101] discovered a power law rule for the dependence of the dispersion of CNT agglomerates on the specific processing energy applied.

While the aforementioned studies provide either qualitative or quantitative results for developing p–s relationships, they consider only the surface energy or the processing energy independently. For non-equilibrium processing methods, we consider the two processing energy parameters simultaneously: *interphase energy* and *processing energy*. Before we formally illustrate the formation of these two energy parameters, we confine the factors that potentially contribute to dispersion. Generally, non-equilibrium processing covers a wide range of processing techniques that result in kinetically trapped microstructures. The most popular one in industrial applications is extrusion because it is inexpensive, fast, and simple. Prior qualitative study has found that to reduce the nanoparticle agglomerate size, the agglomerate cohesive strength must be overcome. With increasing shear energy input, the agglomerate size can be further reduced. Several components would be involved in the processing of particle deagglomeration[101,103–107]:

1. Incorporation of the filler into the matrix.
2. Wetting of the filler with matrix material.
3. Infiltration of the matrix into the agglomerate.
4. Breaking up of the agglomerates and erosion of nanoparticles from the agglomerate surface.
5. Distribution within the matrix.
6. Reagglomeration due to particle collisions during mixing.

These processes are dependent on nine factors:

1. Surface energies of the components;
2. Viscosity of the polymer;
3. Packing density of the agglomerate;
4. Chain stiffness of the polymer;
5. Shear stress;
6. Specific energy input during processing;
7. Agglomerate size;
8. Crystallinity;
9. Agglomerate strength.

Extensive research[101,108–113] has been carried out on the quantitative dependence of some of the listed processes and factors. We encourage readers to refer to this literature for detailed explanations of the dependencies. In quantifying the processing conditions in extrusion, we consider four out of the nine factors: surface energy, polymer viscosity, shear stress, and specific processing energy input (factors 1, 2, 5 and 6). The other factors are ignored in this study either because of the limited number of composite systems in the study (e.g., factors 4 and 8) or difficulty in gaining the needed information during the applied process (e.g., for studying factors 3 and 5). For the demonstration example in this section, we narrow our discussion to the simplest extrusion method, single-screw extrusion as used in Ref. [25] and the corresponding dataset in Refs. [25 and 114]. Data collected from experiments has been curated and stored in NanoMine.

We consider silica nanoparticles (15 nm diameter) as a filler that is surface modified using three monofunctional silanes by the method elucidated by Natarajan *et al.*[102] The silanes were differentiated by the end group of the molecule — octyldimethylmethoxysilane, chloropropyldimethylethoxysilane, and aminopropyldimethylethoxysilane. The polymer matrices of polystyrene, polypropylene, and PMMA were used in powder form for the extrusion process. Each combination of surface modification and polymer matrix provides a range of interfacial energy interactions.[115] Functionalized particles were precipitated using an antisolvent, mixed with the polymer matrix powder, and dried to produce a particle–polymer powder mixture, as explained by Hassinger *et al.*[25]

Prior to extrusion, the polymer–particle powders were jet milled to reduce the size of powder particles. The powders were fed into a single screw extruder, with screw diameter of 12.7 mm, screw length of 342.9 mm, and channel width of 9.8 mm. Extrusion was carried out at 180°C and at varying screw speeds (20, 100, and 195 RPM) to examine the influence of processing parameters on the final dispersion state of the particles. More detailed descriptions of the processing can be found in Ref. [25].

5.2. *Interfacial energy descriptor*

The final dispersion state depends on deagglomeration and reagglomeration of the nanoparticles during processing. The simulation by Starr *et al.*[116] shows that when the particle–polymer interaction is weaker than the particle–particle interaction, the particles agglomerate abruptly. Herein, we adopt Natarjan *et al.*[102] and Khoshkava *et al.*'s[117] quantification of these interactions — the ratio of the work of adhesion between filler and polymer and the work of adhesion of filler to filler (denoted as $W_{\mathrm{PF}}/W_{\mathrm{FF}}$) to represent the interfacial interactions. Table 3 lists the values of the interfacial energy descriptor, $W_{\mathrm{PF}}/W_{\mathrm{FF}}$, for all the composites in the dataset. The cells of the compatible combinations are marked in gray.

5.3. *Processing energy descriptor*

The other descriptor that we introduce is the processing energy descriptor, which essentially measures the energy consumption within the screw during

Table 3: Descriptors describing the interfacial energy of the various material combinations.

Silica modification	Polymer	PP	PS	PMMA
Octyl-mod-silica	$W_{\mathrm{PF}}/W_{\mathrm{FF}}$	0.94	1.15	1.12
Chloro-mod-silica	$W_{\mathrm{PF}}/W_{\mathrm{FF}}$	0.84	1.04	1.05
Amino-mod-silica	$W_{\mathrm{PF}}/W_{\mathrm{FF}}$	0.78	0.95	0.96

Note: The compatible combinations are given in light grey color.

processing. We denote the screw channel depth of the extruder as $H(L)$, in which H is dependent on the length of the screw L. The screw diameter is denoted as d, while the screw speed is represented by N. Then the shear rate could be computed by,[118]

$$\dot{\gamma} = \frac{\pi(d - 2H(L))N}{H(L)}. \tag{8}$$

Given the viscosity η_P, the shear stress is

$$\tau = \eta_P \cdot \dot{\gamma}. \tag{9}$$

The viscosity of the nanocomposite can be estimated by first determining the viscosity of the neat polymer using the Cross Law as shown in Equation (10) and then using the Einstein equation for filled polymers as shown in Equation (11). In these equations, η_P and η_F are the viscosities of the neat polymer and filled polymer, respectively, $\eta_{P,\,lim}$ is the viscosity at infinite shear rate for the neat polymer, $\eta_{P,0}$ is the viscosity at zero shear rate for the neat polymer, α is a fitting-factor, and f is the filler fraction, where the viscosity of the materials could be estimated[119]:

$$\eta_P = \eta_{\text{lim}} + \frac{(\eta_{P,0} - \eta_{P,\text{lim}})}{1 - \alpha \cdot \dot{Y}^{2/3}}, \tag{10}$$

$$\eta_F = \eta_P + f \cdot 2.5 + f \cdot 21.4^2. \tag{11}$$

Using Lai's[120] theoretical model, the processing energy consumption in a circular segment with infinitesimal length along the screw length direction can be calculated as

$$dw = \frac{\pi D \Omega}{60} dF_{by}, \tag{12}$$

where D is the screw diameter, Ω is the screw speed, and dF_{by} is the tangent component of the traction on the screw barrel surface ($F_{by} = S_b \mu_b \dot{\gamma_b} cos\varphi$), in which S_b is the area of traction on the screw barrel, μ_b is the viscosity of the molten polymer and φ is the angle between the resultant traction and the channels. By taking the integral of dw along the length of the screw, the total energy consumption, w, is obtained. The processing energy descriptor, namely the energy consumption per mass unit of the throughput, can be obtained by

$$E_\gamma = \frac{w}{\dot{q}_m}, \tag{13}$$

where $\dot{q}_m$ is the throughput.

5.4. *Microstructure descriptors from characterization*

In addition to quantifying the processing conditions, it is also of great importance to characterize microstructures with a set of low-dimensional parameters. A necessary step before microstructure characterization is material phase segmentation. For filler–matrix composites, it is essentially a binarization process that distinguishes fillers from matrix. While the binarization of TEM images is typically done by setting a global threshold,[34,121] we find it does not work well with the TEM images in our dataset. The major problem is that, in our TEM images, local shades or unevenness would result in a darker/lighter spot in some particular locations, and these spots would be misclassified when a global thresholding algorithm is applied. Therefore, we apply the Niblack algorithm,[35] which is a sliding window algorithm that takes advantage of local pixel statistics to determine the local thresholding value. Figure 14 demonstrates how the Niblack algorithm outperforms the global thresholding algorithm, which is clearly seen in the red boxed area where particles in a shaded region of the original TEM image are better identified using the Niblack method.

While a range of MCR techniques are available as introduced in Section 3,[38] the descriptor-based characterization approach[15] is used as most of the clusters are spherical. Per Xu *et al.*,[15] the descriptors considered are listed in Table 4.

While descriptor-based characterization provides a rich set of physical descriptors that represent the microstructures being studied, the dimensionality is still too high to be correlated to the two processing descriptors discussed above. Therefore, a supervised learning-based descriptor selection


(a) (b) (c)

Figure 14: A demonstration of a sample TEM image with local shades/unevenness (a) and a comparison between the global thresholding algorithm (b) and Niblack algorithm[35] (c).

Table 4: Descriptors being characterized in microstructure characterization.

	Descriptor	Definition	Type	Number moments
Category 1: Composition	VF	Volume fraction	Deterministic	1
Category 2: Dispersion	r_{ncd}	Cluster's nearest centroid distance	Statistical	4
	r_{nbd}	Cluster's nearest boundary distance	Statistical	4
	θ	Principle axis orientation angle	Statistical	4
	I_{filler}	Surface area of filler phase	Deterministic	1
	N	Number of clusters	Deterministic	1
	Local VF	Local Volume fraction of the Voronoi cell	Statistical	4
Category 3: Geometry	r_c	Cluster's equivalent radius	Statistical	4
	r_p	Inscribed circle radius	Statistical	4
	r_{aw}	Area-weighted equivalent radius ($r_{aw} = r_c/A$)	Statistical	4
	A	Cluster area	Statistical	4
	δ_{cmp}	Cluster's compactness	Statistical	4
	δ_{rnd}	Cluster's roundness	Statistical	4
	δ_{ecc}	Cluster's eccentricity	Statistical	4
	δ_{asp}	Cluster's aspect ratio	Statistical	4
	δ_{rect}	Cluster's rectangularity	Statistical	4
	δ_{ttst}	Cluster's tortuosity	Statistical	4

Table 5: Significance of descriptors learned using RReliefF algorithm.

Rank	Descriptor	Significance	Rank	Descriptor	Significance
1	I_{filler}	0.0505	6	A_1	0.0360
2	r_{aw1}	0.0500	7	r_{c2}	0.0345
3	VF	0.0500	8	N	0.0328
4	r_{aw2}	0.0432	9	r_{p1}	0.0324
5	r_{c1}	0.0370	10	$LocalVF_1$	0.0321

process is conducted. Specifically, similar to Xu *et al.*,[17] the RReliefF algorithm is applied to assign significance scores to each descriptor with respect to their impact on the microstructure correlation function.

For each category of descriptors (composition, dispersion, and geometry), the most significant one is selected, i.e., volume fraction (VF), surface area of filler phase (I_{filler}), and area weighted equivalent radius (r_{aw1}), respectively. We also find that these three descriptors are linearly correlated. Therefore, we further condense them into one integrated descriptor — volume fraction normalized filler surface area ($\bar{I}_{\text{filler}} = I_{\text{filler}}/VF$).

5.5. *Building processing–structure relationships*

Table 6 summarizes the values of the processing descriptors and microstructure descriptors presented above. The impact of each processing descriptor on the microstructure descriptor is first studied. Then a predictive p–s model is established by mapping the two sets of descriptors.

First, consider the impact of interfacial energy on the microstructure descriptor. Recall that a larger value of $\bar{I}_{\text{filler}}$ indicates better dispersion. The materials with the best compatibility (e.g., highest values of $W_{\text{PF}}/W_{\text{FF}}$) show the best dispersion (octyl-modified silica and PS) (Figure 15) as indicated by the largest normalized interface area.

The microstructure dispersion descriptor $\bar{I}_{\text{filler}}$ also depends on the processing energy descriptor E_γ (Figure 16). In Figure 16, samples with the same type of polymer and surface modification method are grouped together and marked with the same symbol. Figure 16 shows that the dispersion of the samples with the same polymer and surface modification could be improved by increasing the processing energy.

Table 6: Descriptor values of the composites samples.

Polymer	Particle surface modification	$W_{\text{PF}}/W_{\text{FF}}$	E_γ (J/g)	$\bar{I}_{\text{filler}}$
PS	Octyl	1.15	34.52	0.20
			85.73	0.21
	Chloro	1.04	33.18	0.15
			85.66	0.17
	Amino	0.95	33.28	0.12
			104.34	0.12
			85.76	0.12
PP	Octyl	0.94	0.65	0.16
			3.03	0.13
	Chloro	0.84	0.58	0.07
			3.53	0.17
	Amino	0.78	0.65	0.09
			2.16	0.12
			3.08	0.12
PMMA	Amino	0.96	103.10	0.11
			964.16	0.13
			410.92	0.12

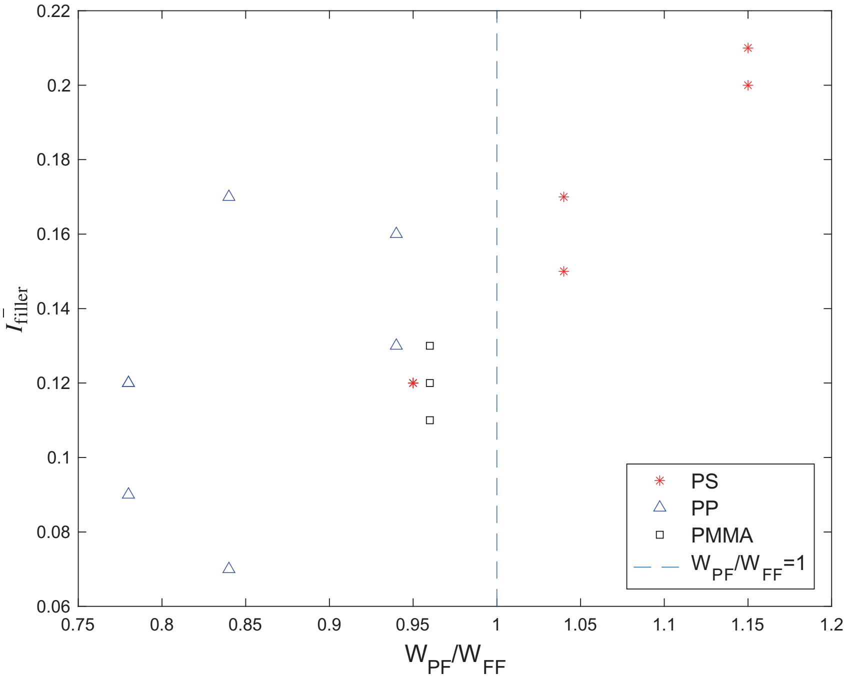


Figure 15: The impact of the filler–matrix compatibility descriptor ($W_{\text{PF}}/W_{\text{FF}}$) on the microstructure dispersion ($\bar{I}_{\text{filler}}$). The dashed line indicates the threshold of 1, beyond which materials have a wetting angle of 0° and the particles wet the polymer.

Natarajan *et al.*,[102] who used solvent mixing, found a very abrupt aggregation of nanoparticles when the compatibility of the nanoparticles with the polymer matrix goes from compatible (e.g., $W_{\text{PF}}/W_{\text{FF}} \geq 1$) to not compatible (e.g., $W_{\text{PF}}/W_{\text{FF}} < 1$). The results in this research show that the aggregation of the nanoparticles with a $\frac{W_{\text{PF}}}{W_{\text{FF}}} < 1$ is less abrupt than was found by Natarajan *et al.* for solvent mixed nanocomposites.[102] This is because the materials that are melt processed are not at equilibrium and the mixing energy and fast cooling prevents aggregation, while for Natarjan *et al.* the samples were annealed to a state of quasi-equilibrium.[102]

These results indicate that dispersion quality, $\bar{I}_{\text{filler}}$, can be correlated to both the material compatibility ($W_{\text{PF}}/W_{\text{FF}}$) and the processing energy (E_γ). The relationship between these variables was further developed using data mining techniques to provide a mathematical expression, which could be used in further analyses and prediction schemes. The results for the three types of polymer matrix are shown in Figure 17, where the R^2 values indicate that $\bar{I}_{\text{filler}}$ is linearly correlated with the combined energetic terms.

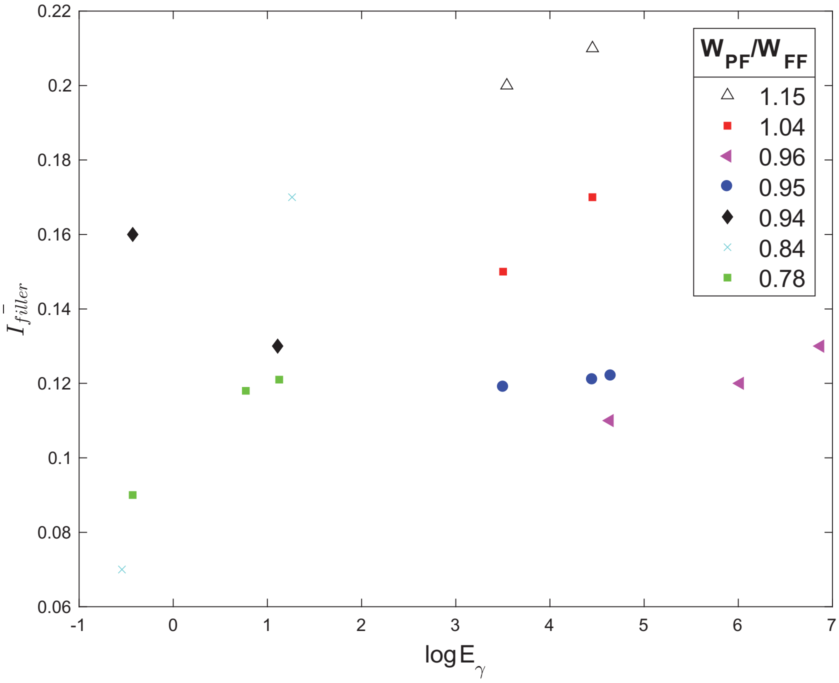


Figure 16: The impact of processing energy descriptor (E_γ) on the microstructure dispersion ($\bar{I}_{\text{filler}}$) in log scale. The polymer types and the surface modification methods that correspond to the data points can be found in Table 6.

Figure 17 reveals a few important observations. First, nanoparticles need more energy input or a better compatibility to be well-dispersed in PS and PMMA due to the smaller slope compared to that for PP. The difference between the slopes of the regression models implies that some properties of the polymer matrix type will have a large influence on the dispersion. The same result has been found in Ref. [92], where the influence of surface energies on the dispersion of CNT in different polymer matrices was analyzed.

Given that all the three composites follow a linear trend in this p–s relationship, we introduce a matrix-dependent term, $f(\text{matrix})$, in the p–s correlation. Then the linear functions could be generalized as

$$\bar{I}_{\text{filler}} = f(\text{matrix}) \sinh^2(2W_{\text{PF}}/W_{\text{FF}} - 1) \log(E_\gamma + 1) + C_0, \qquad (14)$$

where for PS system, $f(\text{matrix}) = 0.009947$ and $C_0 = 0.08798$; for PP system, $f(\text{matrix}) = 0.15039$ and $C_0 = 0.05464$; for PMMA, $f(\text{matrix}) = 0.007548$ and $C_0 = 0.071444$.

By conducting the three steps described above, we successfully establish a generalized, descriptor-based, linear p–s relationship using the dataset of PS,

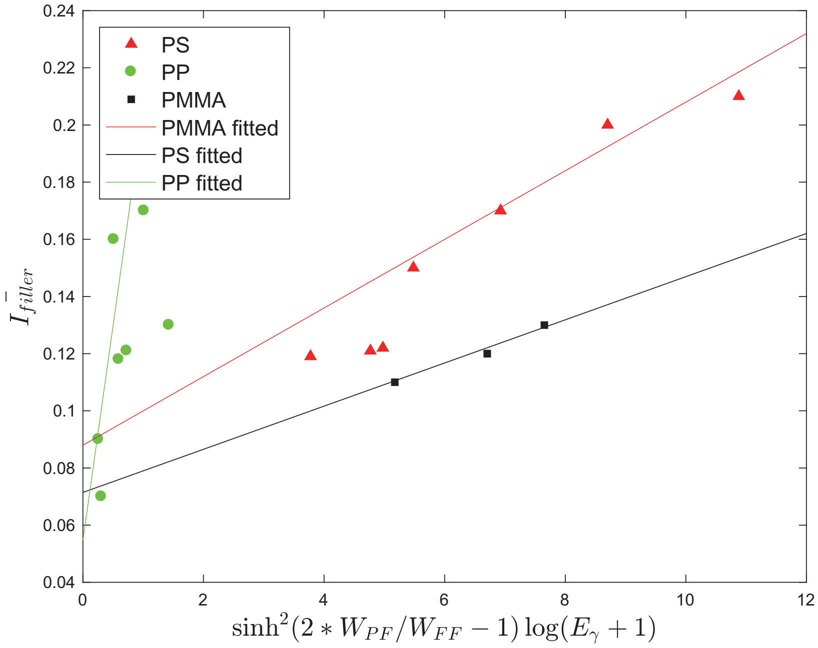


Figure 17: Regression model of the influence of interfacial energetics and processing conditions on the normalized interface of the nanocomposites.

PP, and PMMA composites. This p–s relationship has a matrix-dependent term that can be further investigated for different types of polymer matrices.

6. Structure-Property Evaluation with Adaptive Optimization for Interphase Calibration

In order to simulate and predict continuum level properties for polymer nanocomposites, *finite element* (FE) models have been developed.[16,29] Using data on the material constituents (composition) from the curated data, including polymer matrix and filler particle, combined with knowledge of the filler spatial distribution and interphase properties, a discretized RVE is created. RVEs can be simulated in a number of physical property domains to predict mechanical, electrical, dielectric, or other properties.

One challenge in s–p prediction of nanocomposites is the modeling of interphase behavior. Researchers have investigated interphase properties and its origin both analytically and experimentally.[122–124] Recent experimental efforts have shown adequate evidence that the local polymer properties are significantly altered in the vicinity of polymer surface, through measuring the local mechanical properties and by correlating thin film and nanocomposite data.[125,126] Given the limitations of direct measurements of interphase properties in the experiments, one approach to calculate the interphase properties is inversely through tuning the parameters in micro-scale model constitutive equations or finite element analysis using the bulk properties from experiments.[28,127,128]

It has been shown in our early work[28,29,129] that the interphase properties can often be described by shifting factors based on the pure matrix properties, which can be represented by a Prony Series as a parameter function of multiple relaxation and times and strengths. The interphase properties in those models can be obtained through a trial-and-error-based iterative tuning procedure by matching the simulations given specific experimental data. However, the trial-and-error-based tuning process can be very time-consuming given the complexity of experimental data and simulation cost of FEA.

We present here an automated optimization method to solve the inverse calibration problems. More details about this approach can be found in our previous paper.[26] Using a library of composite-level FE simulations, key interface modeling parameters (shift factors) are determined to match experimental properties. An adaptive sampling approach is implemented in searching for the unknown interphase modeling parameters, and the method has been tested on both dielectric and viscoelastic interphase properties. The interphase properties are assumed to be well-represented by the shifting factors with respect to matrix properties. Therefore, the optimization objective is to find the shifting factors that will optimally match the experimental data through applying adaptive optimization. The key component of our approach is summarized in Figure 18.

(1) Based on our previous understanding of the interphase, empirical bounds of shifting factors are applied to set the range to sample initial training shifting factors using optimal Latin hypercube (OLHC). The FEA model is run using each set of shifting factors to obtain a simulation result. The objective function is formulated as the difference between the simulation and experimental result using mean square error (MSE).
(2) A GP model is applied to fit the metamodel that correlates the objective function (difference F) with features (shifting factors), with uncertainties.

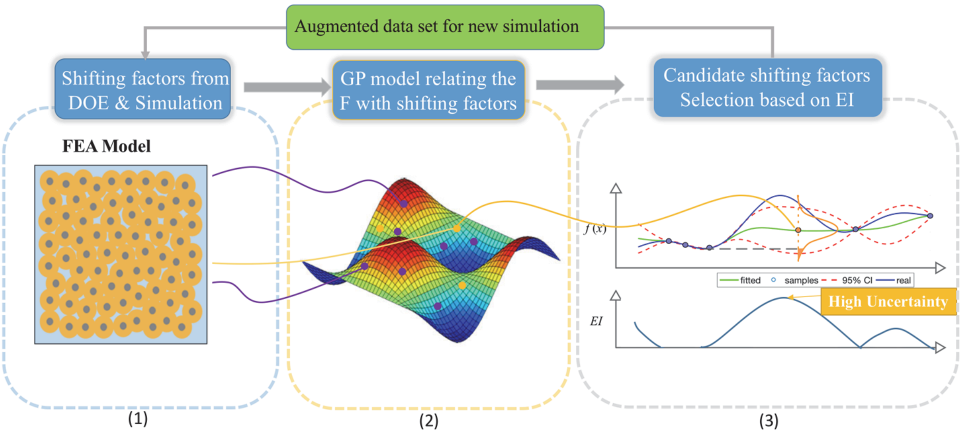


Figure 18: The automated adaptive optimization strategy for searching the interphase properties: (1) FEA is run on each point in the initial shift factor set from DOE and then processed to formulate our objective function by calculating the difference between simulated results and experiments using MSE; (2) a GP model is applied to construct the surrogates for predicting the value of objective function (difference F) within the design space; (3) EI chooses the best candidate points for additional simulation.

(3) Based on the fitted metamodel, new candidate points for the calibration parameters are selected based on feedback from the surrogate model by calculating the expected improvement (EI).[130–132]

This process is usually identified as an adaptive optimization process, also called Bayesian optimization, which augments the training shifting factor sets and drives the subsequent iterative improvement of the surrogate model and prediction accuracy. The adaptive optimizer is applied to construct the surrogate model and suggest the next optimal sampling points in order to minimize the difference between the experiments and simulations and return the optimal solution iteratively.

Using this automated approach and given the target experimental data, the interphase properties on dielectric and viscoelastic studies can be determined automatically. Taking a representative dielectric dataset (for a nanocomposite composed of 2 wt.% bimodal anthracene–PGMA grafted silica in epoxy[133,134]), the method is shown to estimate the interphase properties with only tens of iterations as shown in Figure 19. The prediction accuracy at each step is evaluated by calculating the discrepancy between the simulated data using the predicted shifting factor from the adaptive optimizer and the given experimental properties using MSE. As the iteration

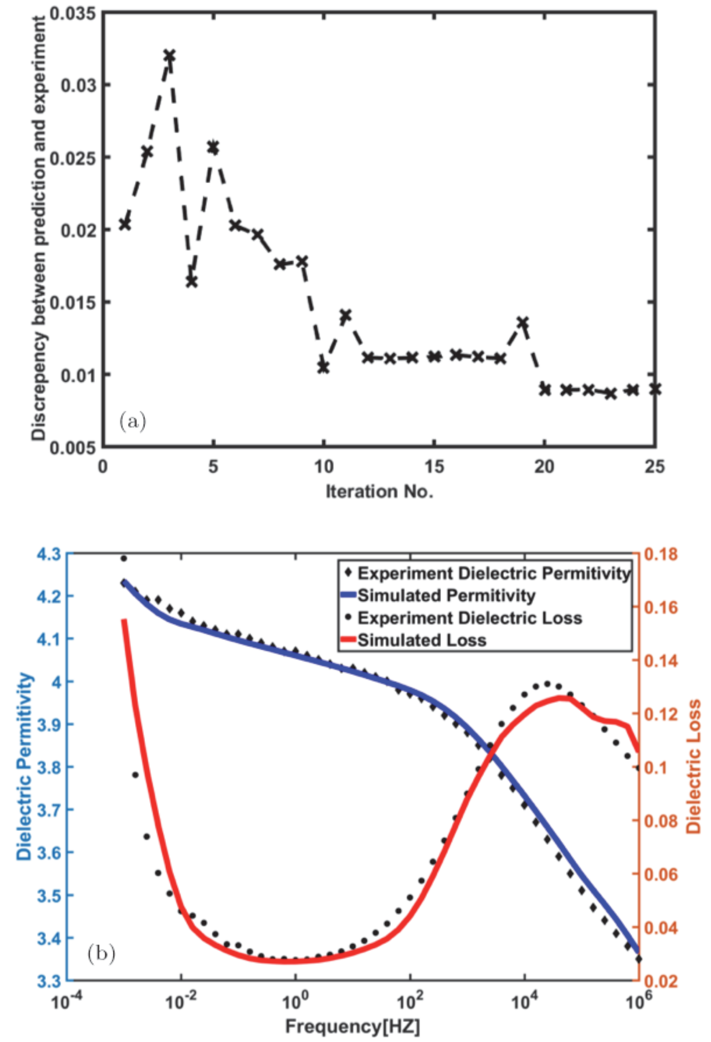


Figure 19: Evolving of optimal solution as a function of successive iterations: (a) the discrepancy between the optimal prediction and the experimental data as a function of iterations; (b) comparison between experimental data and the simulated result using the predicted interphase properties given from the adaptive optimizer.

proceeds, the discrepancy keeps decreasing (though not monotonically, likely due to the model uncertainties) indicating an increase in model accuracy as new sampling points are added based on maximum EI.

Ideally, minimal discrepancy between simulated and experimental data across the entire frequency band is desired and should lead to well-matched data curves. Practically, we find a threshold value C, indicating an acceptable fit between the experimental data and simulated results. C may vary depending the specific applications, but for our case C is set as 0.01. The iteration stops after the difference is smaller than the threshold. While the threshold is met at iteration 20, we plot more iterations in order to show the convergence of our adaptive optimization procedure.

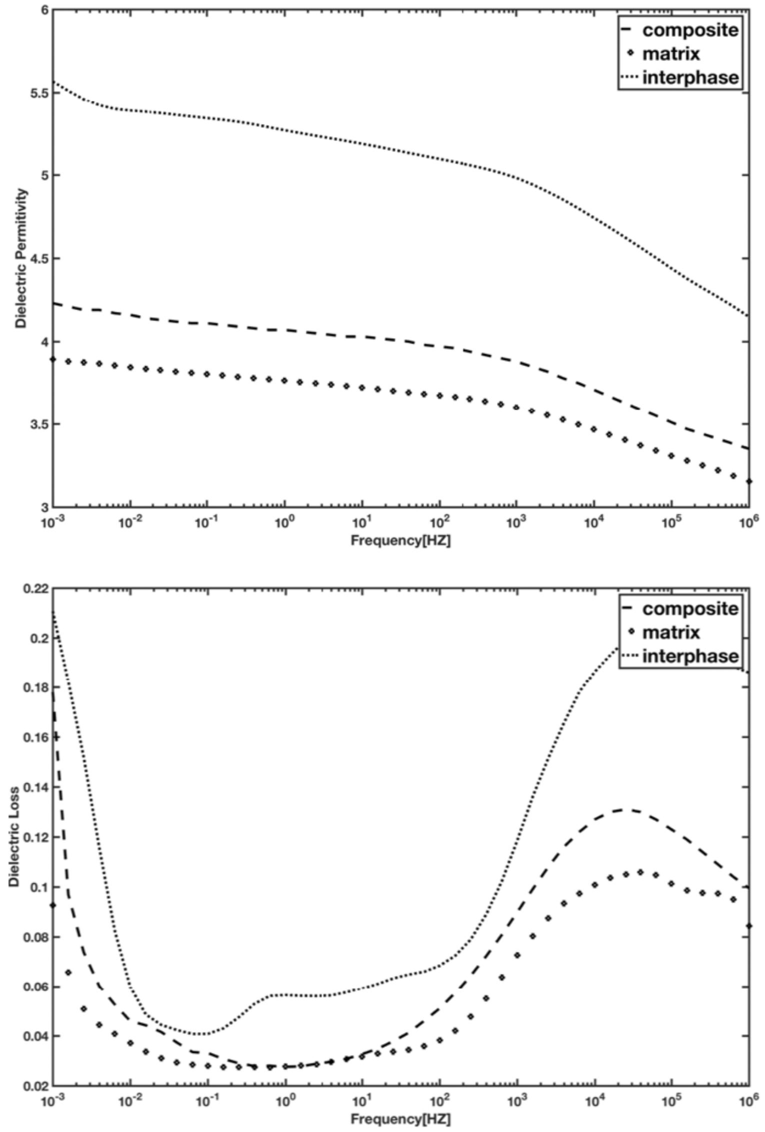


Figure 20: Comparison of matrix, interphase, and composite properties for 2 wt.% bimodal anthracene–PGMA grafted silica in epoxy.

By applying the adaptive optimizer above, the optimal shifting factors that achieve a good match with experimental data can be determined. Because the interphase shifting factors are the descriptors for the interphase properties and completely define those properties, the interphase properties are thus determined. It is important to note that without experimental verification of these local properties, these represent a reasonable approximation to the actual interphase properties. Figure 20 shows the optimal interphase properties together with the matrix and composite properties. The interphase properties determined are significantly different from that of the matrix. This property difference is expected due to mobility changes of polymer chains near the particles causing changes in the local physical properties such as dielectric spectra of the polymer. For this sample, the interphase region shows a higher permittivity and loss compared with matrix and composite data. This interphase property agrees with the experimental data where the addition of functional groups on fillers enhances the relative permittivity.[133] Further work is needed to determine predictive models for the interphase parameters based on sufficiently detailed experimental data.

7. Processing–Structure–Property Integration for Design of Nanodielectric Materials in Capacitors

In this section, the capacitor design problem is used as an example to demonstrate the design of nanodielectric materials by integrating p–s and s–p relationships presented in the earlier sections. Capacitor design is a multi-objective optimization problem where the objectives of maximizing dielectric constant and minimizing its dielectric loss need to be achieved simultaneously. *Dielectric constant*, also known as permittivity, affects the energy storage capability of capacitors. *Dielectric loss* quantifies a dielectric material's inherent dissipation of electromagnetic energy (e.g., heat). Increasing dielectric constant often results in dielectric loss. Multi-criteria optimization technique will be used to make the trade-off between competing design objectives.

The remainder of this section is organized as follows. Section 7.1 introduces the silica-PMMA- material system and the formulation of capacitor design optimization. Section 7.2 provides details of the s–p modeling, along with the multi-criteria microstructure optimization results. In Section 7.3, the optimized structure is mapped to processing conditions based on descriptor-based processing–structure mapping presented in Section 5.

7.1. *Capacitor design case study*

The material system of interest is PMMA/amino-modified silica nanocomposites with silica loading up to 5%. This model system is chosen because we have collected significant data for calibration and validation. The inclusion of silica in PMMA influences the overall permittivity and dielectric loss of the composites, and the dispersion quality of silica determines the strength of the effects. PMMA and silica are mixed using a twin screw extruder, and the high processing input energies tend to make the fillers more evenly dispersed in the polymer matrix.

The key property of a capacitor is the maximum energy, E it can store, which is determined by the following equation:

$$E = \frac{1}{2}\epsilon A d U_d^2, \tag{15}$$

where ϵ and U_d are the permittivity and dielectric breakdown strength of the dielectric material, A is the area of the plate in a parallel plate capacitor, and d is the gap thickness containing the dielectric material. Since we are designing the material, the geometry parameters A and d of the capacitor are not of interest. The breakdown of nanodielectrics is a complex phenomenon and the FEA simulation model is still under development. Therefore, in this case study, we assume the breakdown strength U_d is constant and only consider the permittivity as a major design target for maximizing energy storage.

Another important property in capacitor design is dielectric loss, which essentially quantifies the amount of heat generated during the charging and discharging process. Dielectric loss is usually represented by loss tangent, and is also known as dissipation factor (DF), and is given by

$$\tan\delta = DF = \frac{\epsilon''}{\epsilon'}, \tag{16}$$

where ϵ'' and ϵ' are the imaginary part and real part of the dielectric materials' permittivity. Lower loss tangent is desired to avoid significant energy loss during the usage of the capacitor.

In this case study, the microstructure of the composites is optimized based on FEM-based s–p relationships (Section 6) first and then the empirical p–s relationship (Section 5) is used to find the optimal processing conditions. The mathematical formulation of this multi-objective optimization problem becomes

$$\min_{s \epsilon S} -E, \quad \tan\delta, \tag{17}$$

where s represents the microstructure to optimize, and S is the set of all feasible microstructures. Using the physical-descriptor-based MCR approach introduced in Section 3, the microstructure representation can be reduced to a small number of design variables (descriptors) d. The above problem can be reformulated as a simple bounded optimization

$$\min_{d} -E, \quad \tan\delta; \text{ s.t. } d_L \leq d \leq d_U, \tag{18}$$

where d_L and d_U are the lower and upper bounds of design variables. One common way to deal with multiple objectives in optimization is to use the weighted sum of those objectives as a single objective. If using multi-criteria optimization, the Pareto frontier[135] is identified to form a set of non-dominated solutions based on the trade-off between multiple objectives. Genetic algorithms (GAs)[135] have been used widely to find the Pareto solutions.

7.2. *Nanodielectric microstructure optimization based on s–p relationship*

The FE-based approach as introduced in Section 6 is used to build predictive s–p models. Three major steps are followed:

1. Use SDF parameters θ and VF as design variables to generate optimal DOE sample points;
2. Apply the GRF-based algorithm[78] to reconstruct microstructures based on the parameters generated in step (1);
3. Evaluate the properties of microstructures reconstructed in step (2) using FEA simulation models discussed in Section 6.

It is worth noting here that the normalized filler surface area, $\bar{I}_{\text{filler}}$, which has been found to be the critical structure descriptor in p–s mapping as shown in Section 5, is not directly used as a design variable in our problem. Instead, we use the parameter of SDF distribution as one design variable. As explained in Section 3.3, SDF representation provides significant dimension reduction in microstructural analysis and design. Its applicability is also confirmed by image analysis on the material samples made from the processing technique adopted. By analyzing the microstructures of the PMMA-silica system (Figure 21), the SDFs are found to be close to Gamma distribution with

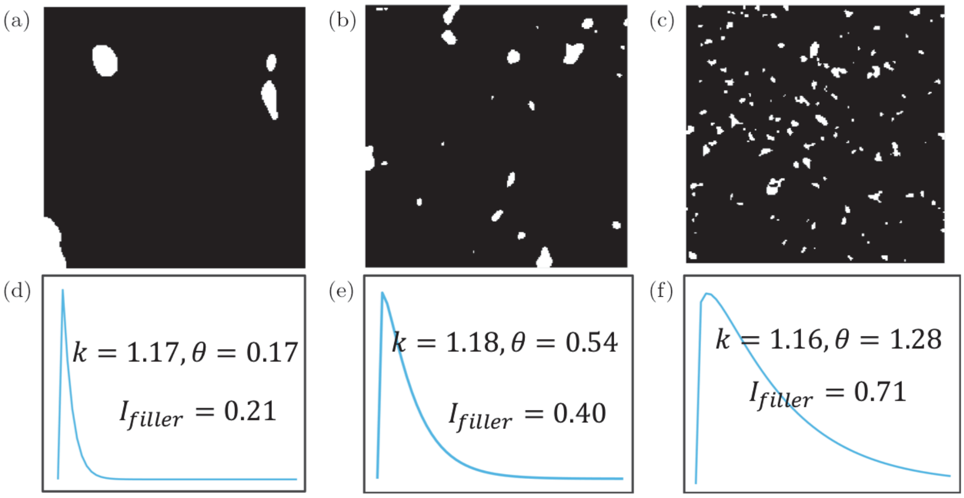


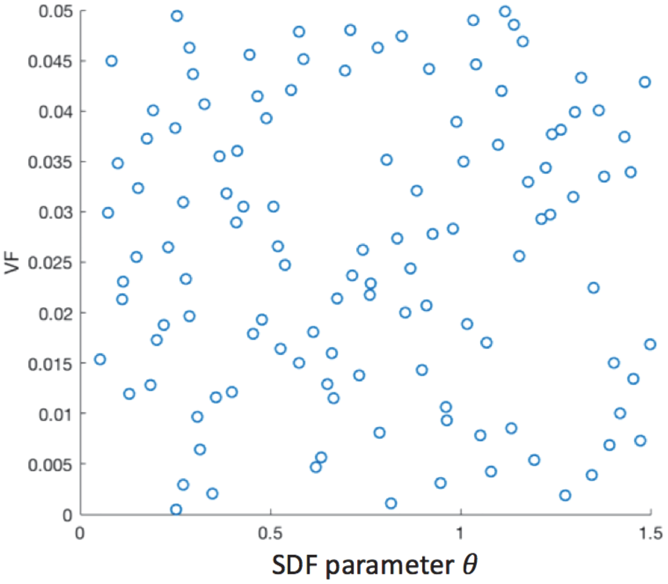
Figure 21: (a)–(c) Microstructures of PMMA–silica system; (d)–(f) corresponding SDF and the fitted parameter k and θ, along with the normalized interphase fraction $\bar{I}_{\mathrm{filler}}$.

two parameters, k and θ

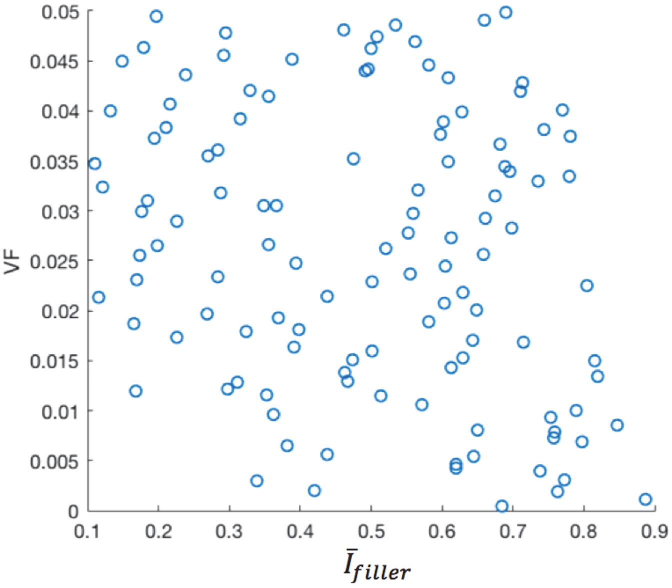
$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, \tag{19}$$

where $\Gamma(\cdot)$ is the Gamma function, k is a shape parameter, and θ is a scale parameter. The peak (mode) of Gamma distribution described in the equation above is $(k-1)\theta$. It is noted from samples collected that the shape parameter k varies in a very small interval (1.15–1.19) and has little influence on the profile of SDF. In contrast, the scale parameter θ varies from 0.05 to 1.5 and changes the width of SDF significantly as shown in Figure 21. In our illustrative design case study, we only use θ as one design variable of microstructure and fix k to its mean value 1.17. It is also found that the scale parameter, θ, is strongly correlated with $\bar{I}_{\mathrm{filler}}$, the important microstructure variable identified earlier, with $\bar{I}_{\mathrm{filler}}$ increasing as θ increases. This strong correlation justifies the use of parameter θ in SDF as a replacement to $\bar{I}_{\mathrm{filler}}$.

We further verify whether the generated DOE sample points using SDF parameter, θ, are evenly distributed in the normalized interphase dimension. Figure 22(a) shows the original DOE space using the SDF parameter. Based on the microstructure reconstructions from step (2), we obtain the normalized filler surface area, $\bar{I}_{\mathrm{filler}}$, of each microstructure and then present the sample points in Figure 22(b). We note that the design space of $\bar{I}_{\mathrm{filler}}$ and VF are covered evenly by those sample points.



(a)



(b)

Figure 22: (a) 120 DOE points covering VF and SDF parameter space; (b) 120 DOE points covering VF and the normalized interphase descriptor $\overline{I}_{\text{filler}}$.

After the DOE samples are created over θ and VF, the next step is to evaluate the properties of these samples. The s–p evaluation approach described in Section 6 using physics-based FEA combined with calibrated interphase parameters is followed. Due to the multiple relaxations observed in the pure matrix, the interphase properties are described using a set of 5D shifting factors (M_α and S_α for alpha relaxation, M_β and S_β for beta relaxation, and C for the intensity shift of permittivity) related to pure matrix with the alpha and beta relaxation modeled separately. The interphase shifting factors are calibrated and obtained inversely by optimally matching a the composite dielectric spectrum from experimental data for amino-modified silica–PMMA composites via an automated procedure applying the adaptive optimization presented in Section 6. Since the composition (filler, matrix, functional group) remains constant for all cases in the DOE, the interphase shifting factors found in this manner can be used for all simulation predictions in the design study. In the interphase optimization process, multiple microstructural images for each sample are used to obtain the appropriatre structural descriptors for reconstruction, which is then used in the optimization procedure. With the interphase properties obtained from the optimization procedure, new microstructure designs are simulated following the DOE samples to obtain the nanocomposite bulk properties. From the histogram in Figure 23, we can see the simulated permittivity ranges from 2.5 to 2.8, and the losses $\tan\delta$ are roughly in the range of 0.04–0.063. The distribution of permittivity is close to normal distribution but slightly right skewed. The dielectric loss has a highly skewed distribution with concentrations in small values.

Based on the simulation samples, machine learning techniques are used to build metamodels to replace the expensive s–p simulations. One common technique applied to approximate expensive computer models is the GP regression model,[136] which is a non-parametric statistical model that captures nonlinear response surfaces and provides flexibility for assessing uncertainties in prediction.

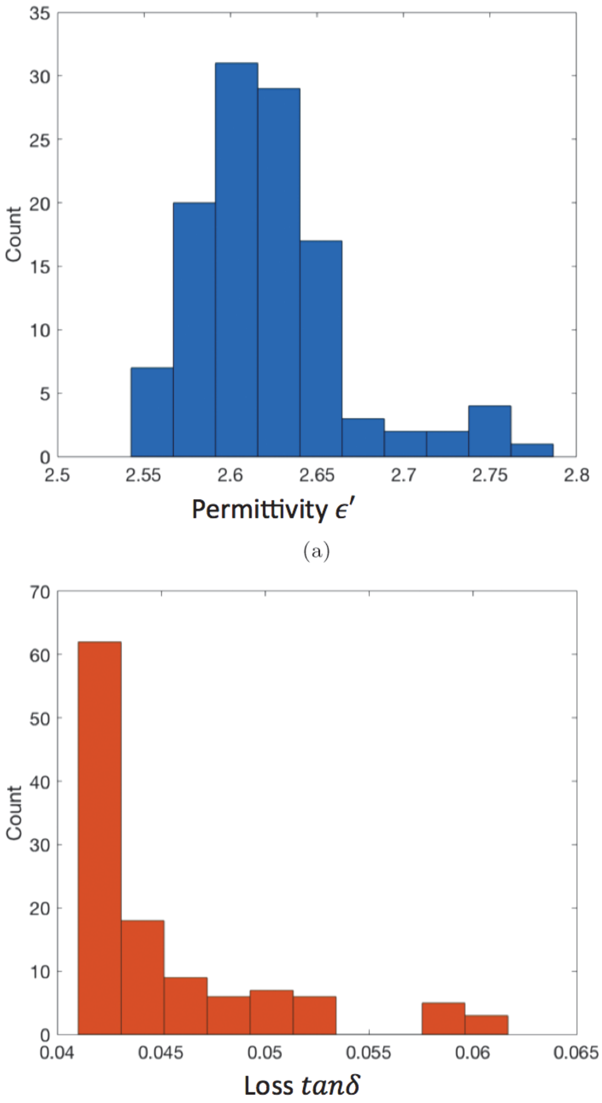
A GP $y(\mathbf{x})$ is defined by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}\mathbf{x}')$:

$$m(\mathbf{x}) = E[y(\mathbf{x})], \tag{20}$$

$$k(\mathbf{x},\mathbf{x}') = E[(y(\mathbf{x}) - m(\mathbf{x}))(y(\mathbf{x}') - m(\mathbf{x}'))], \tag{21}$$

where the vector $\mathbf{x}$ indicates the spatial location, and the response y at each location $\mathbf{x}$ follows the multi-variate normal distribution:

$$y(\mathbf{x}) \sim N(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')). \tag{22}$$



(a)

(b)

Figure 23: (a) Histogram of the simulated permittivity at 1 Hz; (b) histogram of simulated loss $\tan\delta$ at 1 Hz.

For simplicity, the mean function is chosen to be equal to zero and the covariance function is in the squared exponential form, also known as Kriging[137]:

$$k(\mathbf{x}\mathbf{x}') = \sigma^2 \exp(-w(\mathbf{x} - \mathbf{x}')^2), \tag{23}$$

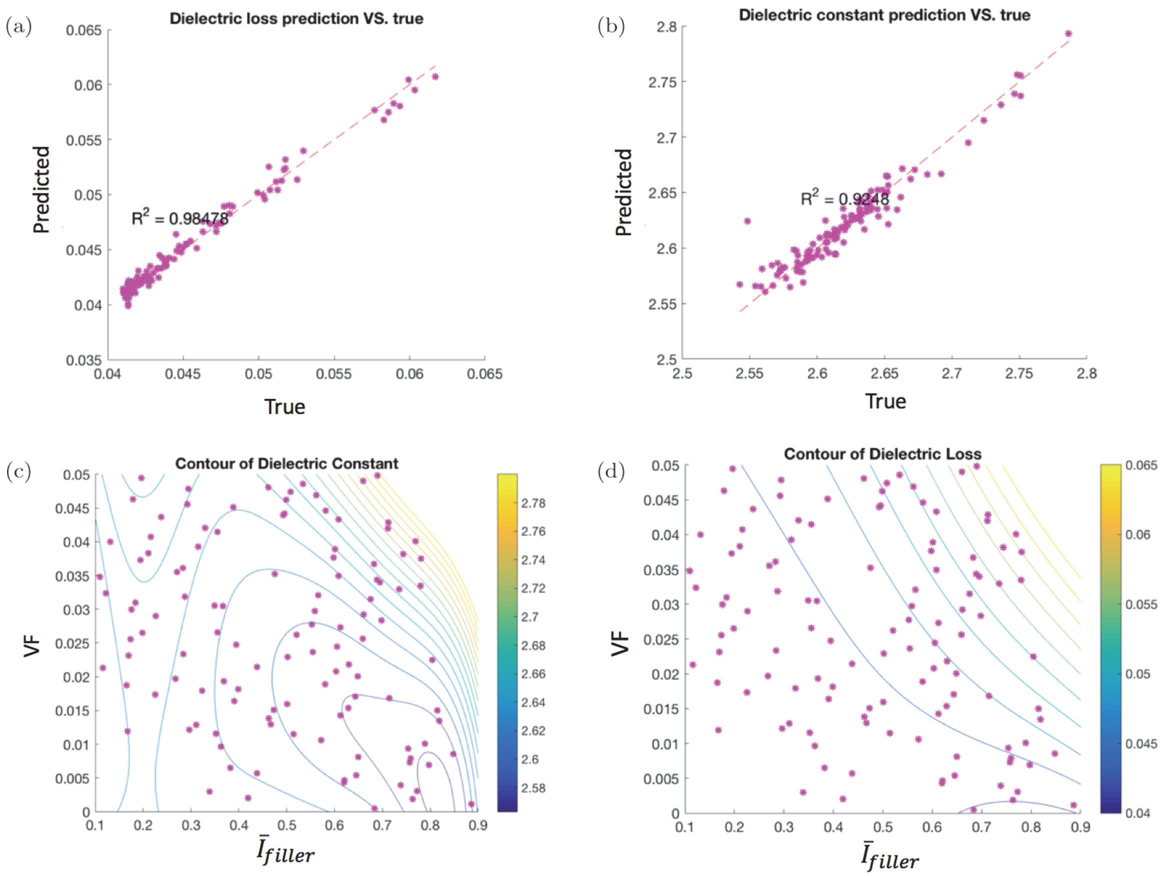


Figure 24: (a) Prediction accuracy of the model for dielectric constant; (b) prediction accuracy of the model for dielectric loss; (c) contour plot of the fitted model for dielectric constant; (d) contour plot of the fitted model for dielectric loss.

where σ is a constant and w is the roughness parameter of GP, whose dimension is the same as x. Both σ and w are obtained by maximum likelihood estimation (MLE). With these notations, the variance of prediction is also tractable, and the detailed mathematical expression can be found in Ref. [138].

From Figures 24(a) and 24(b), we found that the fitted GP models lead to very accurate dielectric property predictions, with R^2 in both models above 0.9, which is strong evidence that the GP models are good surrogates of the more expensive FEA simulation models. To understand the relationship between the dielectric properties and the microstructure design variables VF and $\bar{I}_{\text{filler}}$, we provide the contours of the fitted models in Figures 24(c) and 24(d). It is observed that when VF is low (below 2%), as $\bar{I}_{\text{filler}}$ increases, the dielectric constant decreases, while when VF is relatively high, as I_{filler} increases, the dielectric constant becomes larger, reaching maximum at the upper right region in the contour graph. When the dispersion state is fixed, with higher loading of fillers, we expect higher dielectric constant. For dielectric loss tangent, the lower left region is almost constant, which means that when VF and $\bar{I}_{\text{filler}}$ are low, they have little influence on the loss tangent. In contrast, when VF and $\bar{I}_{\text{filler}}$ both are at a relatively high level, they have strong positive effects on loss tangent. Both properties are maximized around the upper right region in Figure 24, which indicates a trade-off between them, since we want to minimize one objective and maximize the other. We will show how to use a genetic algorithm to tackle the multi-criteria optimization in Section 7.3.

7.3. *Multi-criteria optimization using genetic algorithm*

GA is used as a search algorithm for identifying the solutions to multi-criteria optimization in our design case study. Different from classical optimization methodologies, genetic algorithms do not require gradient information, and the search approaches the set of optimal solutions in an evolutionary manner. In each iteration, a population of solutions is created and the current population is updated into the next generation by four main operators: selection, crossover, mutation, and elite-preservation. Intuitively, GA selects "better" solutions with higher probability in each iteration according to certain criteria, and then combines and perturbs these solutions to create the next population of solutions. In the context of multi-objective optimization, GA favors non-dominant solutions (points in the Pareto frontier). As shown in Figure 25, for a problem where the objective is to minimize both objective 1 f_1 and objective 2 f_2, point E is dominated by points A and C: Point A

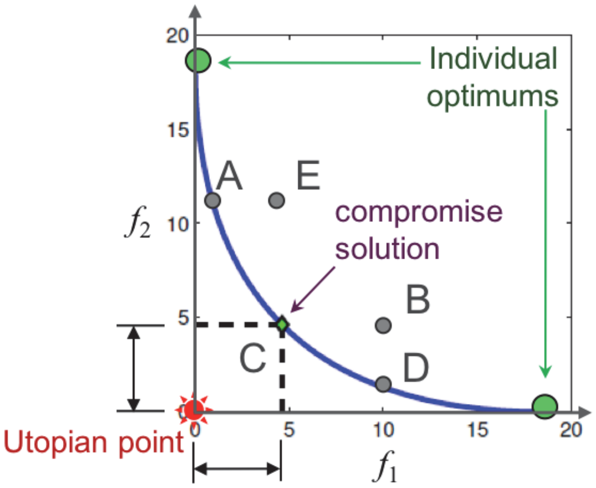


Figure 25: Non-dominant points (Pareto frontier) in two-objective minimization problem.

reduces f_1 without increasing f_2, and at point C, f_1 is improved compared to point E. Point B is dominated by points C and D because C and D both have one objective less (better) than B, while the other objective equals B. Therefore, points A, C, and D form the Pareto frontier of this two-objective optimization example. It is noted that individual optimums are at the two ends of the Pareto frontier and any other points in between on the Pareto curve are compromise solutions. At the lower left corner is the utopian point where both objectives are individually minimized, which is not achievable in most cases when the two objectives compete with each other.

The fitted GP models based on s–p simulations are used for multi-criteria optimization. The obtained Pareto frontier is visualized in Figure 26. Moving along the Pareto frontier from left to the right, the dielectric loss is decreased while the dielectric constant is also reduced, limiting the energy storage capability of the capacitor. Note that the negative (−) value of dielectric constant is plotted. To maximize dielectric constant, the filler needs to be well-dispersed and uniformly distributed in the polymer matrix as illustrated by the upper left image. In contrast, with a few large clusters, the microstructure on the lower right minimizes the dielectric loss while also decreasing dielectric constant (permittivity). The microstructure in the middle with moderate dispersion is a compromise result for both dielectric constant and loss. This observation is consistent with our fitted model shown in Figure 26: with better dispersion (larger $\bar{I}_{\text{filler}}$), both dielectric constant and dielectric loss increase. Which point in the Pareto set should be chosen as the design solution will depend on designer's preference

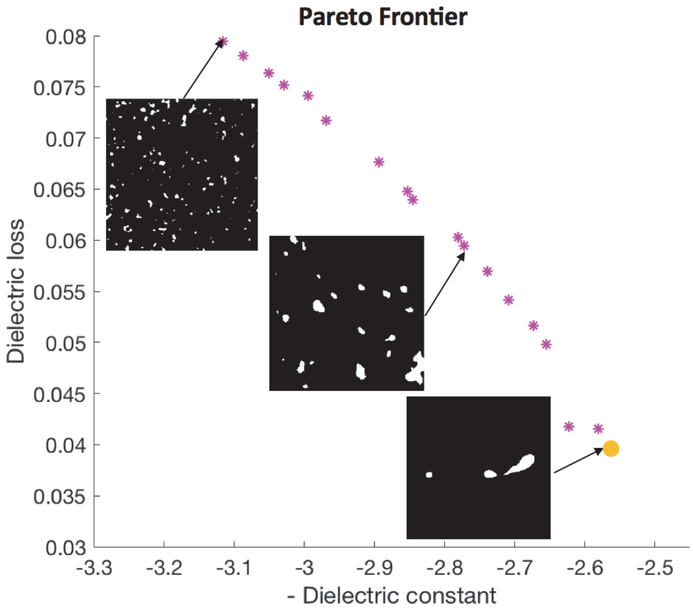


Figure 26: Pareto frontier identified by GA.

function that captures the willingness of making trade-offs between the two objectives.

Once the optimal microstructure design is identified, the corresponding processing condition is obtained by mapping the optimized normalized interphase I_{filler} back to processing energy using the p–s relationship established in Section 5. Here, we choose the point $(-2.623, 0.0418)$ from Figure 26 as a demonstration. This design at one end of the Pareto frontier corresponds to the optimal design for minimum dielectric constant but also minimum dielectric loss. Its corresponding design variables are found to be $VF = 0.0157$, $\bar{I}_{\text{filler}} = 0.2718$. In Section 5, we showed the empirical relationship between processing conditions and the normalized interphase $\bar{I}_{\text{filler}}$ as

$$\bar{I}_{\text{filler}} = 0.07144 + 0.007548 \sinh^2\left(\frac{2W_{\text{PF}}}{W_{\text{FF}}} - 1\right) \log(E_\gamma + 1). \tag{24}$$

As the filler matrix compatibility $\frac{W_{\text{PF}}}{W_{\text{FF}}}$ and the input energy E_γ increase, $\bar{I}_{\text{filler}}$ will also increase. Recall that the material system we studied so far is PMMA/amino-mod silica, whose surface energy $\frac{W_{\text{PF}}}{W_{\text{FF}}} = 0.96$. Inserting this quantity into the above equation, we obtain the relationship between $\bar{I}_{\text{filler}}$

and the input energy E_γ as

$$\bar{I}_{\text{filler}} = 0.07144 + 0.0084 * \log(E_\gamma + 1). \tag{25}$$

To achieve the optimized $\bar{I}_{\text{filler}} = 0.2718$, the input energy needs to satisfy $2 * 10^7$(kJ/g).

8. Conclusion

In this chapter, we present a data-centric approach that integrates physics-based models, empirical data, machine learning approaches, and a calibrated interphase model, to discover new knowledge and accelerate the design of polymer nanocomposites. The implementation of this approach is supported by an open nanocomposite data system ("NanoMine"), where both data resources and tools for microstructural analysis and optimal materials design are integrated. We illustrate that microstructural analysis and optimal materials design is a systematic process in which multiple steps of image preprocessing, microstructure characterization, reconstruction, dimension reduction, machine learning of p–s–p relationships, and optimal design need to be followed.

While material informatics heavily relies on existing machine learning and data mining techniques, a key question of advancing these techniques is what is the proper microstructure representation for the materials systems of interest? We present in this chapter a range of microstructure representations, including correlation-function, physical descriptor-based approach, SDF, and machine learning-based MCR techniques. In addition to considering the specific microstructure geometry, we illustrate, in this chapter, the importance of choosing a representation that allows physically meaningful mappings of p–s–p relationships and rapid microstructure optimization. We used the PMMA–silica-based nanodielectric polymer for a capacitor design as a case study to illustrate the creation of descriptor-based p–s relationships based on experimental data, construction of hybrid physics-based and data-calibrated interphase model for s–p simulation, and the integration of two models using the SDF representation for significant dimension reduction in materials design.

NanoMine now contains over 1,200 distinct material samples, the majority of which are curated from the literature. Most of the data have extensive information on material constituents and composition, synthesis and processing methods, processing conditions and associated properties, as well as micrograph images if reported from the data source. The volume of the data will continuously grow in our future work by expanding on more types of nanocomposite materials, such as high-aspect ratio fillers (carbon nanotubes, nanofibers, clays, and nanoplates) with anisotropic dispersion and extended property domains (optical and structural properties).

We are also working on continuously expanding the capabilities of the NanoMine toolkit and developing more tools based on the p–s–p parameters by exploring curated data in NanoMine. For example, our work has shown a combined simulation and optimization approach to calibrate the interphase property given the bulk composite property and microstructure. We are building connections for this tool with the curated data to create an interphase property library for different combinations of material constituents and surface treatments using the existing material data. This interphase library will help us build the correlations of interphase properties with processing steps and material characteristics. This suite of data and tools could be utilized to achieve a better understanding of the underlying mechanisms behind the interphase material behavior and help design the optimal interphase property by controlling the material constituents or surface treatment. Our modeling and design capability will be further enhanced by introducing multi-scale modeling and multi-phase design search engines to predict and optimize nanocomposite properties with explicit inclusion of constituent chemistry, interfacial energies, intrinsic and extrinsic interface properties, filler dispersion, matrix morphology, and uncertainty quantification.

Acknowledgment

This research is made possible through the support from NSF Dibbs, DMREF, and DEMS grants, 1640840, 1729743, 1729452, and 1818574.

References

1. J. Hill, G. Mulholland, K. Persson, R. Seshadri, C. Wolverton, and B. Meredig, “Materials science with large-scale data and informatics: Unlocking new opportunities,” *MRS Bulletin*, 41(5), 399–409, 2016.
2. A. Jain, G. Hautier, S. P. Ong, and K. Persson, “New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships,” *J Mater Res*, 31(8), 977–994, 2016.
3. A. Jain *et al.*, “Commentary: The materials project: A materials genome approach to accelerating materials innovation,” *Apl Materials*, 1(1), 011002, 2013.
4. K. Rajan, “Materials informatics: The materials “gene” and big data,” *Ann Rev Mater Res*, 45, 153–169, 2015.
5. D. T. Fullwood, S. R. Niezgoda, B. L. Adams, and S. R. Kalidindi, ”Microstructure sensitive design for performance optimization,” *Prog Mater Sci*, 55(6), 477–562, 2010.

6. S. R. Kalidindi, S. R. Niezgoda, and A. A. Salem, "Microstructure informatics using higher-order statistics and efficient data-mining protocols," *JOM*, 63(4), 34–41, 2011.
7. S. Ramakrishna *et al.*, "Materials informatics," *J Intell Manuf*, 30(1), 2307–2326, 2019.
8. K. Rajan, *Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application*. Butterworth-Heinemann, Oxford, UK, 2013.
9. S. R. Kalidindi *et al.*, "Role of materials data science and informatics in accelerated materials innovation," *MRS Bulletin*, 41(8), 596–602, 2016.
10. S. R. Kalidindi, A. J. Medford, and D. L. McDowell, "Vision for data and informatics in the future materials innovation ecosystem," *JOM*, 68(8), 2126–2137, 2016.
11. L. Schadler, L. Brinson, and W. Sawyer, "Polymer nanocomposites: a small part of the story," *JOM*, 59(3), 489–507, 2007.
12. N. R. Council, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security*. National Academies Press, Washington D.C., USA, 2008.
13. M. S. Greene, Y. Liu, W. Chen, and W. K. Liu, "Computational uncertainty analysis in multiresolution materials via stochastic constitutive theory," *Comput Meth Appl Mech Eng*, 200(1), 309–325, 2011.
14. Y. Liu, M. Steven Greene, W. Chen, D. A. Dikin, and W. K. Liu, "Computational microstructure characterization and reconstruction for stochastic multiscale material design," *Comput-Aided Design*, 45(1), 65–76, 2013.
15. H. Xu, D. A. Dikin, C. Burkhart, and W. Chen, "Descriptor-based methodology for statistical characterization and 3D reconstruction of microstructural materials," *Comput Mater Sci*, 85, 206–216, 2014.
16. H. Xu, Y. Li, C. Brinson, and W. Chen, "A descriptor-based design methodology for developing heterogeneous microstructural materials system," *J Mech Design*, 136(5), 051007, 2014.
17. H. Xu, R. Liu, A. Choudhary, and W. Chen, "A machine learning-based design representation method for designing heterogeneous microstructures," *J Mech Design*, 137(5), 051403, 2015.
18. Y. Zhang, H. Zhao, I. Hassinger, L. Brinson, L. Schadler, and W. Chen, "Microstructure reconstruction and structural equation modeling for computational design of nanodielectrics," *Integr Mater Manufact Innov*, 4(1), 14, 2015.
19. A. Bansal *et al.*, "Quantitative equivalence between polymer nanocomposites and thin polymer films," *Nat Mater*, 4(9), 693, 2005.
20. P. Akcora *et al.*, "Anisotropic self-assembly of spherical polymer-grafted nanoparticles," *Nat Mater*, 8(4), 354, 2009.
21. T. Ramanathan *et al.*, "Functionalized graphene sheets for polymer nanocomposites," *Nat Nanotechnol*, 3(6), 327, 2008.
22. S. Abiteboul, P. Buneman, and D. Suciu, *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann, San Francisco, CA, USA, 2000.
23. T. N. Bhat, L. M. Bartolo, U. R. Kattner, C. E. Campbell, and J. T. Elliott, "Strategy for extensible, evolving terminology for the Materials Genome Initiative efforts," *JOM*, 67(8), 1866–1875, 2015.
24. H. Zhao *et al.*, "NanoMine schema: An extensible data representation for polymer nanocomposites," *APL Materials*, 6(11), 111108, 2018.
25. I. Hassinger *et al.*, "Toward the development of a quantitative tool for predicting dispersion of nanocomposites under non-equilibrium processing conditions," *J Mater Sci*, 51(9), 4238–4249, 2016.

26. Y. Wang *et al.*, "Identifying interphase properties in polymer nanocomposites using adaptive optimization," *Compos Sci Technol*, 162, 146–155, 2018.
27. R. D. Bradshaw and L. Brinson, "A sign control method for fitting and interconverting material functions for linearly viscoelastic solids," *Mech Time-Dependent Mater*, 1(1), 85–108, 1997.
28. R. Qiao and L. C. Brinson, "Simulation of interphase percolation and gradients in polymer nanocomposites," *Compos Sci Technol*, 69(3), 491–499, 2009.
29. H. Zhao *et al.*, "Dielectric spectroscopy analysis using viscoelasticity-inspired relaxation theory with finite element modeling," *IEEE Transactions on Dielectrics and Electrical Insulation*, 24(6), 3776–3785, 2017.
30. Y. Huang *et al.*, "Predicting the breakdown strength and lifetime of nanocomposites using a multi-scale modeling approach," *J Appl Phys*, 122(6), 065101, 2017.
31. C. M. Breneman *et al.*, "Stalking the materials genome: A data approach to the virtual design of nanostructured polymers," *Adv Funct Mater*, 23(46), 5746–5752, 2013.
32. S. Torquato, *Random heterogeneous materials: microstructure and macroscopic properties.* Vol. 16. Springer-Verlag, New York, 2013.
33. H. Kumar, C. Briant, and W. Curtin, "Using microstructure reconstruction to model mechanical behavior in complex microstructures," *Mech Mater*, 38(8), 818–832, 2006.
34. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans Sys, Man, Cybernet*, 9(1), 62–66, 1979.
35. W. Niblack, *An Introduction to Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1986, pp. 115–116.
36. B. HKDH, "Neural networks in materials science," *ISIJ Int*, 39(10), 966–979, 1999.
37. V. Sundararaghavan and N. Zabaras, "Classification and reconstruction of three-dimensional microstructures using support vector machines," *Comput Mater Sci*, 32(2), 223–239, 2005.
38. R. Bostanabad *et al.*, "Computational microstructure characterization and reconstruction: Review of the state-of-the-art techniques," *Progress in Materials Science*, 95, 1–41, 2018.
39. P. B. Corson, "Correlation functions for predicting properties of heterogeneous materials. I. Experimental measurement of spatial correlation functions in multiphase solids," *J Appl Phys*, 45(7), 3159–3164, 1974.
40. P. B. Corson, "Correlation functions for predicting properties of heterogeneous materials. II. Empirical construction of spatial correlation functions for two-phase solids," *J Appl Phys*, 45(7), 3165–3170, 1974.
41. S. Torquato and G. Stell, "Microstructure of two-phase random media. I. The n-point probability functions," *J Chem Phys*, 77(4), 2071–2077, 1982.
42. J. G. Berryman, "Measurement of spatial correlation functions using image processing techniques," *J Appl Phys*, 57(7), 2374–2384, 1985.
43. D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms," *Acta Materialia*, 56(5), 942–948, 2008.
44. B. Lu and S. Torquato, "Lineal-path function for random heterogeneous materials," *Phys Rev A*, 45(2), 922, 1992.
45. Y. Jiao, F. Stillinger, and S. Torquato, "A superior descriptor of random textures and its predictive capacity," *Proc Nat Acad Sci*, 106(42), 17634–17639, 2009.
46. S. Torquato, J. Beasley, and Y. Chiew, "Two-point cluster function for continuum percolation," *The J Chem Phys*, 88(10), 6540–6547, 1988.
47. C. Yeong and S. Torquato, "Reconstructing random media," *Phys Rev E*, 57(1), 495, 1998.

48. C. Yeong and S. Torquato, "Reconstructing random media. II. Three-dimensional media from two-dimensional cuts," *Phys Rev E*, 58(1), 224, 1998.
49. E. Aarts and J. Korst, "Simulated annealing and Boltzmann machines," 1988.
50. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans Evolut Comput*, 6(2), 182–197, 2002.
51. B. Collins, K. Matous, and D. Rypl, "Three-dimensional reconstruction of statistically optimal unit cells of multimodal particulate composites," *Int J Multiscale Comput Eng*, 8(5), 2010.
52. N. C. Kumar, K. Matouš, and P. H. Geubelle, "Reconstruction of periodic unit cells of multimodal random particulate composites using genetic algorithms," *Comput Mater Sci*, 42(2), 352–367, 2008.
53. N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recogn*, 26(9), 1277–1294, 1993.
54. J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statist Sci*, 4, 409–423, 1989.
55. T. W. Simpson, D. K. Lin, and W. Chen, "Sampling strategies for computer experiments: design and analysis," *Int J Reliab Appl*, 2(3), 209–240, 2001.
56. R. Jin, W. Chen, and A. Sudjianto, "An efficient algorithm for constructing optimal design of computer experiments," *J Statis Plann Inference*, 134(1), 268–287, 2005.
57. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J Mach Learn Res*, 3(Mar), 1157–1182, 2003.
58. M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, 1997, pp. 296–304.
59. Mathworks®, "Image Processing Toolbox$^{\mathrm{TM}}$: User's Guide (R2017b)."
60. H. Xu *et al.*, "Stochastic reassembly strategy for managing information complexity in heterogeneous materials analysis and design," *J Mech Design*, 135(10), 101010, 2013.
61. Y. Zhang, H. Zhao, I. Hassinger, L. C. Brinson, L. S. Schadler, and W. Chen, "Microstructure reconstruction and structural equation modeling for computational design of nanodielectrics," *Integr Mater Manufact Innov*, 4(1), 14, 2015.
62. M. Baghgar *et al.*, "Morphology-dependent electronic properties in cross-linked (P3HT-b-P3MT) block copolymer nanostructures," *ACS Nano*, 8(8), 8344–8349, 2014.
63. H.-L. Tyan, Y.-C. Liu, and K.-H. Wei, "Thermally and mechanically enhanced clay/polyimide nanocomposite via reactive organoclay," *Chem Mater*, 11(7), 1942–1947, 1999.
64. T. M. Truskett, S. Torquato, S. Sastry, P. G. Debenedetti, and F. H. Stillinger, "Structural precursor to freezing in the hard-disk and hard-sphere systems," *Phys Rev E*, 58(3), 3083, 1998.
65. C.-W. Nan and D. Clarke, "The influence of particle size and particle fracture on the elastic/plastic deformation of metal matrix composites," *Acta Materialia*, 44(9), 3801–3811, 1996.
66. L. Karasek and M. Sumita, "Characterization of dispersion state of filler and polymer-filler interactions in rubber-carbon black composites," *J Mater Sci*, 31(2), 281–289, 1996.
67. M. Yuan and L.-S. Turng, "Microstructure and mechanical properties of microcellular injection molded polyamide-6 nanocomposites," *Polymer*, 46(18), 7273–7292, 2005.
68. B. Dong *et al.*, "Optical response of a disordered bicontinuous macroporous structure in the longhorn beetle Sphingnotus mirabilis," *Phys Rev E*, 84(1), 011915, 2011.

69. W.-K. Lee, C. J. Engel, M. D. Huntington, J. Hu, and T. W. Odom, "Controlled three-dimensional hierarchical structuring by memory-based, sequential wrinkling," *Nano Lett*, 15(8), 5624–5629, 2015.
70. C. Chatfield, The Analysis of Time Series: An Introduction. CRC press, New York, 2016.
71. G. J. Hedley *et al.*, "Determining the optimum morphology in high-performance polymer-fullerene organic photovoltaic cells," *Nat Commun*, 4(2867), 2013.
72. U. Farooq Ghumman *et al.*, "A Spectral Density Function Approach for Active Layer Design of Organic Photovoltaic Cells," *J Mech Design*, 140(11), 111408–111414, 2018.
73. S. C. Yu *et al.*, "Characterization and design of functional quasi-random nanostructured materials using spectral density function," *J Mech Design*, 139(July), 135–145, 2016, https://doi.org/10.1115/1.4036582.
74. W.-K. Lee *et al.*, "Concurrent design of quasi-random photonic nanostructures," *Proc Nat Acad Sci*, 114(33), 8734–8739, 2017.
75. J. S. Huang, W. I. Goldburg, and A. W. Bjerkaas, "Study of phase separation in a critical binary liquid mixture: spinodal decomposition," *Phys Rev Lett*, 32(17), 921, 1974.
76. J. M. Twenge and J. D. Foster, "Birth Cohort Increases in Narcissistic Personality Traits Among American College Students, 1982–2009," *Soc Psychol Person Sci*, 1(1), 99–106, 2010.
77. M. D. Huntington, C. J. Engel, A. J. Hryn, and T. W. Odom, "Polymer nanowrinkles with continuously tunable wavelengths," *ACS Appl Mater Interfaces*, 5(13), 6438–6442, 2013.
78. J. W. Cahn, "Phase separation by spinodal decomposition in isotropic systems," *The J Chem Phys*, 42(1), 93–99, 1965.
79. S. Yu *et al.*, "Characterization and design of functional quasi-random nanostructured materials using spectral density function," *J Mech Design*, 139(7), 071401, 2017.
80. R. Bostanabad, A. T. Bui, W. Xie, D. W. Apley, and W. Chen, "Stochastic microstructure characterization and reconstruction via supervised learning," *Acta Materialia*, 103, 89–102, 2016.
81. R. Bostanabad, W. Chen, and D. Apley, "Characterization and reconstruction of 3D stochastic microstructures via supervised learning," *J Microscopy*, 264(3), 282–297, 2016.
82. D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach learn*, 6(1), 37–66, 1991.
83. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521(7553), 436–444, 2015.
84. R. Cang and M. Y. Ren, "Deep Network-Based Feature Extraction and Reconstruction of Complex Material Microstructures," in *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2016, pp. V02BT03A008–V02BT03A008.
85. R. Cang, Y. Xu, S. Chen, Y. Liu, Y. Jiao, and M. Y. Ren, "Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design," *J Mech Design*, 139(7), 071404, 2017.
86. Doreswamy, "A survey for data mining frame work for polymer matrix composite engineering materials design applications," *Int J Comput Intelli Sys*, 1(4), 313–328, 2008.
87. J. C. Loehlin, *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Lawrence Erlbaum Associates Publishers, 1998.

88. B. Thompson, *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. American Psychological Association, 2004.
89. J. F. Hair Jr, G. T. M. Hult, C. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage Publications, 2016.
90. M. Roy, J. Nelson, R. MacCrone, L. Schadler, C. Reed, and R. Keefe, "Polymer nanocomposite dielectrics-the role of the interface," *IEEE Trans, Dielectr Electric Insulation*, 12(4), 629–643, 2005.
91. T. Ramanathan, H. Liu, and L. Brinson, "Functionalized SWNT/polymer nanocomposites for dramatic property improvement," *J Polym Sci Part B: Polym Phys*, 43(17), 2269–2279, 2005.
92. I. Alig *et al.*, "Establishment, morphology and properties of carbon nanotube networks in polymer melts," *Polymer*, 53(1), 4–28, 2012.
93. D. Paul and L. M. Robeson, "Polymer nanotechnology: nanocomposites," *Polymer*, 49(15), 3187–3204, 2008.
94. B. J. Ash, R. W. Siegel, and L. S. Schadler, "Mechanical behavior of alumina/poly (methyl methacrylate) nanocomposites," *Macromolecules*, 37(4), 1358–1369, 2004.
95. J. Jordan, K. I. Jacob, R. Tannenbaum, M. A. Sharaf, and I. Jasiuk, "Experimental trends in polymer nanocomposites — A review," *Mater Sci Eng: A*, 393(1–2), 1–11, 2005.
96. S. S. Ray and M. Okamoto, "Polymer/layered silicate nanocomposites: a review from preparation to processing," *Prog Polym Sci*, 28(11), 1539–1641, 2003.
97. K. Yano, A. Usuki, A. Okada, T. Kurauchi, and O. Kamigaito, "Synthesis and properties of polyimide–clay hybrid," *J Polym Sci Part A: Polym Chem*, 31(10), 2493–2498, 1993.
98. A. Zhu, A. Cai, J. Zhang, H. Jia, and J. Wang, "PMMA-grafted-silica/PVC nanocomposites: Mechanical performance and barrier properties," *J Appl Polym Sci*, 108(4), 2189–2196, 2008.
99. A. Ophir, A. Dotan, I. Belinsky, and S. Kenig, "Barrier and mechanical properties of nanocomposites based on polymer blends and organoclays," *J Appl Polym Sci*, 116(1), 72–83, 2010.
100. T. Villmow, B. Kretzschmar, and P. Pötschke, "Influence of screw configuration, residence time, and specific mechanical energy in twin-screw extrusion of polycaprolactone/multi-walled carbon nanotube composites," *Compos Sci Technol*, 70(14), 2045–2055, 2010.
101. G. Kasaliwal, "Analysis of multiwalled carbon nanotube agglomerate dispersion in polymer melts," 2011.
102. B. Natarajan, Y. Li, H. Deng, L. C. Brinson, and L. S. Schadler, "Effect of interfacial energetics on dispersion and glass transition temperature in polymer nanocomposites," *Macromolecules*, 46(7), 2833–2841, 2013.
103. M. Wang, "Developing bioactive composite materials for tissue replacement," *Biomaterials*, 24(13), 2133–2151, 2003.
104. M. Ahmed, "Coloring of plastics: Theory and practice," 1979.
105. P. Hartley and G. Parfitt, "Dispersion of powders in liquids. 1. The contribution of the van der Waals force to the cohesiveness of carbon black powders," *Langmuir*, 1(6), 651–657, 1985.
106. Y. Wang and W. C. Lee, "Interfacial interactions in calcium carbonate–polypropylene composites. 2: Effect of compounding on the dispersion and the impact properties of surface-modified composites," *Polymer compos*, 25(5), 451–460, 2004.

107. R. Socher, B. Krause, M. T. Müller, R. Boldt, and P. Pötschke, “The influence of matrix viscosity on MWCNT dispersion and electrical properties in different thermoplastic nanocomposites,” *Polymer*, 53(2), 495–504, 2012.
108. H. Yamada, I. Manas-Zloczower, and D. Feke, “Observation and analysis of the infiltration of polymer liquids into carbon black agglomerates,” *Chem Eng Sci*, 53(11), 1963–1972, 1998.
109. P. Levresse, I. Manas-Zloczower, D. Feke, Y. Bomal, and D. Bortzmeyer, “Observation and analysis of the infiltration of liquid polymers into calcium carbonate agglomerates,” *Powder Technol*, 106(1–2), 62–70, 1999.
110. E. W. Washburn, “The dynamics of capillary flow,” *Phys Rev*, 17(3), 273, 1921.
111. T. Lozano, P. Lafleur, M. Grmela, and C. Thibodeau, “Effect of filler dispersion on polypropylene morphology,” *Polym Eng Sci*, 44(5), 880–890, 2004.
112. P. Atkins, *Physical Chemistry: Thermodynamics, structure, and change*. Macmillan Higher Education, 2014.
113. R. Gendron and D. Binet, “State of dispersion: Polypropylene filled with calcium carbonate,” *J Vinyl Additive Technol*, 4(1), 54–59, 1998.
114. H. Zhao, X. Li, Y. Zhang, L. S. Schadler, W. Chen, and L. C. Brinson, “Perspective: NanoMine: A material genome approach for polymer nanocomposites analysis and design,” *APL Mate*, 4(5), 053204, 2016.
115. R. J. Good and L. Girifalco, “A theory for estimation of surface and interfacial energies. III. Estimation of surface energies of solids from contact angle data,” *The J Phys Chem*, 64(5), 561–565, 1960.
116. F. W. Starr, J. F. Douglas, and S. C. Glotzer, “Origin of particle clustering in a simulated polymer nanocomposite and its impact on rheology,” *The J Chem Phys*, 119(3), 1777–1788, 2003.
117. V. Khoshkava and M. Kamal, “Effect of surface energy on dispersion and mechanical properties of polymer/nanocrystalline cellulose nanocomposites,” *Biomacromolecules*, 14(9), 3155–3163, 2013.
118. T. G. Mezger, *The Rheology Handbook: For Users of Rotational and Oscillatory Rheometers*. Vincentz Network GmbH & Co KG, 2006.
119. C. I. Chung, “Extrusion of polymers,” *Theory and Practice*, Hanser Publishers, Munich, 2000.
120. E. Lai and D. W. Yu, “Modeling of the plasticating process in a single-screw extruder: A fast-track approach,” *Polym Eng Sci*, 40(5), 1074–1084, 2000.
121. J. Kittler and J. Illingworth, “Minimum error thresholding,” *Pattern Recogn*, 19(1), 41–47, 1986.
122. J. S. Jang, B. Bouveret, J. Suhr, and R. F. Gibson, “Combined numerical/experimental investigation of particle diameter and interphase effects on coefficient of thermal expansion and young’s modulus of SiO2/epoxy nanocomposites,” *Polym Compos*, 33(8), 1415–1423, 2012.
123. S. Yu, S. Yang, and M. Cho, “Multi-scale modeling of cross-linked epoxy nanocomposites,” *Polymer*, 50(3), 945–952, 2009.
124. Z. Liu, M. Bessa, and W. K. Liu, “Self-consistent clustering analysis: An efficient multi-scale scheme for inelastic heterogeneous materials,” *Comput Meth Appl Mech Eng*, 306, 319–341, 2016.
125. P. F. Brune *et al.*, “Direct Measurement of Rubber Interphase Stiffness,” *Macromolecules*, 49(13), 4909–4922, 2016.
126. X. Cheng, K. W. Putz, C. D. Wood, and L. C. Brinson, “Characterization of local elastic modulus in confined polymer films via AFM indentation,” *Macromole Rapid Commun*, 36(4), 391–397, 2015.

127. M. G. Todd and F. G. Shi, "Validation of a novel dielectric constant simulation model and the determination of its physical parameters," *Microelectr J*, 33(8), 627–632, 2002.
128. P. Maity, N. Gupta, V. Parameswaran, and S. Basu, "On the size and dielectric properties of the interphase in epoxy-alumina nanocomposite," *IEEE Trans Dielectr Electr Insul*, 17(6), 1665–1675, 2010.
129. H. Deng *et al.*, "Utilizing real and statistically reconstructed microstructures for the viscoelastic modeling of polymer nanocomposites," *Compos Sci Technol*, 72(14), 1725–1732, 2012.
130. D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *J Global Optimiz*, 21(4), 345–383, 2001.
131. D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J Global Optimiz*, 13(4), 455–492, 1998.
132. M. Schonlau, W. J. Welch, and D. R. Jones, "Global versus local search in constrained optimization of computer models," *Lecture Notes-Monograph Series*, 1998, pp. 11–25.
133. S. Virtanen *et al.*, "Dielectric breakdown strength of epoxy bimodal-polymer-brush-grafted core functionalized silica nanocomposites," *IEEE Trans Dielectr Electr Insul*, 21(2), 563–570, 2014.
134. Y. Huang *et al.*, "Prediction of interface dielectric relaxations in bimodal brush functionalized epoxy nanodielectrics by finite element analysis method," in *2014 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, IEEE, 2014, pp. 748–751.
135. A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliab Eng Sys Safety*, 91(9), 992–1007, 2006.
136. J. P. Kleijnen, "Kriging metamodeling in simulation: A review," *Euro J Oper Res*, 192(3), 707–716, 2009.
137. M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, 2012.
138. V. Picheny, T. Wagner, and D. Ginsbourger, "A benchmark of kriging-based infill criteria for noisy optimization," *Struct Multidiscipl Optimiz*, 48(3), 607–626, 2013.