Computational Materials Science 211 (2022) 111491

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full Length Article

Designing active layer of organic solar cells using multi-fidelity molecular simulations and spectral density function

Umar Farooq Ghumman [a,1], Anton van Beek [b,1], Joydeep Munshi [c], TeYu Chien [d], Ganesh Balasubramanian [e], Wei Chen [f,*]

[a] Department of Mechanical Engineering, Northwestern University, 60208 Evanston, IL, United States
[b] Department of Mechanical Engineering, Northwestern University, 60208 Evanston, IL, United States
[c] Argonne National Laboratory, Lemont, IL, United States
[d] Department of Physics and Astronomy, University of Wyoming, 82071 Laramie, WY, United States
[e] Department of Mechanical Engineering and Mechanics, Lehigh University, 18015 Bethlehem, PA, United States
[f] Department of Mechanical Engineering, Northwestern University, 60208 Evanston, IL, United States

ARTICLE INFO

Keywords:
Organic solar cells
OPVC
Coarse-grained molecular dynamics
CGMD
Spectral density function
SDF
Material design
Characterization and reconstruction

ABSTRACT

Molecular dynamics simulations have shown substantial promise in the design of organic photovoltaic cells (OPVC). Despite their potential, the utility of molecular dynamics simulations when designing an OPVC is often limited due to their considerable computational cost and their limited prediction accuracy. To address these challenges, we introduce a three-step multi-fidelity design framework that enables a designer to efficiently explore the space of admissible processing conditions, using coarse-grained molecular dynamics (CGMD) simulations, to identify the optimal OPVC design. Using a novel spectral density based approach to reconstruct microstructures of variable size, the framework is able to sequentially search for the globally optimal microstructure using a low-fidelity CGMD simulation with a smaller window size, followed by the optimization of the processing conditions using the high-fidelity simulation. The division in two steps and two fidelities enables the optimization of CGMD simulations at previously intractable lengths and timescales. We validate our results by demonstrating that the CGMD model predictions are consistent with physical experiments reported in the literature and corroborate that the computational complexity is reduced by one order of magnitude.

1. Introduction

Organic photovoltaic cells (OPVCs) have come a long way from having a power conversion efficiency (PCE) of 1% in 2014 [1] to 18.4% in 2021 [2]. While this is still less than the 27% that can be achieved with inorganic solar cells [3], OPVCs have additional properties that make them a competitive alternative (e.g., lighter weight, higher flexibility, lower cost, and environmentally friendly). Because of these advantages, OPVCs are deemed a promising candidate for the next generation of solar cells [4]. In addition, the recent increasing trend of PCE achieved with OPVCs based on non-fullerene acceptors [5,6] is an indicator that perhaps there is still significant potential for improvement. However, to access this latent potential, novel design methods are required to explore the space of alternative OPVC compositions and processing conditions. In this paper, we are concerned with the use of molecular dynamics simulations to explore the PCE performance of OPVCs manufactured through spin coating. Within this process, we consider two processing conditions as our design variables, (i) material composition as the fraction of PCBM in the mixture of PCBM and P3HT, and (ii) annealing temperature. The main reasons we choose PCBM/P3HT over other available organic molecules are, (i) the molecular dynamics simulation for these molecules are mature [7,8]; and (ii) there are plenty of experimental results in literature that can be used for validation. The framework introduced in the paper can be applied to other molecular combinations.

Viewing the optimization of OPVCs as a design problem enables the use of a plethora of methods, specifically interesting are those developed under the Materials Genome Initiative (MGI). The MGI provided a driver

* Corresponding author.
E-mail addresses: umarghumman2018@u.northwestern.edu (U.F. Ghumman), antonvanbeek2022@u.northwestern.edu (A. van Beek), jmunshi@anl.gov (J. Munshi), tchien@uwhy.edu (T. Chien), bganesh@lehigh.edu (G. Balasubramanian), weichen@northwestern.edu (W. Chen).
[1] Shared co-first authors.

https://doi.org/10.1016/j.commatsci.2022.111491
Received 29 January 2022; Received in revised form 28 April 2022; Accepted 1 May 2022
Available online 17 May 2022

for the scientific community to leverage computational resources to expedite the discovery of new materials at a fraction of the cost compared to traditional methods. These computational tools not only augment the generated knowledge with additional insights but also facilitate the identification of new materials with superior properties. One example where these computational tools have provided a noticeable impact is the elucidation of how processing conditions influence a material's properties [9]. While the conceptual framework has been established, the lack of data and computational resources often inhibits its successful implementation. One approach to address this limitation is to leverage the microstructure information to construct a causal link between the processing conditions, the material's structure, and its properties [10,11]. Not only are the microstructures indicative of a material's properties [12–14], but it also provides physical insights [15] to further benefit the design process. This sequence of processing conditions, microstructure, and property is commonly referred to as the process-structure-property (PSP) link.

In literature, the consideration of the microstructure in the design framework has been called microstructure sensitive design (MSD) [14]. The goal of MSD is to identify new materials that either satisfy a prespecified target property or maximize the target property. The MSD framework has been successfully applied to design a wide variety of materials, including polymers [16], composites [17], metals [18], and alloys [19]. In a previous study [16], we introduced an MSD framework that relied on Spectral Density Function (SDF) to use microstructure information for the design of the active layer of an OPVC and maximize its PCE. The motivation for using SDF to characterize the microstructure is to reduce its dimensionality. In addition, SDF holds two major advantages over other microstructural characterization techniques: (i) its ability to be approximated by a parametric function, and (ii) its fast microstructure reconstruction capability [20]. However, a limitation of including SDF in the design process is that its parameters need to be constrained to represent microstructures that can be manufactured from admissible processing conditions. While this can often be achieved by analyzing microstructure images obtained from imaging techniques (i.e., scanning tunneling microscopy), the acquisition of these images is typically too monetary intensive, and it can even be noisy because of the poor contrast in the rendered images.

Here, we present a design framework that circumvents the need for expensive experimental images by simulating the influence of processing conditions on the active layer of an OPVC through coarse-grained molecular dynamics (CGMD) [21–23]. Compared to all-atom molecular dynamic (AAMD) simulations, CGMD provides a significant computational advantage by lumping repeating groups of atoms into beads. However, even with this computational advantage, the maximum length scales of CGMD simulations (less than 60 nm [24–26]) are generally much smaller than physical experiments. For example, in a previous study [27], we explored the OPVC performance using CGMD to simulate $20 \times 20 \times 20$ nm^3 systems and found that our results were not consistent with experiments reported in the literature. The reason for this is that some physical phenomena were not accounted for as they manifest at larger length scales. In this manuscript, we built on these results and show that by decreasing the simulation time steps and increasing the total number of interaction sites in a $100 \times 100 \times 100$ nm^3 system along with a variable cutoff scheme we can achieve CGMD predictions that are consistent with physical experiments.

To address the computational burden of simulating larger systems, we introduce a three-step multi-fidelity design framework. Specifically, we leverage the insight that our structure to property model, uses the microstructure reproduced from SDF parameters as well as the material composition to provide the incident photon to converted electron (IPCE) ratio. Additionally, we introduce a novel technique that includes the SDF of a sample to reconstruct statistically equivalent microstructures of variable sizes. Consequently, we can first identify the optimal material composition by maximizing the IPCE with respect to the material composition and the range of admissible SDF parameters. Where the admissible range of the SDF parameters can be approximated from low-fidelity CGMD simulations (evaluation of $20 \times 20 \times 20$ nm^3 systems). Second, we identify the optimal annealing temperature by maximizing the PSP link using our high-fidelity CGMD simulations (evaluation of $100 \times 100 \times 100$ nm^3 systems). Note that we achieved a considerable computational speedup as: (i) we optimized our high-fidelity CGMD model with respect to only one processing condition, and (ii) the CGMD simulations at one fixed composition and multiple annealing temperatures require considerably less time than performing similar number of simulations at both different compositions and annealing temperature. The novelty of the framework is three-fold: (i) we present large-scale ($100 \times 100 \times 100$ nm^3) simulations whose performance predictions conform with experiments, (ii) we introduce a multi-fidelity design framework that allows us to explore the properties of an OPVC at length scales previously intractable with CGMD simulations, and (iii) we propose a scheme for enhancing the resolution of SDF to reconstruct microstructures of variable sizes.

The remainder of the paper is organized as follows. First, we introduce microstructure characterization and reconstruction through SDF and elaborate on how this is used in the three-step multi-fidelity design framework (Section 2). Next, we elaborate on the developed CGMD simulations and the effect of increasing the interaction distance between particles (Section 3). Subsequently, we present the obtained results when using the introduced design framework and CGMD model to design an OPVC (Section 4). Finally, we conclude this paper by summarizing the presented contributions and delineating directions for future work (Section 5).

2. SDF-based microstructure design framework

In this section, we start by introducing an SDF-based approach to reduce the dimension of microstructure design representation based on images. Then we briefly talk about reconstructing microstructure of variable size using SDF. Lastly, we introduce the three-step multi-fidelity design framework in Section 2.4.

2.1. Introduction to SDF

Microstructure characterization and reconstruction (MCR) [20] is an essential step in the MSD framework as it reduces the dimensionality of the microstructures to facilitate data-efficient exploration of design alternatives. To characterize the microstructure of any material using SDF, a two-phase digital image (2D or 3D; grayscale or colored) representing a material's fine-scale morphology is required. These images can be acquired through high-resolution imaging techniques (e.g., scanning electron microscopy), or simulation models (e.g., molecular dynamics). Typically, the pixel values of the acquired images range from 0 to 255, and they need to be binarized into a matrix $\boldsymbol{X}$. The SDF is then computed as the square of the magnitude of its Fourier transform.

$$\rho(\boldsymbol{\kappa}) = |F(\boldsymbol{X})|^2, \tag{1}$$

where $F(\cdot)$ represents the Fourier transform operator, and $\boldsymbol{\kappa}$ is a vector denoting frequency. If the image is isotropic (i.e., it has no directional dependence) then the SDF can be averaged radially over the frequency domain to retrieve a one-dimensional functional representation of a microstructure's morphology. For more details on SDF for microstructure characterization and reconstruction, we refer the reader to [20]. The advantage of converting a microstructure image into its frequency space is that the SDF can often be approximated by a function with a manageable number of variables. In some examples, the microstructure complexity can be reduced to as little as one [28] or two [16] parameters. The periodicity of the microstructure features dictates how many parameters are required to accurately approximate its SDF. It is worth noting that SDF has a one-to-one correlation with the statistical two-

point correlation function and therefore captures the same material features. Even though SDF cannot capture complex microstructure features that require three-point or higher correlations functions, it has proven to be a powerful tool for the design of new material systems [16,28,29].

The second aspect of MCR methods is their ability to reconstruct statistically equivalent microstructures. We consider the microstructure to be a signal which is manipulated using linear shift invariant system to produce the statistically equivalent microstructure (for more details the reader is referred to [30]). Compared to alternative tools (e.g., descriptors [31–33] or statistical correlation functions [34]), reconstruction using the SDF is highly efficient. For example, a 400×400-pixel^2 image can be reconstructed in less than a second using its SDF. In contrast, descriptor-based methods can take minutes to hours (depending upon the volume fraction). Efficient reconstruction is critical for material design as optimization typically requires many reconstructions.

For visualization, an example of MCR of a two-dimensional microstructure is presented in Fig. 1. Note that in the target image (Fig. 1A), there is a strong underlying spatial correlation between the two phases and is also observed in its corresponding SDF by the spike isolated at a specific range of frequencies (Fig. 1B). The SDF of this example can be approximated through a step function with only three parameters (the amplitude, start, and endpoint of the rectangular function). The approximated SDF can then be used to produce a statistically equivalent image. The reconstructed image (Fig. 1C) is statistically equivalent to the original image as it shares the same features as the target image. For example, the length of the convoluted channels can be observed to be identical to the original image. In this research, a double-peak SDF function is identified from the microstructure images obtained from CGMD simulations. More on this will be explained in Section 4.1.

2.2. Microstructure enlargement by enhancing SDF resolution

A benefit of using statistical methods such as SDF for material design is that the significant features of a microstructure can be captured from a smaller sample of a larger microstructure image. Given that the smaller sample is large enough to capture to material features. To elaborate, the details inside this smaller sample should include all features of the larger microstructure image. Otherwise, if there are long-range features which are not present in the small sample, then this will lead to an inaccurate characterization. As shown in Fig. 2, the SDF of a large microstructure and the smaller sample have the same approximated shape i.e., a step function, with the same width. Only the magnitudes differ and is inconsequential for reconstruction. Consequently, we observe how a smaller microstructure can be used for characterization and reconstruction.

To reconstruct statistically equivalent microstructures of variable size, SDF is constrained by the sampling frequency. The sampling frequency depends on the number of samples, and that in turn depends on the number of pixels of an image. Taking examples from Fig. 2, the original image of the microstructure has 200 pixels in each direction, and the smaller one has only has 100 pixels in each direction. As a result, SDF has 100 frequency bins from original image, and 50 frequency bins from the smaller image. As per the Nyquist frequency limit, the maximum frequency for both is: $f_{max} = \frac{1}{2\Delta}$, where Δ is the physical distance represented by one pixel in the microstructure image. Since Δ is same for both images, f_{max} will also be the same. However, because the number of bins is different, each of these bins has a size of 3.33×10^{-6} nm^{-1} and 6.67×10^{-6} nm^{-1}, for the larger and smaller image respectively. Because of the difference in the two SDFs, if we were to reconstruct an image using the SDF from a smaller sample, we would only be able to reconstruct a microstructure of similar size i.e., 50 *nm*.

To overcome this limitation, we propose an SDF frequency enhancing technique to reconstruct microstructure with variable size from the original sample image. Since the final size of the microstructure depends on the frequency-resolution of the SDF function, we propose a new technique to enhance it. Specifically, by duplicating the elements in each frequency bin multiple times to add the missing information. As seen in Fig. 3A, we repeat each element in a bin two times to extend the SDF to double its size. For the given example it means that we have divided the increment of 6.67e−6 nm^{-1} into two bins of 3.33e−6 nm^{-1}, and thus doubled the number of total bins to 100. If we compare this new extended SDF with the original SDF of the entire image from Fig. 2A, we see that their SDF are very similar Fig. 3B.

To demonstrate the efficacy of this approach, we present a 3D microstructure of the OPVC in Fig. 4. We start by taking a 20 *nm* CGMD simulation and characterize its microstructure using SDF. Then we extend the SDF by a factor of four and use it for reconstruction. The reconstructed microstructure is thus $20 \times 4 = 80$ *nm*. If we crop a 20 *nm* sample from the large microstructure and compare its SDF, we can observe that it looks identical to the original SDF of the 20 *nm* simulation. Hence in this section we validated the efficacy of using frequency enhancement to reconstruct microstructures of variable size from a single microstructure image. Note that a smaller reconstruction can always be achieved by taking a smaller section from a larger reconstruction.

A caveat with this approach is that we are extending the SDF curve by trying to add artificial information for the small frequency indices so there is some approximation of the SDF curve of the larger sample. Thus, these reconstructions cannot entirely replace the larger simulations. Another approach to extend the SDF could be to interpolate between the missing frequencies to get a smoother curve, but the authors believe that the result would only be slightly different.

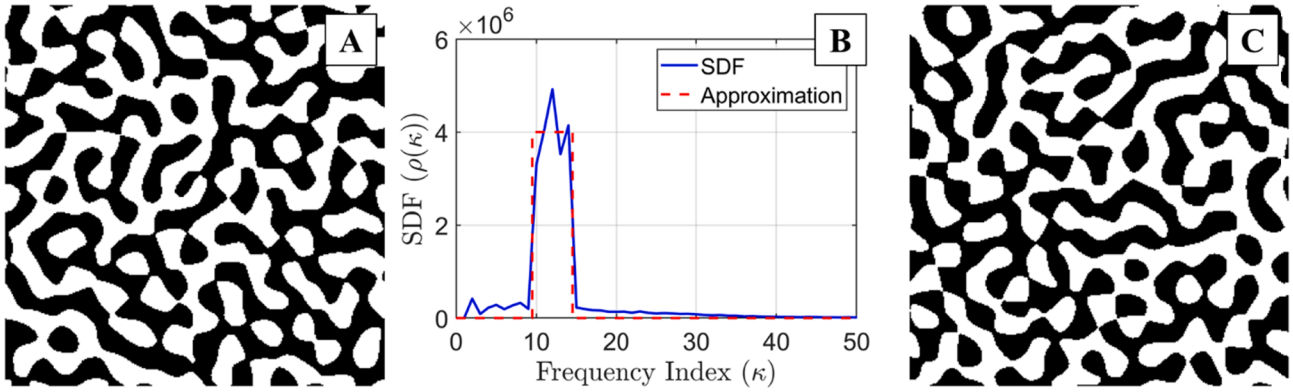

Fig. 1. Characterization and reconstruction of a microstructure through SDF. The target microstructure (A) has prominent spatial features that are represented by the SDF (B) through an approximate rectangular function. In addition, the SDF can be used to reconstruct statistically equivalent microstructures (C).

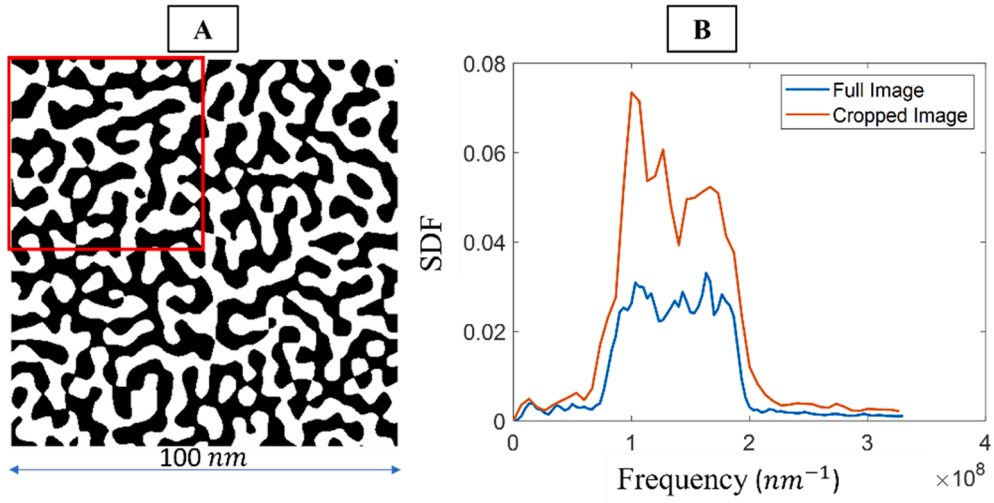

Fig. 2. Comparison of the SDF at different length scales of the same material. The SDF of the full microstructure (A) is represented by the blue function (B), and the SDF for the red inlay (A) is represented by the red function (B). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

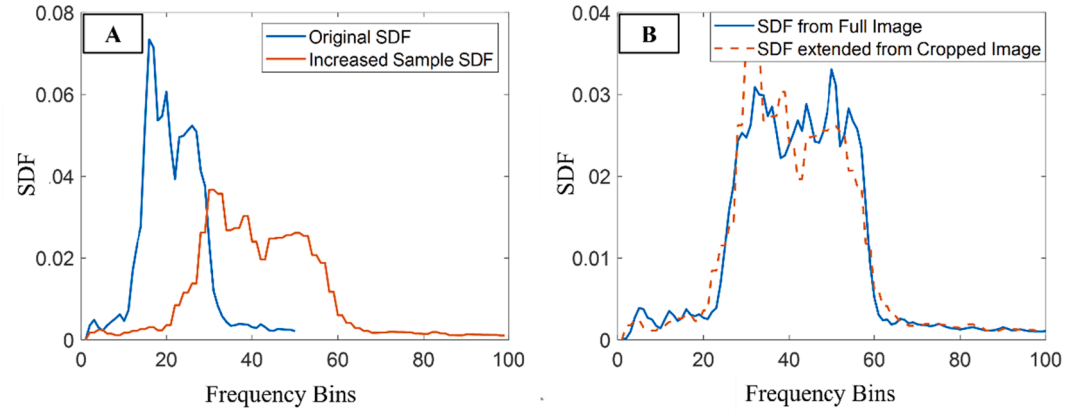

Fig. 3. SDF frequency resolution enhancement. (A) Duplicating values in each bin of SDF to increase the number of frequency bins by two. (B) Comparing the SDF of the original full image of microstructure with the enhanced SDF of the cropped image of microstructure. Note that x-axis here is only the index of frequency bin and not the frequency itself.

2.3. SDF-based design framework

Under the framework of MSD [14], materials are viewed as a complex structural systems that can be optimized to achieve improved performance. In this section, we present an SDF based microstructure design framework (Fig. 2) that can be employed for the design of quasi-random microstructural systems considering its PSP relation. For the framework discussion that follows, we select OPVC as our material of interest.

The key idea of the framework (Fig. 5) is to use SDF to reduce the dimensionality of the OPVC microstructures and enable inverse design through forward PSP prediction. The framework begins by fabricating a set of samples by changing the processing conditions such that the produced samples are representative for the space of admissible processing conditions. Next, microstructure details are captured by imaging techniques that enable the microstructure to be characterized by SDF. A crucial aspect of SDF is that it can often be parameterized by a handful of parameters (2 to 4 for certain materials). This allows us to link the microstructure to processing conditions (i.e., process-structure mapping). Another benefit of SDF is that it enables us to efficiently reconstruct statistically equivalent representative volume elements (RVEs).

To evaluate the performance of an RVE, a model is developed that accounts for structural features in addition to device physics and material properties. For OPVCs, the core performance indicator is IPCE. This can be achieved by using a computational tool that predicts the IPCE value for a given microstructure $\boldsymbol{X}$ [16]. This tool computes the influence of microstructure on known physical phenomenon of light conversion to establish the structure-property relationship that forms the basis for performance optimization. Depending on the size of the RVE and its resolution, the computational cost of the performance model could be considerable. For example, an RVE having 1000 pixels in each of the 3 dimensions can take up to 1 h to compute. To mitigate this cost, a surrogate model can be built and used for design optimization.

Design optimization is performed with the pre-determined design variables: volume fraction and parameterized SDF variables. The output from the design optimization are the optimal microstructure parameters corresponding to the best IPCE value. As the optimal microstructure parameters are the SDF parameters, they can be mapped back to the processing parameters using the already established process to structure model.

In the presented SDF based material design framework, the role of SDF is to leverage its characterization, parameterization, and reconstruction ability to enable Process-Structure and Structure-Property mapping. This in turn facilitates the design of materials with quasi-random microstructures. A key assumption of this framework is that the material microstructure can be sufficiently characterized by SDF.

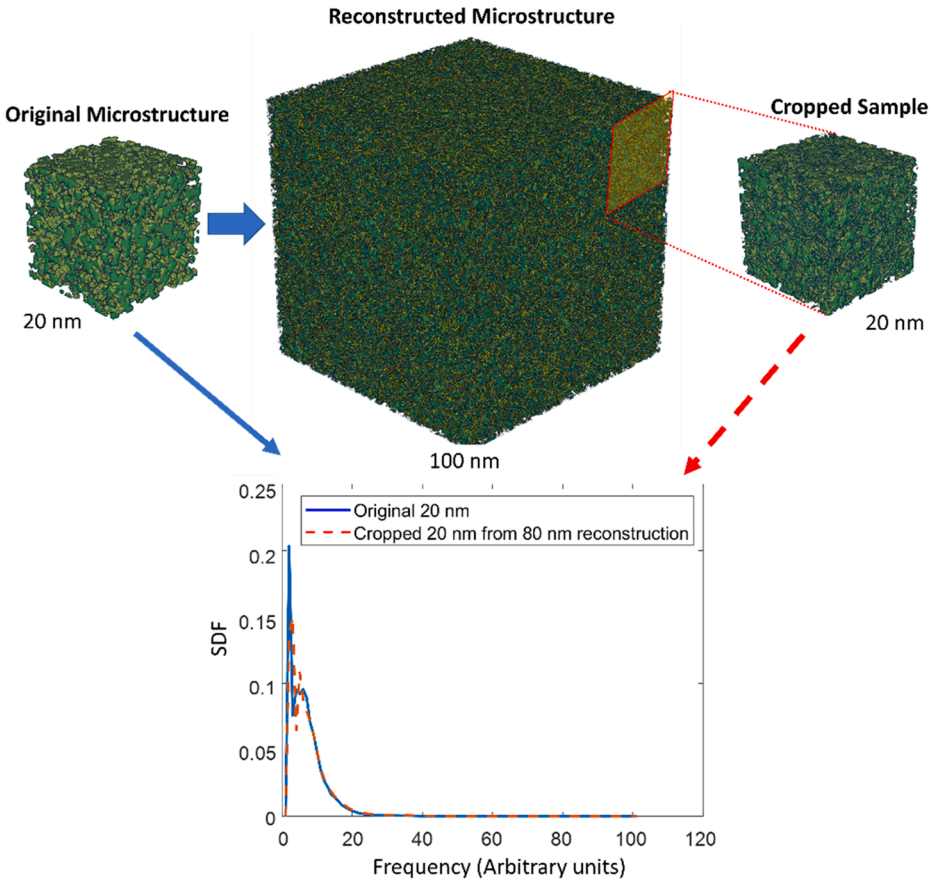

Fig. 4. Visualization of using frequency enhancing to reconstruct statistically equivalent microstructure of variable size from a single observed image. Top: Starting from a small 20 × 20 × 20 nm CGMD simulated microstructure, a large 80 × 80 × 80 nm is reconstructed using SDF frequency resolution enhancement. From the large reconstruction a smaller 20 × 20 × 20 nm sample is cropped for comparison. Bottom: SDF comparison of the original microstructure with cropped microstructure.

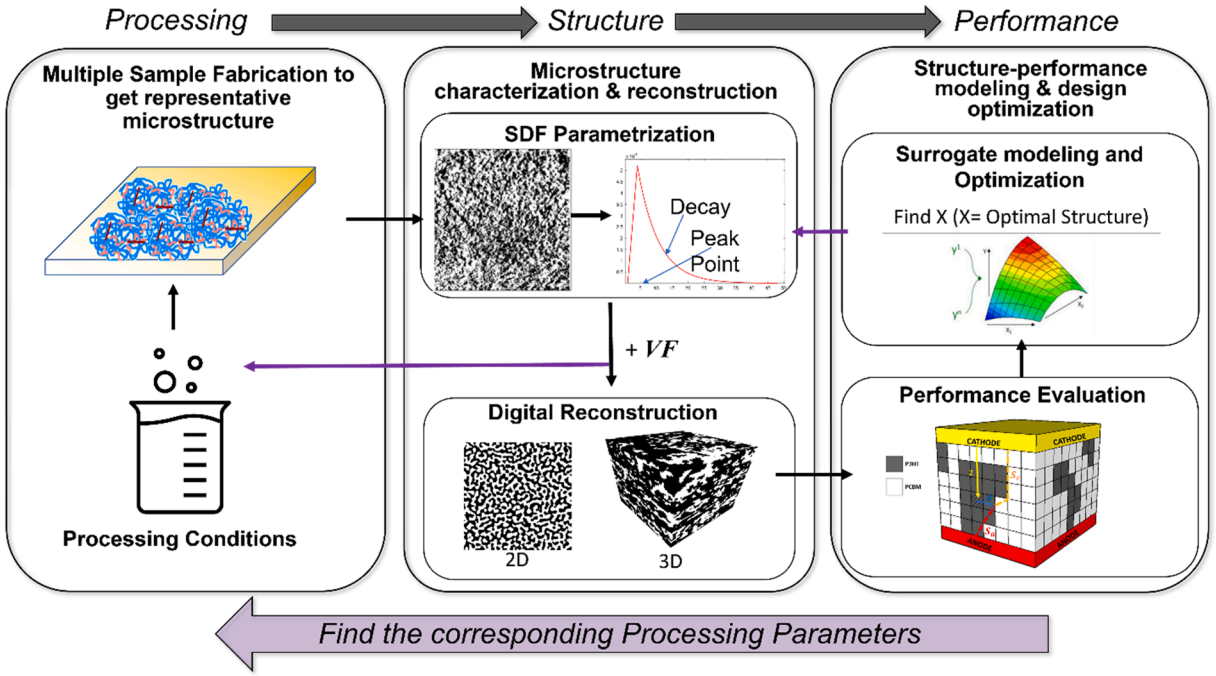

Fig. 5. SDF-based design framework to find optimal processing conditions of any quasi-random material. The framework links processing conditions to a material's performance by considering its microstructure morphology, and is facilitated by SDF to reduce the dimensional representation of these microstructures.

2.4. Three-step multi-fidelity inverse design scheme

Based on the SDF-based design framework, we attempted to find the optimal processing conditions of a solar cell in a previous study [35]. However, using the high-fidelity simulations (i.e., 100 *nm* structures) directly, would be too computationally exhaustive. In this section, we

address this challenge by introducing a three-stage multi-fidelity design scheme that includes the use of low-fidelity simulations (i.e., 20 *nm* structures) to improve tractability as well as high-fidelity simulations to improve computational accuracy.

The PSP linkage as presented in Fig. 6 is an acyclic multimodal design problem [36] that contains three models: (i) low-fidelity process to structure model $\{P,R,D\}^T = \psi_l(T,VF)$ where P,R and D are respectively the SDF peak point, relative high difference of the first two peaks, and decay parameters (Section 4.1), (ii) high-fidelity process to structure model $\{P,R,D\}^T = \psi_h(T,VF)$, and (iii) structure to property model $IPCE = \phi(P,D,VF)$. Subsequently, by solving the optimization problem

$$\max_{T,VF\in\Omega} \phi(\psi_h(T,VF),VF), \tag{2}$$

for the domain $\{T,VF\} \in [50,170]\times[0.25,0.75] = \Omega$, a designer can identify the optimal processing settings. We infer this range by from literature; T usually varies between 50 °C and 170 °C in experiments. However, the range of VF in experiments is a bit narrow in the authors opinion, so a wider range is selected from 0.25 to 0.75 to search the unfamiliar design space. However, optimizing Equation (1) using evolutionary (e.g., genetic algorithm, simulated annealing, and particle swarm) or gradient-based methods (e.g., gradients descent, Newton's methods, and sequential quadratic programming) is unpractical as they would require too many costly function evaluations. A computationally more efficient class of methods is Bayesian optimization, but for problems of this dimension even Bayesian optimization requires dozens of function evaluations in the case of homoscedastic noise [37] and hundreds in the case of heteroscedastic noise [27]. In our problem, we need to consider noise as the process to structure model (i.e., the CGMD simulation model) that has randomized initial conditions that are evaluated over a finite length and timescale.

To reduce the computational complexity, we make two observations: (i) the computational cost (core-hours) of the high-fidelity process to structure model is more than 12 (Table 1) times larger than the low-fidelity model and the structure to property model, and (ii) all simulations modes share the composition input VF. Leveraging this insight, we propose the following three-step multi-fidelity inverse design scheme:

1. Identify the domain ω of intermediate model variables $P,R,D \in \omega$ (i.e., the range of the SDF peak parameter P and decay parameter D) through a small set of samples obtained from the low-fidelity structure to property model $\psi_l(\cdot)$.
2. Identify the optimal composition VF^* by optimizing the structure to property model $\phi(\cdot)$ using the domain of intermediate model variables ω, as identified in Step 0.
3. Identify the optimal annealing temperature T^* by optimizing the high-fidelity linked process to structure to property model $\phi(\psi_h(\cdot),\cdot)$ over a reduced dimensional design space.

Table 1
Computational cost for individual CGMD simulations using low and high-fidelity approaches.

	Total wall time (hrs.)	Number of CPUs (cores)	Total core-hours
Low fidelity (20 nm)	120	560	67,200
High fidelity (100 nm)	300	2800	840,000

Concerning the initial Step 0, we start by creating a design of experiments for the low-fidelity CGMD process to structure simulation $\psi_l(T,VF)$. Specifically, using an optimal Latin hypercube design [38] containing nine samples we uniformly cover the domain of processing settings Ω. By evaluating these samples using the low-fidelity CGMD process to structure simulation we can identify the range of the intermediate modeling variables as $P,R,D \in \omega \subset \mathbb{N}\times\mathbb{R}^2$. Assuming that this range is representative for the intermediate modeling variables, we can identify the optimal composition VF^* and structure $\{P^*,R^*,D^*\}^T$ (Step 1) as

$$VF^*,P^*,R^*,D^* = \underset{\substack{VF\in[0.3,0.8]\\ P,R,D\in\omega}}{\operatorname{argmax}} \phi(P,R,D,VF). \tag{3}$$

With the optimal composition VF^* known we need to identify the optimal annealing temperature T^*. Originally, this would have been a costly procedure as it required the evaluation of the high-fidelity CGMD process to structure simulation at different levels of composition VF and annealing temperature T in the domain Ω. However, by identifying the optimal composition using Equation (2), we can simplify the problem into a one-dimensional optimization (Step 2) as

$$T^* = \underset{T\in[50,170]}{\operatorname{argmax}} \phi(\psi_h(T,VF^*),VF^*). \tag{4}$$

Optimizing Equation (4) will require fewer function evaluations than Equation (1) for two reasons: (i) the search space has been reduced from two to one dimension without changing the roughness of the objective function, and (ii) we only need to create one structure as an input to the high-fidelity CGMD structure to property simulation that we can evaluate at different temperatures T. Although we have significantly reduced the complexity of the optimization problem, the computational cost for evaluating $\psi_h(\cdot)$ is still nontrivial, and so we use a set of nine high-fidelity processes to structure to property simulations, obtained from an optimal Latin hypercube design for $T \in [50,170]$, to train a Gaussian process model $\widehat{f}(\cdot)$. Consequently, by substitution of $\widehat{f}(\cdot)$ into Equation (4) we can find the optimal annealing temperature as

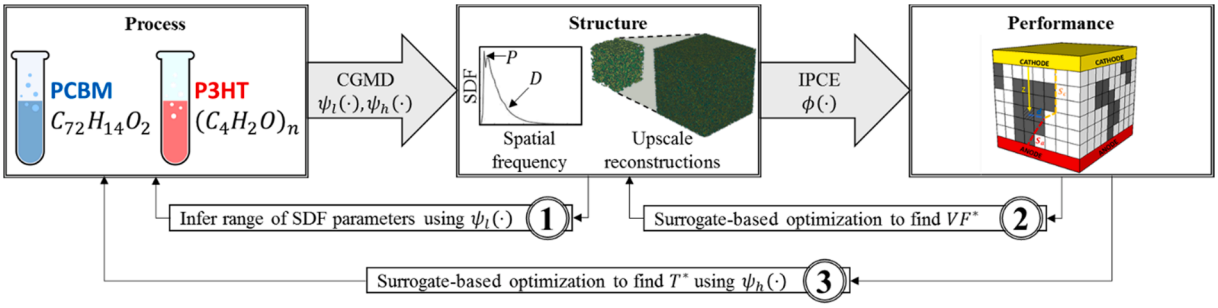

Fig. 6. Three-step multi-fidelity optimization framework. The initialization of the optimization problem is to use the low-fidelity model to infer the range of intermediate structure parameters. Subsequently, in the next step, we find the optimal performance by maximizing the IPCE with respect to the volume fraction VF^* and the intermediate modeling parameters $\{P,D\}$. Finally, the optimal annealing temperature T^* is found by maximizing the entire PSP link using the high-fidelity model.

$$T^* = \underset{T\in[50,170]}{\operatorname{argmax}} \widehat{f}(T). \tag{5}$$

The validity of the three-step multi fidelity inverse design scheme introduced in this section depends on two implicit assumptions. First, by identifying the optimal structure (Step 1), we assume that all structures described by $P, R, D \in \boldsymbol{\omega}$ can be synthesized through processing settings in the domain Ω. This decentralized approach greatly limits the computational cost but selecting a domain that is too small risks converging to a suboptimal design. In contrast, selecting a domain of intermediate modeling variables that is too large risks converging to a structure that cannot be synthesized from the admissible processing settings. Second, it is assumed that the domain of intermediate modeling parameters $\boldsymbol{\omega}$ is independent of composition VF. Specifically, when solving Equation (2), we assume that for any value of the composition VF there exists a structure with any SDF parameters P, R, D in the domain $\boldsymbol{\omega}$. Despite these limitations, the introduced three-step multi-fidelity inverse design scheme can identify the processing settings associated with strong properties through a manageable number of function evaluations.

3. CGMD simulations

Due to the reduced number of total particles (coarse-grained interaction sites) relative to AAMD, CGMD allows us to simulate larger systems (~10–1000 *nm* each side) containing millions of beads (also known as superatom) for longer time scales (~1–10 μs) spanning millions of timesteps. Hence the investigation of morphological features based on different processing conditions such as the annealing temperature and volume fraction (weight ratio) is achievable using chemically informed CGMD simulations [7,8]. One can also investigate the trend in morphological evolution during the typical solvent evaporation and thermal annealing process to construct correlations between processing, structure, and performance. In this work, all the CGMD simulations are carried out using Gromacs molecular dynamics package [39], visualized using VMD [40], and post-processed in MATLAB.

3.1. Multi-fidelity simulations

Low-fidelity CGMD simulations are employed to simulate the blend morphology formation and evolution within a smaller length scale (~20 *nm* each side). Organic semiconducting polymer Poly(3-hexylthiophene) (P3HT) in combination with phenyl-C61-butyric acid methyl ester (PCBM) are initially solvated in Chlorobenzene (CB) solution. Finer-than-traditional CG beads and the corresponding interatomic interactions, as described by Martini force fields [41], are considered to model P3HT, PCBM, and CB molecules. While larger domains simulated for a longer time scale (high-fidelity) can describe the morphology comparable to experiments, performing these high-fidelity simulations would be arduous given the computational time that they require. Additionally, we believe that our low-fidelity simulations, as described in earlier efforts by Munshi et al. [21,42], can unravel the fundamental physical trends in the microstructure evolution driven by the intermolecular interactions at finite temperature. That is why we believe they are sufficient for finding the representative microstructure features to extract the underlying SDF shape (Section 2.4).

3.2. Large scale high fidelity simulations

While the low-fidelity CGMD simulations are an excellent choice to extract microstructural features and provide fundamental insights into the trends in overall exciton diffusion to charge transport process, these are seldom comparable to the experimental characterization. To address this challenge, high-fidelity CGMD simulations are employed to mimic the physical experiments of OPVC design. Akin to the low-fidelity approach, P3HT and PCBM are initially solvated in organic CB solution. Following energy minimization and equilibration at constant temperature and pressure conditions (also referred to as the number of particles, pressure, and temperature equilibration), CB molecules are evaporated gradually from the system to simulate a typical solvent evaporation process. Finally, the 100 *nm* thick solvent-free P3HT: PCBM bulk heterojunction nanomorphology is heated to a specific temperature T (thermal annealing), and gradually cooled to room temperature (T = 25 °C). For all these simulations, the x and y-dimension are kept constant at 100 *nm* while the z-direction (thickness) is initialized at ~800 *nm* for the solvated ternary mixture. The final solvent-free and thermally annealed structure, in absence of the CB solvent molecules, reduce to the thickness of ~100 *nm* leading to simulation box volume of ~100 nm^3. In contrast to the low-fidelity CGMD simulations, large-scale simulations consider large number of interaction site thus making the high-fidelity simulations highly sensitive to the interaction parameters between the CG beads. To stabilize the large-scale systems and accurately capture the dynamics during evaporation and annealing, we varied the Martini bead radius for the P3HT and PCBM molecules along with a subtle variation in the Lennard-Jones (LJ) cutoff distance implemented using Verlet scheme. While in our previous efforts for low-fidelity CGMD simulations [41], we found a LJ cutoff distance of 1.1 *nm* to achieve P3HT: PCBM system density in agreement with experiments, for the high-fidelity CGMD simulations a cutoff distance ~1.4 *nm* was found to be optimum. The increase in the total number of interaction sites and the cutoff distance account for the large difference between the overall computational cost between the two variants of CGMD simulations. Finally, these large-scale high-fidelity simulations, in this work, are directly compared with experimental observations such as cross-sectional scanning tunneling microscope (X/STM) images to investigate microstructure evolution through solution processing techniques, such as spin coating, typically employed in experiments of OPVC design.

3.3. Post-processing through Voronoi tessellation

Both the low-fidelity and high-fidelity simulations output the final positions of all the coarse-grained beads inside the simulation box. This final state of the simulation can be visualized for inspection purposes but it cannot be processed through SDF or the structure to performance model (i.e., the IPCE simulation). To this end, we require a post-processing scheme to convert the vector of coarse-grained bead centers into a binary matrix $\boldsymbol{X}$.

In the proposed scheme, the first step is to import the positioning information of all the beads in the form of cartesian coordinates of the center of all molecules (both phases i.e., P3HT and PCBM). Next, we initiate a three-dimensional square matrix $\boldsymbol{X}$ with a total of n_p elements in each dimension. Moreover, the element $X_{ijk} \in \mathbb{Z}_2, ijk = 1, \cdots, n_p$ at the ijk^{th} index in $\boldsymbol{X}$ is a voxel for which we must assign a material phase. For this material structure, we set the size of each voxel equal to 1 $Å^3$/voxel, this provides an adequate balance between the accuracy of representing the complexity of the microstructure and the computational complexity required to evaluate the IPCE value of the structure. Consequently, for the $100 \times 100 \times 100$ nm^3 CGMD simulation, we end up with a matrix $\boldsymbol{X}$ of size 1000 by 1000 by 1000 voxels.

For visualization, we have shown a two-dimensional slice of a three-dimensional structure in Fig. 7A, where the black circles indicate the centers of the elements in $\boldsymbol{X}$. Next, we place the center locations of the molecules obtained from the CGMD simulations in this structure, as visualized by the colored pentagrams. The color of these pentagrams is associated with the material phase of the CGMD response (i.e., PCBM vs P3HT). Subsequently, through the use of Voronoi tessellation [43], we can assign binary values to the empty elements in $\boldsymbol{X}$ based on the material phase of the nearest molecule. If both the P3HT and PCBM molecules had the same size, then we could directly apply Voronoi tessellation as shown in Fig. 7A. However, as the molecules are different in size, we need to account for this by weighting the distance calculation

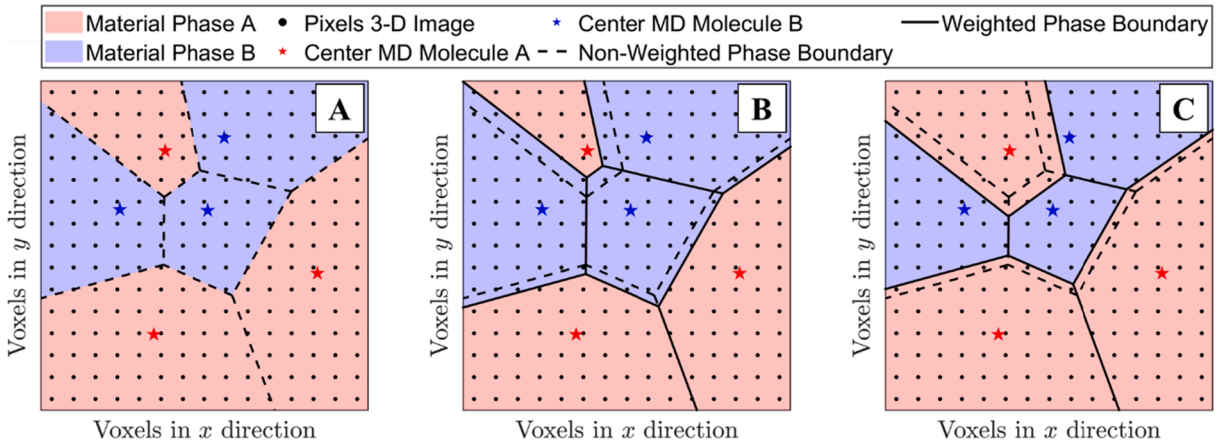

Fig. 7. Voronoi Tessellation for CGMD postprocessing. Using equal weights (A) might result in a volume fraction that is inconsistent with the CGMD inputs. Consequently, the material phase A molecules (B) or phase B molecules (C) can be weighted to be consistent with the volume fraction used as the CGMD model input.

in the Voronoi tessellation according to the specific material phase. The selection of the weight in the Voronoi tessellation processes is guided by matching the volume fraction of the two materials in X with the volume fraction used as an input to the CGMD simulation. In Fig. 7B and 7C we have shown the scenario where the weight in the Voronoi tessellation is adjusted to increase and decrease the volume of material phase B, respectively. Observe how the boundary only changes at the interphase between the two different materials and remains constant between two materials of the same phase.

4. Results and discussion

In this section, we discuss the results obtained when we apply the design framework introduced in Section 2.3 to design an OPVC with optimal IPCE performance. First, we need to identify the range of SDF parameters estimated from a small number of low-fidelity CGMD simulations (Section 4.1). Next, we identify the optimal microstructure morphology by optimizing the structure to property model with respect to the SDF parameters and volume fraction (Section 4.2). Subsequently, we optimize the property with respect to the processing conditions using the high-fidelity CGMD model (Section 4.3). Finally, we conclude this section by comparing the optimal processing conditions with experimental results reported in the literature (Section 4.4).

4.1. Step 1: Inferring the range of SDF parameters

In Section 2.3 it was assumed that the SDF functions could be characterized by a two-dimensional parametric function. This approximate SDF function was established by observing the similarity between the SDF of the microstructures obtained from the nine low-fidelity CGMD simulations, as documented in Table 2. Specifically, in Fig. 8A we have plotted all nine normalized SDF functions (i.e., the area under the plots has been normalized to equal one). Note that we can normalize the SDF function as the reconstruction depends on the relative frequency of material features and not the absolute frequency as discussed in Section 2.2. We extracted the salient features of the nine curves to match them with the properties (IPCE values), using the following approximation

$$\widehat{\rho}(\kappa, D, R, P) = \begin{cases} 0, \kappa = 0 \\ 1 - \dfrac{\kappa - 1}{3}, 0 < \kappa < P \\ Re^{(-D(\kappa - P))}, P \leq \kappa \end{cases} \tag{6}$$

where D is the frequency of the second peak, R is the ratio between the first and second peak, and D is a variable that governs the decay of the SDF at higher frequencies. Note that the lower frequencies contain the largest degree of inconsistency when compared with the target structure in Fig. 5A. However, we believe that this noise influences the IPCE of the reconstructed microstructures only marginally. The reason that the low frequency has limited influence on the IPCE is that they are associated with features that have a long length scale and less samples, and hence are more prone to noise and are less consequential to the reconstructed microstructure morphology. It is for this reason that the exponential decay term of the parametrized SDF $\rho(\cdot)$ captures the most important features. In Fig. 8B we have plotted the SDF of one low-fidelity CGMD structure (blue function) and show that we can get a good fit with the approximate SDF (red dashed function). Note that in this plot we have normalized the plot to show its consistency with the experimental structure. Moreover, minimizing the mean squared error for each of the nine low-fidelity CGMD structures we can find the P, R and D parameters as shown in the fourth, fifth and sixth column of Table 2. Using these intermediate modeling parameters, we reconstructed the initial nine low-fidelity structures and simulated their IPCE. Subsequently, we

Table 2
Design of experiments used to explore the potential microstructures using the low-fidelity CGMD simulation model.

	Process		Structure			Property		
#	VF	T(°C)	P	R	D	$\phi(\psi_1(\mathrm{T},\mathrm{VF}),\mathrm{VF})$	$\phi(\mathrm{P},\mathrm{R},\mathrm{D},\mathrm{VF})$	Relative error (%)
1	0.375	50	5	0.522	0.217	0.3558	0.3649	2.56
2	0.5625	65	4	0.337	0.244	0.3333	0.3379	1.38
3	0.625	170	5	0.385	0.295	0.3133	0.3216	2.65
4	0.25	140	4	1.245	0.224	0.3403	0.3481	2.29
5	0.6875	125	3	1.332	0.321	0.2863	0.2909	1.61
6	0.4375	155	4	0.675	0.213	0.3531	0.3584	1.50
7	0.3125	95	4	0.344	0.259	0.3512	0.3563	1.45
8	0.5	110	4	0.564	0.276	0.3446	0.3494	1.39
9	0.75	80	4	0.271	0.402	0.2543	0.2591	1.89

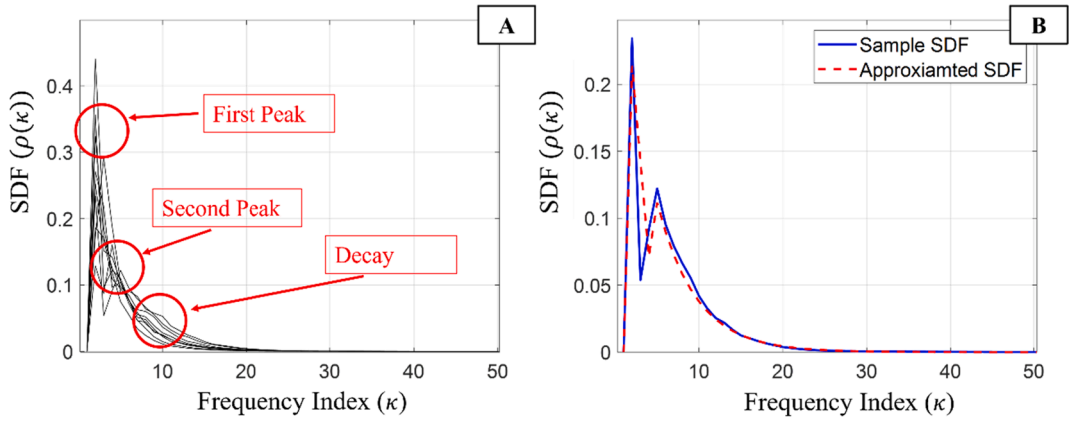

Fig. 8. Parametrization of the SDF of low-fidelity CGMD microstructures. The SDF of the nine microstructures obtained from low fidelity CGMD simulations (A) shows a remarkable consistency among the microstructures. Subsequently, this similarity can be captured through a parametric function that can be used to approximate the original SDFs (B).

calculated the relative error with respect to the IPCE values of the target microstructure responses as shown in the ninth column of Table 2. The maximum observed relative error is $< 3\%$ and as such we can conclude that the parametrized SDF provides an accurate approximation of the microstructures.

4.2. Step 2: Structure-property optimization

To find the optimal microstructure using Equation (3), we need to identify the range of intermediate modeling parameters ω. From the fourth, fifth and sixth column in Table 2 we can infer the range for the intermediate modeling parameters by expanding the design space from the original samples as $P,R,D \in [3,5] \times [0.271, 1.332] \times [0.213, 0.402] = \boldsymbol{\omega}$. Specifically, the second index of the peak frequency variable in the SDF is a natural number in the range $P \in [3,5]$, the exponential decay variable in the SDF is a real number in the range $D \in [0.213, 0.402]$, and the ratio between the first and second peak is a real number in the range $R \in [0.271, 1.332]$. Using the range of the intermediate modeling parameters $\boldsymbol{\omega}$ we can optimize Equation (3), to find the optimal microstructure. Before optimization, we first need to increase the sample size and make it comparable to experiments. The typical thickness of spin coated OPVC is of 200 *nm*, so we use the SDF frequency enhancement (Section 2.2.2) to increase the size of the smaller samples 10^3-fold.

While the SDF improves the efficiency of the microstructure reconstruction, the computational cost for evaluating the IPCE is still not trivial. Consequently, we use a surrogate-based optimization framework by training a Gaussian process model on a set of 30 IPCE function evaluations assigned through an optimal Latin hypercube design. Subsequently, using a multi-start sequential quadratic programming approach we find the optimal microstructure as $\{VF^*, P^*, R^*, D^*\} = \{0.4, 5, 1.33, 0.213\}^T$, for which the Gaussian process model predicts an optimal IPCE value of 0.195. To validate these results, we reconstructed a new microstructure using the optimal microstructure parameters and find an optimal IPCE value of 0.189. Compared with the Gaussian process model prediction, we have a relative error of $<0.4\%$, and thus we can conclude that the Gaussian process is sufficiently accurate to approximate the globally optimal microstructure.

4.3. Step 3: Process-structure-property optimization

Using the optimal volume fraction VF^*, we can identify the remaining optimal processing condition by optimizing the high-fidelity CGMD simulations with respect to a reduced dimensional design space. Specifically, we can use a surrogate model-based optimization approach by training a Gaussian process model on nine equally spaced annealing temperatures and the predicted IPCE values. It should be noted that this means we train a surrogate model to directly predict the IPCE value as a function of the processing conditions. In addition, to run the low-fidelity CGMD simulations we set the volume fraction at $VF^* = 0.4$ as identified in Step 1. The resulting Gaussian process model predictions are shown in Fig. 9. The motivation for using a Gaussian process is that it enables us to account for the prediction uncertainty in the CGMD simulations. Specifically, this can be observed in Fig. 9 from the nonzero prediction uncertainty (blue shaded regions) at the observed training samples. The advantage of Gaussian process over other machine learning methods is its Bayesian assumption on the space of potential functions that provides a statistically rigorous approach to avoid over fitting or underfitting the training data.

From the response surface approximation shown in Fig. 9, we can observe that the globally optimal IPCE value is found at an annealing temperature of T* = 170 °C. In addition, the optimal IPCE value for the identified optimal processing conditions is $IPCE^* = 0.3953$. It should be noted that the optimal IPCE value is higher than what was predicted with the low-fidelity model. This is likely a result of the difference in the size of the simulated structures (100 *nm*, and 200 *nm*), as thinner structures are expected to trap the light more effectively. Since we are interested in the trend rather than the exact value of IPCE for optimization, it is less consequential. Nevertheless, the introduced three-step multi-fidelity design framework facilitates microstructure sensitive design using previously intractable models. Since the globally optimal IPCE was found at an already simulated annealing temperature we do not need further validation. Lastly, the SDF parameters ($P^* = 4, D^* =$

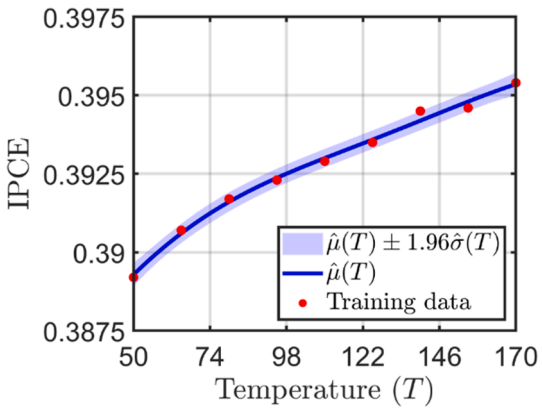

Fig. 9. Gaussian process model prediction for the IPCE as a function of temperature. The high-fidelity CGMD simulations used the optimal volume fraction $VF^* = 0.40$ as identified in Step 1 of the introduced design framework.

$0.36, R^* = 0.33$) of the optimal CGMD simulation are also well within the design space ω. This is evidence of the validity of step 1, where smaller CGMD simulations were used to extract the range of SDF parameters.

4.4. Validation of results

To validate the findings of the introduced three-step multi-fidelity design framework we compare the optimal processing conditions (i.e., T^* and VF^*) with those reported in the experimental literature. First, we note that the identified optimal volume fraction $VF^* = 0.4$ is relatively similar to 0.8:1 PCBM:P3HT (0.38) loading as reported in [44]. Concerning the annealing temperature, comparison to literature is more difficult as reported values range from 140 to 160 °C [45,46]. While this matches relatively well with the optimal annealing temperature in this study (i.e., T = 170 °C, additional investigation is warranted. From a modeling perspective, this difference could come from three sources: (i) a systematic model bias because of the inherent assumptions in the CGMD model, (ii) the Voronoi tessellation approximation that represents the PCBM and P3HT molecules as irregular polyhedral, or (iii) the inherent assumptions in the IPCE model. In contrast, we can also note from comparing the results presented in the seventh column of Table 2 with the high-fidelity observations plotted in Fig. 9, that annealing temperature has only a small influence on the IPCE compared to the volume fraction. This could explain why multiple optimal values for the annealing temperature have been reported in the literature, as observations would become more sensitive to experimental noise. Nevertheless, the presented CGMD is highly consistent with the experimental observations and as such provides designers with a reliable OPVC design tool.

To validate the efficiency of the three-step multi-fidelity design framework we compared the computational cost. In the presented study we used nine low-fidelity and nine high-fidelity simulations. Evaluating the low-fidelity simulations and the high-fidelity simulations required approximately 120 and 300 h, respectively. Comparing this with two hypothetical scenarios: (A) the same study using only high-fidelity simulations, and (B) using Bayesian optimization for stochastic functions [27] in conjunction with the high-fidelity model.

(A) Doing high-fidelity simulations to identify the range of the intermediate modeling parameters requires function evaluations for different values of *VF* and *T*. Evaluating the high-fidelity response for the same design of experiments would have taken approximately nine times more core hours as seen in Table1. While this would have improved the accuracy of the identified intermediate modeling parameters, its influence on the identified range would have been marginal. In addition, it would not have been enough to train an accurate surrogate model and find the optimal processing conditions directly.

(B) Our previous study required 305 simulations at 46 unique volume fractions. If both length-scales have a similar response surface we can expect an equal number of function evaluations. Consequently, the expected computational time will be about 91,000 (300x305) hours using low-fidelity simulations compared to the approximately 3780 (120x9 + 300x9) hours for this study.

The substantial difference in cost not only provides practical validation for the efficiency of the introduced approach but also shows how information about a material's microstructure can be leveraged to expedite the design process. Note that in the above comparison we did not consider the cost of the IPCE model as its computational cost is marginal compared the CGMD simulations.

5. Summary

In this work, we presented a coarse-grained molecular dynamic (CGMD) based three-step multi-fidelity design framework for the design of an organic photovoltaic cell (OPVC) that has optimal performance with respect to acceptable processing conditions. The contributions of the presented work are: (i) simulating large-scale ($100 \times 100 \times 100 nm^3$) high-fidelity simulation whose performance predictions conform with experimental observations, (ii) we introduced a multi-fidelity design framework that enables design of an OPVC with respect to its processing conditions at a predictive fidelity that was previously intractable with CGMD simulations, and (iii) we used an SDF frequency-enhancement technique to reconstruct microstructures with variable sizes. The introduced design framework reduces the computational complexity by using low-fidelity CGMD simulations to gain insight into the potential microstructures that can be achieved from admissible processing conditions. Consequently, we leverage this insight to reduce the input dimensionality of the high-fidelity CGMD simulations by finding the optimal microstructure using the faster structure to performance model. Reducing the input dimensionality of the high-fidelity model facilitates the identification of the optimal processing conditions with significantly fewer costly high-fidelity CGMD function evaluations.

The introduced design framework can be employed for other microstructure design applications when the following four assumptions are valid: (i) the process to structure model and the structure to performance model have shared design variables, i.e., volume fraction of the OPVC compositions (ii) the structure to performance model is much faster to evaluate than the process to structure model, (iii) the microstructure features can be accurately characterized through their spectral density function (SDF), and (iv) the range of the SDF parameters used to characterize the microstructures are independent of the design variables shared by the Process-Structure and Structure-Property models. The authors believe that (iv) is the most limiting assumption which can be accounted for by establishing the underlying relationship between the parameters and design variables by performing more simulations in the next step of this study.

CRediT authorship contribution statement

Umar Farooq Ghumman: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Anton van Beek:** Writing – original draft, Visualization, Formal analysis, Writing – review & editing. **Joydeep Munshi:** Resources, Data curation, Writing – original draft. **TeYu Chien:** Writing – review & editing, Supervision, Funding acquisition. **Ganesh Balasubramanian:** Writing – review & editing, Supervision. **Wei Chen:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Wei Chen reports financial support was provided by National Science Foundation. Wei Chen reports a relationship with National Science Foundation that includes: funding grants.

Acknowledgements

This material is based on the work supported by the National Science Foundation (NSF) under Award No. CMMI-1662435, 1662509 and 1753770 under the Design of Engineering Material Systems (DEMS) program.

Data availability

The raw/processed data required to reproduce these findings cannot

be shared at this time due to technical or time limitations.

References

[1] H. Hoppe, M. Niggemann, C. Winder, J. Kraut, R. Hiesgen, A. Hinsch, D. Meissner, N.S. Sariciftci, Nanoscale Morphology of Conjugated Polymer/Fullerene-Based Bulk-Heterojunction Solar Cells, Adv. Funct. Mater. 14 (10) (2004) 1005–1011.

[2] Y. Lin, A. Magomedov, Y. Firdaus, D. Kaltsas, A. El-Labban, H. Faber, D. R. Naphade, E. Yengel, X. Zheng, E. Yarali, N. Chaturvedi, K. Loganathan, D. Gkeka, S.H. AlShammari, O.M. Bakr, F. Laquai, L. Tsetseris, V. Getautis, T. D. Anthopoulos, 18.4% Organic Solar Cells Using a High Ionization Energy Self-Assembled Monolayer as Hole-Extraction Interlayer, ChemSusChem 14 (17) (2021) 3569–3578.

[3] NREL, "NREL" [Online]. Available: https://www.nrel.gov/pv/cell-efficiency.html.

[4] P.R. Berger, M. Kim, Polymer Solar Cells: P3HT:PCBM and Beyond, J. Renew. Sustain. Energy 10 (1) (2018) 013508, https://doi.org/10.1063/1.5012992.

[5] Y. Cui, P. Zhu, X. Shi, X. Liao, Y. Chen, Theoretical Study of Excited State Charge Transfer Characteristics Based on A-D-A and A-DA′D-A Type Nonfullerene Acceptors, J. Phys. Chem. C 125 (19) (2021) 10250–10259.

[6] G. Zhang, X.K. Chen, J. Xiao, P.C.Y. Chow, M. Ren, G. Kupgan, X. Jiao, C.C.S. Chan, X. Du, R. Xia, Z. Chen, J. Yuan, Y. Zhang, S. Zhang, Y. Liu, Y. Zou, H. Yan, K. S. Wong, V. Coropceanu, N. Li, C.J. Brabec, J.L. Bredas, H.L. Yip, Y. Cao, Delocalization of Exciton and Electron Wavefunction in Non-Fullerene Acceptor Molecules Enables Efficient Organic Solar Cells, Nat. Commun. 11 (1) (2020) 1–10.

[7] J. Munshi, W. Chen, T.Y. Chien, G. Balasubramanian, Machine Learned Metaheuristic Optimization of the Bulk Heterojunction Morphology in P3HT:PCBM Thin Films, Comput. Mater. Sci. 187 (2021) 110119, https://doi.org/10.1016/j.commatsci.2020.110119.

[8] J. Munshi, T.Y. Chien, W. Chen, G. Balasubramanian, Elasto-Morphology of P3HT:PCBM Bulk Heterojunction Organic Solar Cells, Soft Matter 16 (29) (2020) 6743–6751.

[9] C. Wang, H. Fu, L. Jiang, D. Xue, J. Xie, "A Property-Oriented Design Strategy for High Performance Copper Alloys via Machine Learning", npj Comput, Mater. 5 (1) (2019), https://doi.org/10.1038/s41524-019-0227-7.

[10] E.T. Thostenson, T.-W. Chou, Processing-Structure-Multi-Functional Property Relationship in Carbon Nanotube/Epoxy Composites, Carbon N. Y. 44 (14) (2006) 3022–3029.

[11] J. Smith, W. Xiong, W. Yan, S. Lin, P. Cheng, O.L. Kafka, G.J. Wagner, J. Cao, W. K. Liu, Linking Process, Structure, Property, and Performance for Metal-Based Additive Manufacturing: Computational Approaches with Experimental Support, Comput. Mech. 57 (4) (2016) 583–610.

[12] B.S. Fromm, B.L. Adams, S. Ahmadi, M. Knezevic, Grain Size and Orientation Distributions: Application to Yielding of α-Titanium, Acta Mater. 57 (8) (2009) 2339–2348.

[13] M. Fátima Vaz, M.A. Fortes, Grain Size Distribution: The Lognormal and the Gamma Distribution Functions, Scr. Metall. 22 (1) (1988) 35–40.

[14] D.T. Fullwood, S.R. Niezgoda, B.L. Adams, S.R. Kalidindi, Microstructure Sensitive Design for Performance Optimization, Prog. Mater. Sci. 55 (6) (2010) 477–562.

[15] H.-S. Ma, J.-H. Prévost, R. Jullien, G.W. Scherer, Computer Simulation of Mechanical Structure-Property Relationship of Aerogels, J. Non. Cryst. Solids 285 (1-3) (2001) 216–221.

[16] U. Farooq Ghumman, A. Iyer, R. Dulal, J. Munshi, A. Wang, T. Chien, G. Balasubramanian, W. Chen, A Spectral Density Function Approach for Active Layer Design of Organic Photovoltaic Cells, J. Mech. Des. Trans. ASME 140 (11) (2018).

[17] L.V. Gibiansky, O. Sigmund, Multiphase Composites with Extremal Bulk Modulus, J. Mech. Phys. Solids 48 (3) (2000) 461–498.

[18] A. Paul, P. Acar, W.-K. Liao, A. Choudhary, V. Sundararaghavan, A. Agrawal, Microstructure Optimization with Constrained Design Objectives Using Machine Learning-Based Feedback-Aware Data-Generation, Comput. Mater. Sci. 160 (2019) 334–351.

[19] A.J. Kulkarni, K. Krishnamurthy, S.P. Deshmukh, R.S. Mishra, Microstructural Optimization of Alloys Using a Genetic Algorithm, Mater. Sci. Eng. A 372 (1-2) (2004) 213–220.

[20] R. Bostanabad, Y. Zhang, X. Li, T. Kearney, L.C. Brinson, D.W. Apley, W.K. Liu, W. Chen, Computational Microstructure Characterization and Reconstruction: Review of the State-of-the-Art Techniques, Prog. Mater. Sci. 95 (2018) 1–41.

[21] J. Munshi, U. Farooq Ghumman, A. Iyer, R. Dulal, W. Chen, TeYu Chien, G. Balasubramanian, Composition and Processing Dependent Miscibility of P3HT and PCBM in Organic Solar Cells by Coarse-Grained Molecular Simulations, Comput. Mater. Sci. 155 (2018) 112–115.

[22] B.R. Luginbuhl, P. Raval, T. Pawlak, Z. Du, T. Wang, G. Kupgan, N. Schopp, S. Chae, S. Yoon, A. Yi, H. Jung Kim, V. Coropceanu, J.-L. Brédas, T.-Q. Nguyen, G. N.M. Reddy, Resolving Atomic-Scale Interactions in Nonfullerene Acceptor Organic Solar Cells with Solid-State NMR Spectroscopy, Crystallographic Modelling, and Molecular Dynamics Simulations, Adv. Mater. 34 (6) (2022) 2105943, https://doi.org/10.1002/adma.v34.610.1002/adma.202105943.

[23] G. Han, Y. Yi, Z. Shuai, From Molecular Packing Structures to Electronic Processes: Theoretical Simulations for Organic Solar Cells, Adv. Energy Mater. 8 (28) (2018) 1702743, https://doi.org/10.1002/aenm.v8.2810.1002/aenm.201702743.

[24] M. Razi, A. Narayan, R.M. Kirby, D. Bedrov, Force-Field Coefficient Optimization of Coarse-Grained Molecular Dynamics Models with a Small Computational Budget, Comput. Mater. Sci. 176 (2020) 109518, https://doi.org/10.1016/j.commatsci.2020.109518.

[25] H. Eslami, How Thick Is the Interphase in an Ultrathin Polymer Film? Coarse-Grained Molecular Dynamics Simulations of Polyamide-6, 6 on Graphene, J. Phys. Chem. C., **117**(10) (2013) 5249–5257.

[26] J. Yang, D. Custer, C.C. Chiang, Z. Meng, X.H. Yao, Understanding the Mechanical and Viscoelastic Properties of Graphene Reinforced Polycarbonate Nanocomposites Using Coarse-Grained Molecular Dynamics Simulations, Comput. Mater. Sci. 191 (2021) 110339, https://doi.org/10.1016/j.commatsci.2021.110339.

[27] A. van Beek, U.F. Ghumman, J. Munshi, S. Tao, T.Y. Chien, G. Balasubramanian, M. Plumlee, D. Apley, W. Chen, Scalable Adaptive Batch Sampling in Simulation-Based Design with Heteroscedastic Noise, J. Mech. Des. Trans. ASME 143 (3) (2021).

[28] S. Yu, Y. Zhang, C. Wang, W.K. Lee, B. Dong, T.W. Odom, C. Sun, W. Chen, Characterization and Design of Functional Quasi-Random Nanostructured Materials Using Spectral Density Function, J. Mech. Des. Trans. ASME 139 (7) (2017).

[29] S. Yu, C. Wang, Y. Zhang, B. Dong, Z. Jiang, X. Chen, W. Chen, C. Sun, Design of Non-Deterministic Quasi-Random Nanophotonic Structures Using Fourier Space Representations, Sci. Rep. 7 (1) (2017) 1–10.

[30] A. Iyer, R. Dulal, Y. Zhang, U.F. Ghumman, TeYu Chien, G. Balasubramanian, W. Chen, Designing Anisotropic Microstructures with Spectral Density Function, Comput. Mater. Sci. 179 (2020) 109559, https://doi.org/10.1016/j.commatsci.2020.109559.

[31] H. Xu, D.A. Dikin, C. Burkhart, W. Chen, Descriptor-Based Methodology for Statistical Characterization and 3D Reconstruction of Microstructural Materials, Comput. Mater. Sci. 85 (2014) 206–216.

[32] Y. Zhang, H.e. Zhao, I. Hassinger, L.C. Brinson, L.S. Schadler, W. Chen, Microstructure Reconstruction and Structural Equation Modeling for Computational Design of Nanodielectrics, Integr. Mater. Manuf. Innov. 4 (1) (2015) 209–234.

[33] S.D. Sintay, A.D. Rollett, Testing the Accuracy of Microstructure Reconstruction in Three Dimensions Using Phantoms, Model. Simul. Mater. Sci. Eng. 20 (7) (2012) 075005, https://doi.org/10.1088/0965-0393/20/7/075005.

[34] C.L.Y. Yeong, S. Torquato, Reconstructing Random Media, Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 57 (1) (1998) 495–506.

[35] U.F. Ghumman, A. Van Beek, J. Munshi, G. Balasubramanian, T. Chien, W. Chen, Two Stage Optimization of Organic Photovoltaic Cells Using Large Scale Coarse-Grained Molecular Dynamics Simulations, 2021.

[36] S. Tao, A. Van Beek, D.W. Apley, W. Chen, Multi-Model Bayesian Optimization for Simulation-Based Design, J. Mech. Des. Trans. ASME 143 (11) (2021).

[37] D. Huang, T.T. Allen, W.I. Notz, N. Zeng, Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models, J. Glob. Optim. 34 (3) (2006) 441–466.

[38] A. Olsson, G. Sandberg, O. Dahlblom, On Latin Hypercube Sampling for Structural Reliability Analysis, Struct. Saf. 25 (1) (2003) 47–68.

[39] N.K. Hansoge, T. Huang, R. Sinko, W. Xia, W. Chen, S. Keten, Materials by Design for Stiff and Tough Hairy Nanoparticle Assemblies, ACS Nano 12 (8) (2018) 7946–7958.

[40] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual Molecular Dynamics, J. Mol. Graph. 14 (1) (1996) 33–38.

[41] M. Vögele, C. Holm, J. Smiatek, Coarse-Grained Simulations of Polyelectrolyte Complexes: MARTINI Models for Poly(Styrene Sulfonate) and Poly (Diallyldimethylammonium), J. Chem. Phys. 143 (24) (2015) 243151, https://doi.org/10.1063/1.4937805.

[42] J. Munshi, R. Dulal, TeYu Chien, W. Chen, G. Balasubramanian, Solution Processing Dependent Bulk Heterojunction Nanomorphology of P3HT/PCBM Thin Films, ACS Appl. Mater. Interfaces 11 (18) (2019) 17056–17067.

[43] F. Aurenhammer, Voronoi Diagrams—a Survey of a Fundamental Geometric Data Structure, ACM Comput. Surv. 23 (3) (1991) 345–405.

[44] M. Reyes-Reyes, K. Kim, D.L. Carroll, High-Efficiency Photovoltaic Devices Based on Annealed Poly(3-Hexylthiophene) and 1-(3-Methoxycarbonyl)-Propyl-1-Phenyl- (6,6) C61 Blends, Appl. Phys. Lett. 87 (8) (2005) 083506, https://doi.org/10.1063/1.2006986.

[45] G. Li, V. Shrotriya, J. Huang, Y. Yao, T. Moriarty, K. Emery, Y. Yang, High-Efficiency Solution Processable Polymer Photovoltaic Cells by Self-Organization of Polymer Blends, in: Materials for Sustainable Energy: A Collection of Peer-Reviewed Research and Review Articles from Nature Publishing Group, World Scientific Publishing Co., 2010, pp. 80–84.

[46] B. Kadem, A. Hassan, W. Cranton, Efficient P3HT:PCBM Bulk Heterojunction Organic Solar Cells; Effect of Post Deposition Thermal Treatment, J. Mater. Sci. Mater. Electron. 27 (7) (2016) 7038–7048.