ARTICLE   OPEN

Microstructure design using graphs

Pengfei Du[1], Adrian Zebrowski[2], Jaroslaw Zola[3], Baskar Ganapathysubramanian[1] and Olga Wodo[4]

Thin films with tailored microstructures are an emerging class of materials with applications such as battery electrodes, organic electronics, and biosensors. Such thin film devices typically exhibit a multi-phase microstructure that is confined, and show large anisotropy. Current approaches to microstructure design focus on optimizing bulk properties, by tuning features that are statistically averaged over a representative volume. Here, we report a tool for morphogenesis posed as a graph-based optimization problem that evolves microstructures recognizing confinement and anisotropy constraints. We illustrate the approach by designing optimized morphologies for photovoltaic applications, and evolve an initial morphology into an optimized morphology exhibiting substantially improved short circuit current (68% improvement over a conventional bulk-heterojunction morphology). We show optimized morphologies across a range of thicknesses exhibiting self-similar behavior. Results suggest that thicker films (250 nm) can be used to harvest more incident energy. Our graph based morphogenesis is broadly applicable to microstructure-sensitive design of batteries, biosensors and related applications.

npj Computational Materials (2018)4:50 ; doi:10.1038/s41524-018-0108-5

INTRODUCTION

Microstructure design, or morphogenesis, is the decision making process of determining material distribution to optimize performance.[1,2] Since the microstructure critically affects transport of mass,[3] charge[4] as well as reaction rate,[5] there has been a sustained focus on microstructure-sensitive design. The design goal here is to identify tailored microstructures that result in maximization of desired properties. Examples include designing the porous structure of membranes to enhance the filtering process for water reclamation,[6] designing the electrode microstructure in batteries to improve energy transport,[7] and controlling the microstructure in thin film organic electronics to improve energy harvesting[8] or sensing[9] capabilities.

Morphogenesis can broadly be decomposed into two distinct stages: the representation stage and the mapping stage. The first stage (representation stage) is to carefully construct a mathematical framework for representing and, most importantly, synthetically generating microstructures (by 'synthetically' we mean generated in silico). The second stage is to build the mathematical construct that maps the microstructure to a property. The first stage allows sampling candidate microstructures and perturbing them, while the second stage enables evaluating the performance of a microstructure. These two stages are then wrapped into an optimization framework to iteratively identify a microstructure with extreme performance. Current approaches to morphogenesis are limited in two ways: (a) the use of computationally complex methods for microstructure representation[10,11]—field based (phase field, level set, random fields), or feature based approaches—which make reconstruction or perturbing microstructures non-trivial,[12] and (b) the necessity of solving complex, non-linear partial differential equations based mappers[7,13] for evaluating the property of a microstructure. These limitations

bottleneck rapid exploration of the search space, especially for geometrically constrained (thin films) and anisotropic systems where the number of constraints make exploration quite non-trivial.

Here, we show how morphogenesis can be substantially simplified and generalized by using concepts from graph theory. Labeled, weighted, undirected graphs can be used to represent microstructures. They provide a fast, accurate and extensible way to represent, perturb and generate microstructures. Furthermore, graph measures (or queries) can be used to construct viable surrogates for the microstructure-property mappers, which enable fast and efficient property evaluation. This is especially useful when considering microstructure-sensitive processes (like transport and reactions) that can be naturally represented in terms of pathways, connected components and interfaces, which are foundational concepts in graph theory. Additionally, expert intuition, available knowledge and data trends can be naturally fused into these surrogate graph metrics allowing easy integration of domain knowledge with data.

Finally, a graph based approach merges both stages of morphogenesis (microstructure representation and structure-property mapping) in a computationally elegant and efficient way. We show that one can frame morphogenesis as a graph optimization problem and illustrate the approach by designing optimized morphologies for photovoltaic applications.[14] We demonstrate how this fast approach allows us to evolve an initial morphology into an optimized morphology exhibiting substantially improved short circuit current. We show optimized morphologies for thicker films (~250 nm) than currently prevalent, which allows harvesting more incident energy. Our graph based morphogenesis is broadly applicable for microstructure-sensitive design for batteries, biosensors and filtration.

[1]Department of Mechanical Engineering, Iowa State University, Ames, IA, USA; [2]Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, USA; [3]Department of Computer Science and Engineering, University at Buffalo, Buffalo, USA and [4]Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY, USA
Correspondence: Baskar Ganapathysubramanian (baskarg@iastate.edu) or Olga Wodo (olgawodo@buffalo.edu)
These authors contributed equally: Pengfei Du, Adrian Zebrowski.
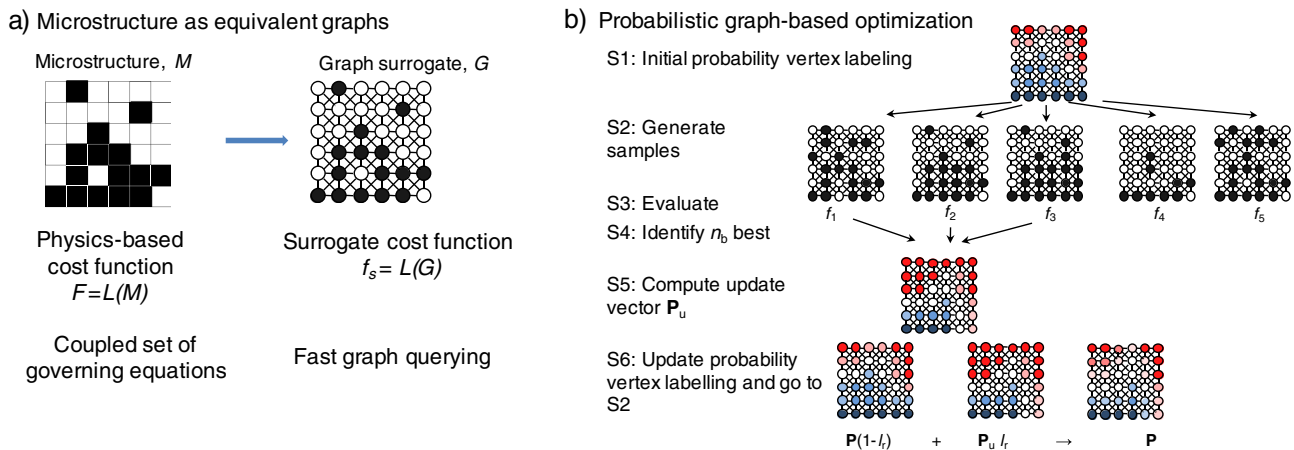
Fig. 1 Elements of graph-based microstructure design framework. **a** The concept of representing morphology as an equivalent graph allows construction of fast graph based surrogates to full-physics based device performance indicators; **b** Scheme of population based incremental learning: An initial (or prior) distribution is sampled to generate candidate graphs (i.e., microstructures), which are rapidly evaluated and ranked. This ranked information is used to update the distribution. Iterating this process evolves the distribution towards graphs exhibiting better performance. Such an approach naturally allows incorporation of expert information into the prior

RESULTS AND DISCUSSION

Graph-based microstructure design framework

Standard approaches for computationally predicting the performance of microstructure sensitive systems involve interrogating the microstructure of interest using computationally intensive frameworks. Incorporating these strategies into any inverse design problem becomes computationally intractable. We address this challenge in two stages. In the first stage, we develop a physics-based 'graph surrogate model' of the performance of the device. The 'surrogate model' extracts the microstructure-aware performance measures and is built (a) using the underlying equivalence between a multi phase morphology and a labeled, weighted, undirected, graph, (b) using concepts and algorithms from graph theory, (c) to seamlessly allow incorporating domain expert knowledge and intuition. In the second stage, this simple and fast surrogate model is linked with a parallel probabilistic optimization algorithm that enables the inverse design of morphologies to maximize certain performance criteria. The choice of the probabilistic approach again allows seamless incorporation of domain knowledge in terms of the prior distribution (see Fig. 1).

The key idea here is to represent microstructure as a graph. A graph-based representation allows us to leverage well studied algorithms from graph theory along with their optimized implementations. We note that graphs have been used in many material science models ranging from molecular structure[15,16] to grain boundaries[17,18] and porous microstructure.[19]

Discretized microstructure is represented as a labeled, weighted, undirected graph $G = (V,E,W,L)$ (Fig. 1a). The graph captures all imperative information inherent to a digitized morphology. Each pixel (or voxel) in the microstructure becomes a graph vertex, with V being the set of all vertices in the microstructure. An associated labeling function, L, assigns a label to each vertex in V. For instance, considering a two-phase microstructure, L could associate a color (red) to represent one phase, and another color (blue) to represent the second phase. Thus L acts on each vertex in V. It should be noted that the set of vertex labels can be easily extended to account for multi phase system or to represent multiple characteristics (like electrical, thermal and magnetic properties, or crystalline vs amorphous properties) of each vertex. Each vertex is connected to its neighboring vertices through edges. The set of all edges in the graph is denoted as E. Finally, the weight function, $W : E \rightarrow R_+$, assigns a non-negative real weight to each edge in E. For instance,

the weight function could simply encode the physical distance between voxels. Alternate weight functions could encode hopping frequencies of charge/mass/momentum transfer between vertices or other transport characteristics. This $G = (V,E,W,L)$ based representation provides a lot of flexibility in representing a variety of microstructures and associated spatially varying properties. On top of the versatility of this graph representation of a microstructure, we gain the ability to use very efficient graph algorithms to extensively characterize the graph (morphology). In general, most transport and reaction characteristics that depend on the underlying morphology can be recast as graph measures.[19,20]

Given graph-based representation of the microstructure, a surrogate model of material properties is constructed directly on the graph-based representation. Traditionally, there have been a spectrum of approaches for constructing a mapping from a given microstructure to its associated device performance indicator. On one end of the (complexity) spectrum are the full-physics based approaches that use detailed microstructure-aware property-structure simulators as the mapping function. These typically come at a very high computational cost precluding optimization except on large computing resources. On the other end of the spectrum are the approaches that use a small set of salient microstructural features from the microstructural data that is then mapped to the performance indicator. These features encode domain knowledge, are intuition driven and/or a result of carefully designed hypothesis driven structure-property studies. A classic example of the latter approach is the Hall-Petch relationship, where the average grain size (salient feature) governs the yield strength of the material.[21] We emphasize that representing a digitized morphology as an equivalent graph allows one to construct a 'graph surrogate model' (based on graph metrics) that can span this full spectrum of structure-property mappers.

A surrogate model is an approximate (but fast) mapping from the microstructure to the output, usually constructed with the aid of statistical and/or data-driven approaches.[22] If a small set of full-physics structure-property simulation results are available, the surrogate model is created using this limited set of results. The surrogate model construction (or 'training') can be done by extracting a large set of graph descriptors/measures from the graph and identifying the best set of features that create a good regression to the full-physics predictions.[23] Representing the full microstructure as a graph helps avoid any a priori feature extraction, and builds the surrogate model on the same graph-based representation. Additionally, intuition driven or domain
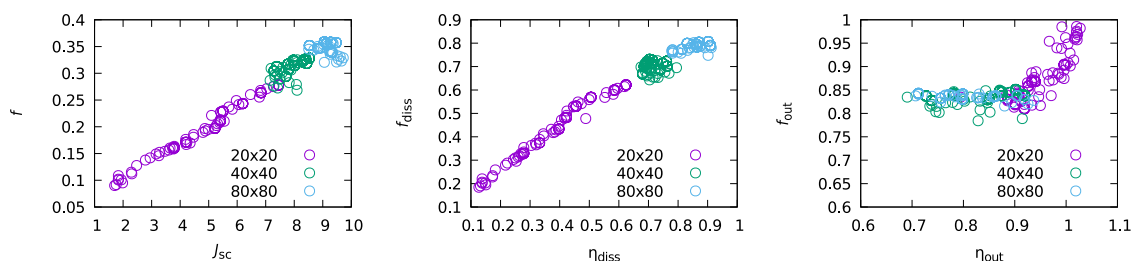
Fig. 2 Correlations between surrogate model and physics based performance characteristics. (Left) correlation between device performance metric (short circuit current in mA/cm^2) and descriptors ($f = f_{abs}f_{diss}f_{out}$), (Middle and right) Individual contributions from graph features describing exciton dissociation and charge transport, respectively. These results are consistent across different graph sizes

knowledge based features can be easily represented as graph metrics and incorporated into a surrogate mapper.[20] Examples include evaluating the connectivity in porous media as graph connectivity metrics and graph paths, domain sizes in ceramics and alloys as graph connected components, among several other examples that originated from network-based computational models for characterizing transport properties.[24–26]

To keep the approach generic and allow diverse applications to a variety of problem specific traits, we separate construction of the surrogate model into two steps: defining a set of physics relevant graph descriptors that are generic (i.e., application agnostic), and weighting functions that encapsulate the specifics of the property of interest (i.e., application dependent). The generic graph descriptors include traits such as path lengths (from specific locations, say, boundaries to all the internal nodes or from one phase to another), centrality measures (of, say, reactants or initiators), connectivity measures (of, say, phases in the micro-structure with the boundary or to each other), domain sizes, among others.[27,28] Next, weighting functions are constructed to capture the detailed trends (say, kinetics aspects, or material specific aspects) of the transport. For example, effect of Brownian motion (or random walks) can be encoded by defining a weighting function that associates smaller weights with larger distances that an ion/charge/molecule has to travel on the graph. This is integrated with the graph path length measures to construct a viable surrogate for the property of interest. This generality provides substantial flexibility to creatively construct reliable surrogate mappers.

Probabilistic graph-based optimization

The graph based surrogate model provides an efficient map from microstructure to performance. We link this surrogate model with a probabilistic optimization algorithm. We specifically choose a probabilistic approach for the following reasons: (a) the outcome of the optimization is a distribution, i.e., a microstructure class (rather than specific microstructure), (b) such probabilistic approaches perform global searches in a gradient free manner and are embarrassingly parallelizable, (c) it is straightforward to encode domain knowledge as a prior distribution, and (d) the optimization is performed directly on graphs, hence eliminating any conversion steps.

In this work, we use a probabilistic optimizer called population-based incremental learning (PbIL).[29] The basic steps of the algorithm are outlined in Fig. 1 (b). PbIL is an optimization algorithm that estimates the explicit distribution of the optimal solution rather than searching for a specific solution, as in typical meta-heuristics (e.g., genetic algorithms,[30] hill-climbing[31] or simulated annealing[31]). The multi variate probability distribution is stored as a probability vector **P** of vertex labeling. In particular, with each vertex in the graph (variable) we associate the probability of a given phase (or component, or more generally the label, L) appearing at this location in the morphology. The probability vector **P** is updated in each iteration of the

optimization based on the feedback from the cost functional. The optimization starts with a given probability vector of vertex labeling (step S1 in Fig. 1(b)), which is either uniform or reflects domain expert knowledge (see Fig. 3). In each iteration, n morphology realizations (i.e., instances of a colored graph) are sampled from the current probability vector **P** (step S2 in Fig. 1(b)). For each realization, j, the graph surrogate model is deployed to evaluate the performance indicator (cost functional), $f(j)$ (step S3 in Fig. 1(b)). The n_b best samples ($n_b \ll n$) are identified (step S4 in Fig. 1(b)). These best samples are used to calculate the update vector, $\mathbf{P}_u$, of the probability distribution (step S5 in Fig. 1(b)). Specifically, the update vector $\mathbf{P}_u$ represents the maximum likelihood estimation of vertex labeling from the n_b best samples. An iteration ends by updating the probability vector $\mathbf{P} = \mathbf{P} \cdot (1 - l_r) + \mathbf{P}_u \cdot l_r$, where l_r is the learning rate (step S6 in Fig. 1(b)). Intuitively, the update reinforces features present in the morphology, and dampens those missing, at the constant rate prescribed by l_r. The algorithm iterates until standard termination criteria are met (iteration limits, improvement bounds). The resulting method has a small memory footprint with only one vector of vertex labels being stored and updated at each iteration. The method is also embarrassingly parallel, since multiple samples drawn from the evolving distribution vector **P** can be evaluated simultaneously. The cost functional f expresses a single objective to optimize. However, we note that in practical settings, f can encapsulate multiple criteria, as we demonstrate in the next section.

Application-specific exploration

We illustrate the power of this approach by optimizing the morphology of organic solar cells (OSC). OSCs have the potential for widespread usage due to their low cost-per-watt, mechanical flexibility, and ease of manufacturability.[14] Their widespread use and commercialization, however, are bottlenecked primarily by relatively low solar efficiencies. Efficiency can be improved either by changing the material chemistry (molecular engineering), microstructure tailoring (microstructure engineering) or architecture (device engineering). It has been argued that most advantages from tuning material energy levels (i.e., molecular engineering) have already been achieved,[32] leaving microstructure design as a promising avenue for efficiency increase.[33,34]

Traditional approaches to microstructure engineering in OSC have relied on systematic trial-and-error experiments to generate diverse microstructures and their device performance (which is resource intensive), or via using computationally intensive simulators to virtually interrogate the candidate morphologies to identify features that an optimal morphology should exhibit.[35,36] It is generally believed that an interdigitated columnar morphology (Fig. 3) is an optimal morphology. This is because the inter-digitated morphology exhibits a graceful combination of two competing requirements—having a large donor–acceptor inter-face to enable efficient exciton dissociation, while still providing straight, fully connected domains to enable efficient charge
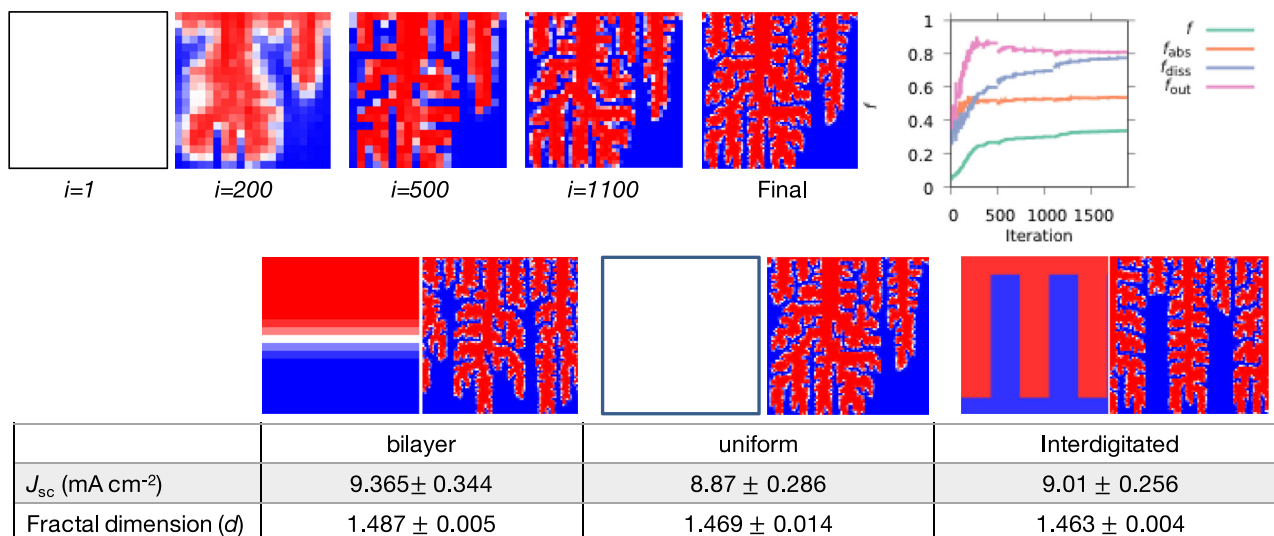
Fig. 3 Microstructure optimization for a typical organic solar cell device configuration. (Top) The morphogenesis from an initial guess towards the final optimized structure. (Bottom) Effect of initial probability field on the morphogenesis outcome. Three initial probability fields: bilayer, uniform and interdigitated initial field evolve towards consistently similar morphologies with statistically similar properties and fractal dimensions

	bilayer	uniform	Interdigitated
J_{sc} (mA cm^{-2})	9.365 ± 0.344	8.87 ± 0.286	9.01 ± 0.256
Fractal dimension (d)	1.487 ± 0.005	1.469 ± 0.014	1.463 ± 0.004

transfer and collection. We substantially improve upon this current state-of-the-art using our graph based morphogenesis (Fig. 3) and show that a whole class of morphologies exist that not only exhibit a much higher photovoltaic performance, but which may also be easier to fabricate.

Following the template for graph based morphogenesis outlined in subsection 'Graph-based microstructure design framework', we represent a OSC microstructure—consisting of two phases, electron-donating and electron-accepting—as an labeled, weighted, undirected graph.[20] We then utilize a finite set of full physics simulations to identify three physically meaningful graph descriptors that together produce an exceptional surrogate for device performance (specifically, the short circuit current)[23] (see Supplementary Information for more details). Figure 2 illustrates the quality of the graph surrogate model. Our surrogate for the short circuit current produced by a morphology is given as: $F = f_{abs} f_{diss} f_{out}$, where these three descriptors are:

(i) f_{abs}: weighted fraction of photoactive material. This accounts for that fraction of the microstructure that absorbs incident light and produces excitons. This also accounts for decaying light intensity as light propagates through a semi-transparent media (light intensity decays with distance from the top electrode);

$$f_{abs} = \sum_{i=0}^{n_D} w_{abs}(d_H) \qquad (1)$$

where n_D is the total number of electron donor vertices, d_H is the distance to the top surface, w_{abs} is the weighting function that encodes the probability of light absorption. We consider three scenarios here: $w_{abs} = 1$ homogeneous light absorption, $w_{abs} = exp(-d_H/\lambda)$ and $w_{abs} = \sum_{p=1}^{P} a_p(d_H)^p$ is the polynomial of order p fitted to the results of Maxwell equation for varying thickness of the device and reflection from the bottom surface accounted for (details provided in Supplementary Information).

(ii) f_{diss} : weighted fraction of donor with weighting based on the distance to the donor–acceptor interface. This feature accounts for the ability of the microstructure to successfully dissociate the produced excitons. Excitons generated in the donor phase diffuse (i.e., via Brownian motion) and are dissociated only if they reach a boundary. This graph based

metric accounts for the average (weighted) distance from any donor vertex to the nearest donor-acceptor interface;

$$f_{diss} = \sum_{i=0}^{n_D} w_{diss}(d_l) \qquad (2)$$

where d_l is the shortest distance from any donor vertex to the nearest acceptor, $w_{diss} = exp(-d_l/L_d)$ is the weighting function that encodes the probability of exciton reaching the interface, and L_d is the exitonic diffusion length. Note how domain knowledge (the exciton dissociation length) is gracefully incorporated into the surrogate model.

(iii) f_{out}: fraction of complementary paths to the electrodes. This graph feature encodes the availability (and length) of charge transport pathways for the produced electrons and holes to move through microstructure and reach the electrodes where they are collected to form useful current:

$$f_{out} = \sum_{i=0}^{n_D^C} w_{tort}(\tau_D) + \sum_{i=i}^{n_A^C} w_{tort}(\tau_A) \qquad (3)$$

where n_D^C is the total number of donor vertices that are connected to anode, n_A^C is the total number of acceptor vertices connected to cathode. Both set of vertices are identified using the connected component algorithm,[27] which is a standard graph algorithm. For each vertex the shortest path towards the respective electrode is determined along with the tortuosity. Tortuosity is defined as the ratio between the shortest path via the appropriate domain (donor or acceptor) versus the shortest straight path without any obstacles. A tortuosity of one corresponds to a straight rising path, where the shortest path is equivalent to the straight path without any obstacles. The tortuosity is calculated for each vertex and subsequently weighted by the function $w_{tort} = exp(-4(\tau - 1))$. The weighting function penalizes the curvy pathways where more recombination can occur and reinforce the straight paths.

To demonstrate the robustness of the surrogate model, Fig. 2 depicts the correlation between graph-based descriptors (surrogate model) and performance metrics from exitonic drift diffusion model (physics-based model).[37] We plot three metrics, full performance evaluation and two major contributions to the

performance, namely exciton dissociation and charge transport. The correlation coefficient for three metrics are 96.65%, 96.95% and 58.81%, respectively. In each subplot, we plot data for three resolutions of the adaptive exploration: 20×20, 40×40, and 80×80. One data point corresponds to one iteration of the optimization. This analysis was performed on data collected during optimization. In particular, for each iteration of the optimization, we saved the best performing intermediate candidate morphology and run the physics-based simulation. Our correlation studies show that the surrogate model is well correlated with the physics-based model at all stages of the optimization.

Figure 3 (top panel) illustrates results for the evolution of the morphology from an initially uniform random distribution. We consider a device thickness of 100 nm, which is typical for OSC devices. We assumed a homogeneous light absorption scenario (i.e., a non-reflecting bottom surface). The PbIL algorithm identifies a dendritic type structure, which is uncommon in OSC's. Interestingly, this structure exhibits a bicontinuous interpenetrated network of well balanced donor and acceptor domains with domain size comparable to the exciton diffusion length (which is set as $L_d = 10$ nm). This dendritic structure exhibits a graceful combination of two competing requirements—having a large donor–acceptor interface to enable efficient exciton dissociation, while still providing straight, fully connected domains to enable efficient charge transfer and collection. In fact, the currently believed optimal morphology—an interdigitated columnar morphology—is though to be optimal for precisely this reason.

We virtually interrogate the dendritic microstructure using a morphology-aware physics-based device simulator.[37] The device simulator calculates the short circuit current density, J_{sc} by solving the associated exciton-drift-diffusion equations (see Supplementary Information). The predicted J_{sc} for the dendritic structure is 8.9 mA cm^{-2}, which is substantially larger than the $J_{sc} = 5.28$ mA cm^{-2} exhibited by the interdigitated columnar morphology. As anticipated, this is due to enhanced exciton dissociation with minimal increase in recombination. More specifically, the increased donor–acceptor interfacial area results in enhanced exciton dissociation, while the charge transport pathways remain relatively non-tortuous, resulting in minimal performance degradation due to recombination.

Incorporating domain knowledge into morphogenesis

We next account for domain knowledge about features of an optimal microstructure into the graph based optimization. Instead of starting from a random initial distribution, we repeat the morphogenesis by initializing the material distribution with the interdigitated columnar morphology (Fig. 3). The ensuing optimized microstructure improves upon the initial guess, and exhibits a $J_{sc} = 9$ mA cm^{-2}. Interestingly, the optimization only changes the material distribution very slightly (see movie M1 in Supplementary Information) by creating the secondary dendritic arms that increase the donor–acceptor interfacial area. We also used a bad initial guess by using a bilayer morphology (which exhibits a $J_{sc} = 0.8$ mA cm^{-2}) as the initial distribution. Here, the optimization produces both primary and secondary dendrites (see movie M2 in Supplementary Information) as the optimization proceeds (Fig. 3).

Universal scaling

Searching for a distribution rather than a specific solution seems to make the framework agnostic to initial distributions, with all three optima exhibiting similar J_{sc}. We qualitatively check this by rerunning the optimization five times, with different random seeds, for each of the three initial distributions considered (random, bilayer, interdigitated). All 15 morphologies exhibited very similar J_{sc} with a 3% variability across the optima. While visually dissimilar, all exhibited similar statistical properties, especially a fractal dimension around 1.5. This suggests that a whole class of morphologies could potentially be optimal. Moreover, we found the fractal dimension to be close to that of quadratic cross (1.49[38]). This is consistent with dendritic structure where perpendicular side branches mimic quadratic cross topology. Although morphologies we report in this paper have columnar dendritic structure, it is interesting to see consistent fractal dimension with quadratic cross topology.

Effect of changing device architecture

We next explore variations in the optimal morphology when light reflecting boundary conditions are imposed at the bottom electrode. As we emphasized earlier, accounting for this new aspect of physical phenomena is trivial: we simply augment the graph surrogate with a weighting term that encodes a thickness dependent light intensity (see Supplementary Information for

Effect of light absorption scenario and device thickness
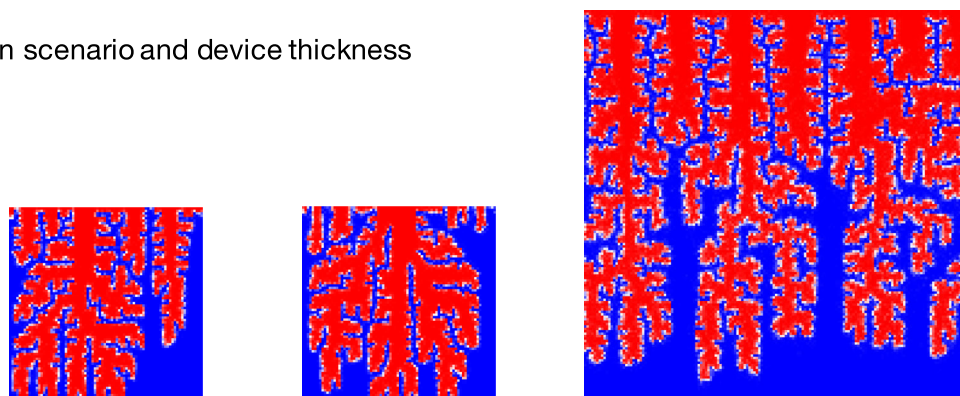
	uniform H=100 nm	bottom reflection H=100nm	bottom reflection H=250 nm
J_{sc} (mA cm^{-2})	8.87 ± 0.286	8.05 ± 0.159	10.09 ± 0.71
Fractal dimension (d)	1.469 ± 0.014	1.471 ± 0.004	1.515 ± 0.006

Fig. 4 Effect of changing device architecture: light absorption scenario and device thickness. Two device thicknesses and three light reflection scenarios evolve towards consistently similar morphologies. Optimized morphologies demonstrate superior properties for a range of device configurations

details). Figure 4 depicts the optimized microstructure for this scenario. Notice that the absorption of the reflected light favors more donor phase in the middle of the active layer. However, the morphology still exhibits a dendritic structure with nearly the same fractal dimensions.

Optimal morphology for thicker devices

We finally explore variations in the optimized microstructure with increasing device thickness. This is a pressing problem in commercializing OSCs as increasing the device thickness can potentially enable absorbing (and converting) more incident radiation, but usually results in poor performance due to the concomitant increased recombination resulting from increased charge transport pathways. Figure 4 compares the optimized microstructures for thin (100 nm) and thicker devices (250 nm). The thicker device has 28.9% more incident radiation that is absorbed. Interestingly, most of this additional incident energy is converted to collected current, with the short circuit current improving from $J_{sc} = 8.05$ mA cm^{-2} for the 100 nm device to $J_{sc} = 10.09$ mA cm^{-2} for thicker devices. Notice that the morphology for the thicker device exhibits a similar dendritic structure with similar fractal dimensions ($\sim$1.5). However, we notice two types of dendrites within the thicker devices. Dendrites with only first order side branches are promoted close to the top, while higher order branching is promoted in the bottom of the device. The fractal dimension of this structure is marginally larger than that exhibited by the thinner device, due to this complexity. The heterogeneity of the optimized structure demonstrates the robustness of the surrogate model in the search of structures with spatially varying complexity.

In summary, microstructure-sensitive design has become an indispensable part of the materials discovery paradigm. However, the identification of optimal microstructures (morphogenesis) has traditionally been hindered by twin bottlenecks, namely non-trivial approaches for microstructure perturbations (i.e., searching the microstructure space), and complex approaches for microstructure-to-property mappings. Here, we illustrate how both these issues can be efficiently resolved via a graph-based strategy. Treating microstructures as graphs allows efficient, modular, and extensible representation which results in simple approaches to perturb and explore the microstructure space. Simultaneously, graph based measures (that can be rapidly, and efficiently computed) make excellent surrogates of microstructure-to-property mapping. This graph based representation and graph-to-property surrogate mapper can be integrated with a probabilistic optimization strategy to efficiently identify optimal microstructures. A graph based approach furthermore allows natural incorporation of domain knowledge into the design process. We illustrate a proof-of-concept application by designing the optimal structure for OSC. This resulted in the identification of a new class of microstructures that exhibit better performance than the currently hypothesized optimal microstructure. We anticipate and look forward to widespread use of this microstructure design strategy by the materials community.

METHODS

For the design framework, an in-house C++ code is used that implements probabilistic graph-based optimization. In each iteration of the optimization, $n = 4000$ samples are generated from the current distribution. Each realization is evaluated according to the cost function, and $n_b = 100$ best microstructures are chosen to compute the update vector $\mathbf{P}_u$. The learning rate is set as 0.1. The search terminates when no improvement in cost is recorded over 100 consecutive iterations. The cost function is defined through the surrogate model evaluated directly on the graph as discussed in the text. The model is constructed to approximate a short circuit current of the device. Specifically, the full model consists of the exitonic drift diffusion model and is used to calculate the short circuit current J_{sc}. More details can be found in[23] and Supplementary Information. To determine statistical similarity, all microstructures are characterized with fractal dimension. In particular, the Hausdorff (box-counting) fractal dimension has been computed for binary image using MATLAB function hausDim. More details are given in Supplementary Information.

AUTHOR CONTRIBUTIONS

P.D. and A.Z. contributed to results generation and subsequent analysis. J.Z., B.G., and O.W. contributed to problem formulations and project oversight. All authors contributed to the manuscript preparation.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Computational Materials* website (https://doi.org/10.1038/s41524-018-0108-5).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Aage, N., Andreassen, E., Lazarov, B. S. & Sigmund, O. Giga-voxel computational morphogenesis for structural design. *Nature* **550**, 84 (2017).
2. Fullwood, D. T., Niezgoda, S. R., Adams, B. L. & Surya, R. K. Microstructure sensitive design for performance optimization. *Prog. Mater. Sci.* **55**, 477–562 (2010).
3. Torquato, S. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*, Volume 16 (Springer Science & Business Media: New York, 2013).
4. Alberto Salleo, R. J., Kline, D. M., DeLongchamp & Michael, L. C. Microstructural characterization and charge transport in thin films of conjugated polymers. *Adv. Mater.* **22**, 3812–3838 (2010).
5. Wilson, J. R. et al. Three-dimensional reconstruction of a solid-oxide fuel-cell anode. *Nat. Mater.* **5**, 541–544 (2006).
6. Quirk, J. P. The significance of the threshold and turbidity concentrations in relation to sodicity and microstructure. *Soil Res.* **39**, 1185–1217 (2001).
7. Garca, R. E., Chiang, Y-M., Carter, W. C., Limthongkul, P. & Bishop, C. M. Microstructural modeling and design of rechargeable lithium-ion batteries. *J. Electrochem. Soc.* **152**, A255–A263 (2005).
8. John, E. A.Organic electronics: addressing challenges. *Nat. Mater.* **13**, 773–775 (2014).
9. Yao, H-B. et al. A flexible and highly pressure-sensitive graphene-polyurethane sponge based on fractured microstructure design. *Adv. Mater.* **25**, 6692–6698 (2013).
10. Kalidindi, S. R., Landi, G. & Fullwood, D. T. Spectral representation of higher-order localization relationships for elastic behavior of polycrystalline cubic materials. *Acta Mater.* **56**, 3843–3853 (2008).
11. Saheli, G., Garmestani, H. & Adams, B. L. Microstructure design of a two phase composite using two-point correlation functions. *J. Comput.-Aided Mater. Des.* **11**, 103–115 (2004).
12. Yeong, C. L. Y. & Salvatore, T. Reconstructing random media. *Phys. Rev. E* **57**, 495 (1998).
13. Suzue, Y., Shikazono, N. & Kasagi, N. Micro modeling of solid oxide fuel cell anode based on stochastic reconstruction. *J. Power Sources* **184**, 52–59 (2008).
14. Brabec, C. J. et al. Plastic solar cells. *Adv. Funct. Mater.* **11**, 15–26 (2001).
15. Blatov, V. A., Shevchenko, A. P. & Proserpio, D. M. Applied topological analysis of crystal structures with the program package topospro. *Cryst. Growth Des.* **14**, 3576–3586 (2014).
16. Sumpter, B. G. & Noid, D. W. Neural networks and graph theory as computational tools for predicting polymer properties. *Macromol. Theory Simul.* **3**, 363–378 (1994).

17. Johnson, O. K., Lund, J. M. & Critchfield, T. R. Spectral graph theory for characterization and homogenization of grain boundary networks. *Acta Mater.* **146**, 42–54 (2018).

18. Moghadam, M. M., Rickman, J. M., Harmer, M. P. & Chan, H. M. The role of boundary variability in polycrystalline grain-boundary diffusion. *J. Appl. Phys.* **117**, 045311 (2015).

19. Cecen, A. et al. Microstructure analysis tools for quantification of key structural properties of fuel cell materials. *ECS Trans.* **41**, 679–687 (2011).

20. Wodo, O., Tirthapura, S., Chaudhary, S. & Ganapathysubramanian, B. A graph-based formulation for computational characterization of bulk heterojunction morphology. *Org. Electron.* **13**, 1105–1113 (2012).

21. Armstrong, R. W., Codd, I., Douthwaite, R. M. & Petch, N. J. The plastic deformation of polycrystalline aggregates. *Philos. Mag.* **7**, 45–58 (1962).

22. Forrester, A. et al. *Engineering Design via Surrogate Modelling: A Practical Guide*. (John Wiley & Sons: West Sussex, United Kingdom, 2008).

23. Wodo, O., Zola, J., Pokuri, B. S. S., Du, P. & Ganapathysubramanian, B. Automated, high throughput exploration of process-structure-property relationships using the mapreduce paradigm. *Mater. Discov.* **1**, 21–28 (2015).

24. Bryant, S. L., Mellor, D. W. & Christopher, A. C. Physically representative network models of transport in porous media. *AIChE J.* **39**, 387–396 (1993).

25. Cecen, A. et al. 3-d microstructure analysis of fuel cell materials: spatial distributions of tortuosity, void size and diffusivity. *J. Electrochem. Soc.* **159**, B299–B307 (2012).

26. Vogel, H. J. A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models. *Eur. J. Soil Sci.* **51**, 99–105 (2000).

27. Cormen, T. H. *Introduction to Algorithms*. (MIT press: Cambridge, Massachusetts, 2009).

28. West, D. B. et al. *Introduction to Graph Theory*, Volume 2 (Prentice Hall: Upper Saddle River, New Jersey 2001).

29. Baluja, S. Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Technical report, School of Computer Science, Carnegie Mellon University, 1994.

30. Stoecklein, D., Wu, C.-Y., Kim, Donghyuk, Carlo, D. D. & Ganapathysubramanian, B. Optimization of micropillar sequences for fluid flow sculpting. *Phys. Fluids* **28**, 012003 (2016).

31. Russell, S. J. & Norvig, P. *Artificial intelligence: a modern approach*. (Pearson Education Limited, Malaysia, 2016).

32. Jackson, N. E., Savoie, B. M., Chen, L. X. & Ratner., M. A. A simple index for characterizing charge transport in molecular materials. *J. Phys. Chem. Lett.* **6**, 1018–1021 (2015).

33. Eperon, G. E., Burlakov, V. M., Docampo, P., Goriely, A. & Henry, J. S. Morphological control for high performance, solution-processed planar heterojunction perovskite solar cells. *Adv. Funct. Mater.* **24**, 151–157 (2014).

34. Huang, Y., Kramer, E. J., Heeger, A. J. & Bazan., G. C. Bulk heterojunction solar cells: morphology and performance relationships. *Chem. Rev.* **114**, 7006–7043 (2014).

35. Li, G. et al. High-efficiency solution processable polymer photovoltaic cells by self-organization of polymer blends. *Nat. Mater.* **4**, 864–868 (2005).

36. Shaheen, S. E. et al. 2.5% efficient organic plastic solar cells. *Appl. Phys. Lett.* **78**, 841–843 (2001).

37. Kodali, H. K. & Ganapathysubramanian, B. A computational framework to investigate charge transport in heterogeneous organic photovoltaic devices. *Comput. Methods Appl. Mech. Eng.* **247**, 113–129 (2012).

38. Vicsek, T. S. *Fractal Growth Phenomena* (World scientific, 1992).