

# How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics

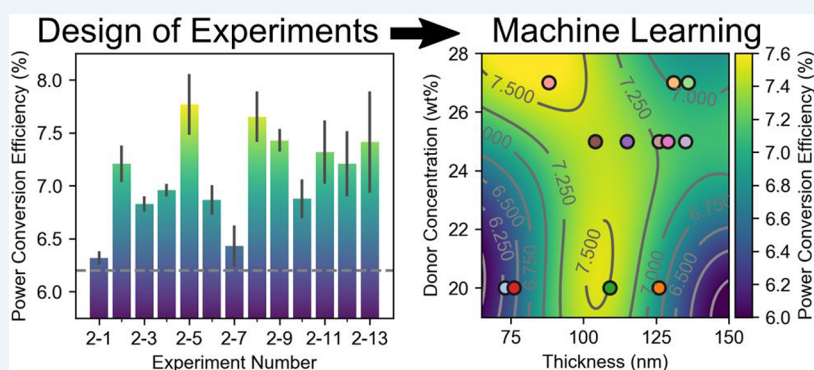
Bing Cao,<sup>†,‡</sup> Lawrence A. Adutwum,<sup>\*,†,§,‡</sup> Anton O. Oliynyk,<sup>†,‡</sup> Erik J. Lubber,<sup>†,‡</sup> Brian C. Olsen,<sup>\*,†,‡</sup> Arthur Mar,<sup>\*,†,‡</sup> and Jillian M. Buriak<sup>\*,†,‡</sup>

<sup>†</sup>Department of Chemistry, University of Alberta, 11227 Saskatchewan Drive, Edmonton, AB T6G 2G2, Canada

<sup>‡</sup>National Institute for Nanotechnology, National Research Council Canada, 11421 Saskatchewan Drive, Edmonton, AB T6G 2M9, Canada

<sup>§</sup>Department of Pharmaceutical Chemistry, College of Health Sciences, University of Ghana School of Pharmacy, P.O. Box LG 43, Legon, Ghana

## Supporting Information



**ABSTRACT:** Most discoveries in materials science have been made empirically, typically through one-variable-at-a-time (Edisonian) experimentation. The characteristics of materials-based systems are, however, neither simple nor uncorrelated. In a device such as an organic photovoltaic, for example, the level of complexity is high due to the sheer number of components and processing conditions, and thus, changing one variable can have multiple unforeseen effects due to their interconnectivity. Design of Experiments (DoE) is ideally suited for such multivariable analyses: by planning one's experiments as per the principles of DoE, one can test and optimize several variables simultaneously, thus accelerating the process of discovery and optimization while saving time and precious laboratory resources. When combined with machine learning, the consideration of one's data in this manner provides a different perspective for optimization and discovery, akin to climbing out of a narrow valley of serial (one-variable-at-a-time) experimentation, to a mountain ridge with a 360° view in all directions.

The recent rise of materials informatics and machine-learning-based computational discovery of materials, such as the Materials Genome Initiative, provides new sources of fuel (ideas, directions) for materials science.<sup>1–5</sup> The ability to screen millions or more possible candidates computationally for a given application or set of properties is resulting in new leads for experimentalists as vast parameter spaces are screened, opening up previously unconsidered avenues of research. One of the most important features of machine-assisted methods is the ability to predict a wide range of materials properties, even when fundamental understanding of the chemistry or physics behind the property is lacking. Even when fundamentals are well-understood, materials

informatics can provide scientific insights, thus enhancing chemical intuition. Machine-learning models are becoming more available and user-friendly, whether or not researchers have a background in informatics (e.g., Citrine's materials informatics, MI, platform<sup>6</sup>). The materials informatics community is working toward the principle of inclusivity, making tools available for the community, such as the Matminer MI library,<sup>7</sup> Magpie materials descriptors,<sup>8</sup> scikit-learn machine-learning package,<sup>9</sup> and COMBO Bayesian optimization library.<sup>10</sup> These tools have been used extensively in materials science and engineering to target novel materials

Published: July 20, 2018

and optimization strategies and to tackle classification problems.<sup>11–22</sup>

Lagging behind machine-learning-based discovery is the experimental side of the equation, leading to the following question: Can experimental optimization keep pace with machine learning? Edisonian or empirical screening of new materials or devices in an experimental laboratory takes considerable time (time scales of months to years) and resources (many thousands of dollars for salaries, supplies, and instrument time).<sup>23</sup> Even if machine learning suggests a new family of materials with given properties, synthesis and optimization in the laboratory requires consideration of experimental parameters not limited to precursor choice, synthetic method, temperature, atmosphere, molar ratios, additives, and many others. Chemical intuition is valuable but is also convoluted with preconceived bias and, in a multidimensional system with many variables, may be flawed.



**Figure 1.** So many options—how do we choose? The authors pondering a small subset of the myriad of published components that have been tested in organic photovoltaic (OPV) devices. Image credit: Kelli Lubner.

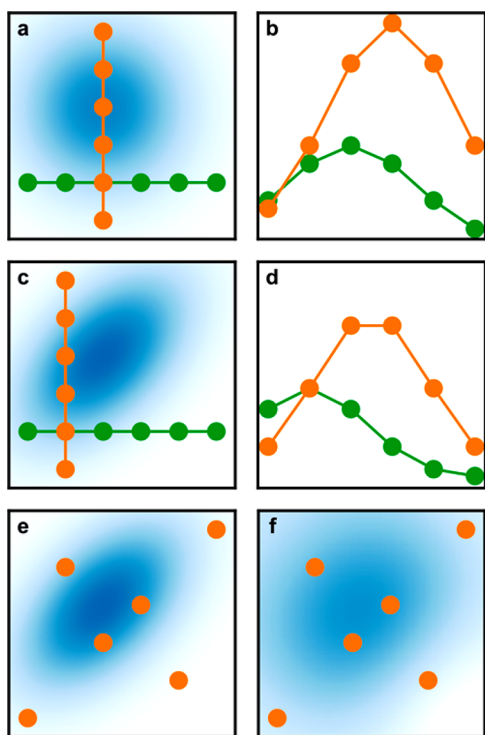
In academic laboratories, we typically teach our students to change one variable at a time in order not to confound the roles of different factors that lead to the observed result. This approach to experimentation is termed one factor/one variable at a time (OFAT, or OVAT) and is limited not only by speed, as it is slow, but it also rarely results in discovery of optima.<sup>24</sup> The sheer number of variables requires the researcher to make educated guesses throughout, and this process infrequently leads to truly optimized parameters.<sup>25</sup> Put more succinctly, one cannot even know if one's "best" results are indeed the overall best that can be achieved with the material or device at hand, as vast areas of parameter space would have been missed. In this Perspective, we describe how Design of Experiments, termed DoE, combined with machine-learning analysis, can dramatically increase the rate of screening and optimization of materials properties and devices. Several variables can be changed and investigated simultaneously if carried out in an appropriate statistical framework that is determined before starting experiments, which can lead to much faster screening of experimental parameters and optimization. The DoE approach enables the experimentalist to sample a large, multidimensional parameter space in a rational manner, which can then be coupled with machine learning to map approximately the parameter space. This methodology can enable faster and higher fidelity exploration of new systems, as

well as optimizing conditions during the development of established materials.

In this Perspective, we describe how Design of Experiments, combined with machine-learning analysis, can dramatically increase the rate of screening and optimization of materials properties and devices.

The basic ideas behind DoE date back to the early 20th century, with the work of British statistician and geneticist, Ronald Fisher, who wrote *The Design of Experiments*, published in 1935.<sup>26</sup> The chemical statistician, George Box (who happened to be married to Fisher's second daughter), was greatly influential throughout the second half of the 20th century with his work on statistical modeling and DoE, among many other topics.<sup>27–32</sup> DoE has since demonstrated its utility in manufacturing, engineering, and other fields.<sup>33</sup> Several excellent tutorials outlining the use of DoE for analytical<sup>34,35</sup> and pharmaceutical<sup>24</sup> chemistry have been written. In this Perspective, we describe the fundamentals of DoE for materials optimization and provide an example of optimization of the bulk heterojunction of organic photovoltaic (OPV) devices through two rounds of parameter optimization. We will demonstrate the enormous potential for experimentalists in materials science to arrive at "real" optima more rapidly, be it device performance or another characteristic. By applying machine-learning methods to the results of DoE, multidimensional maps can be rendered to enable the experimentalist to see, without bias or preconceived ideas, not only the areas or domains of "best" performance but also new areas that have not been previously considered. Design of Experiments, combined with machine-learning methods, is a powerful tool not only to improve your system or materials in a much more directed manner but also to elevate the experimentalist to be able to view the landscape of possibilities clearly and to make new discoveries that had previously been obscured.

**Fundamentals of Design of Experiments (DoE): Optimization of Multiple (Dependent and Independent) Variables.** Take a simple material or a device to optimize that has two uncorrelated parameters, like the one seen in Figure 2a,b. The blue cloud represents the region of best/optimum performance that the experimentalist would like to find. Optimizing one variable at a time (OFAT or OVAT), one of the two parameters would first be chosen to be screened (green line), and then using that optimum on the green line as the starting point for orthogonal screening of the second variable, the experimentalist would then optimize in an orthogonal direction (orange line). In this simple example, with a sequential experimental approach, the experimentalist would find the optimum as this approach does locate the center of the blue region that represents best performance or characteristics. For more complex systems that have a greater number of parameters or correlations, this simple experimental approach would require a larger number of experiments and yet could miss the optimum, as demonstrated in Figure 2c,d. In the first series of tests (shown in green), the experimentalist observes a maximum, but unlike the previous uncorrelated system shown in Figure 2a,b, building from this maximum for the orthogonal second set of experiments (orange line) does

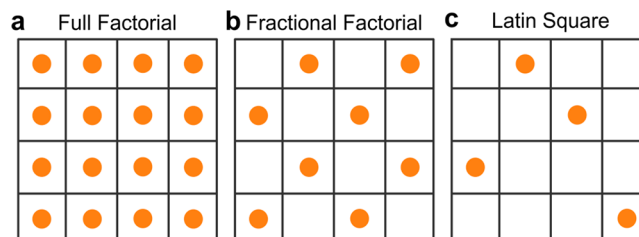


**Figure 2.** Optimization of a two-parameter system: Design of Experiments (DoE) vs one-factor-at-a-time. (a,c,e) Map of the true values (blue gradient) of an output parameter of interest (e.g., yield, power conversion efficiency, polydispersity, etc.) as a function of the two input parameters (horizontal and vertical axes). (a,b) One-factor-at-a-time sampling of an *uncorrelated* system will generally lead to finding the optimum value of the output parameter. (c,d) One-factor-at-a-time sampling of a *correlated* system will generally not result in finding the optimum value of the output parameter. (e) DoE approach with orthogonal sampling of a correlated system. (f) Approximation of the true output parameter map produced by machine-learning fitting methods applied to the six data points acquired by DoE in (e).

not yield the optimum value for the system. The experimentalist could be self-deceived into thinking that they had found the best performance, when in fact, they did not. This method cannot reveal if the actual maximum has indeed been found because a simple OFAT optimization approach reveals little about correlations between the system parameters. Consider the case where the experimentalist tackles the same correlated system, but instead of choosing points on lines *via* serial optimization, they choose specific points in parameter space using DoE principles, where the points are distributed orthogonally and evenly like those found in Figure 2e. With the data from these points, data analysis techniques from machine learning can be used to build a map of the whole parameter space (Figure 2f), instead of only having linear series of data along a small number of lines (directions). This map from the first six experiments also does not currently contain the maximum, but it shows an area of interest and reveals correlations between the parameters. At this point, the experimentalist would then use this map to devise a second set of experimental points focused on the area of interest; the ensemble of experiments needed to reach this point is not greater than the number of experiments carried out in Figure 2a–d. Continuing in this way, the experimentalist will likely

find the system optimum, with some degree of confidence that they have done so.

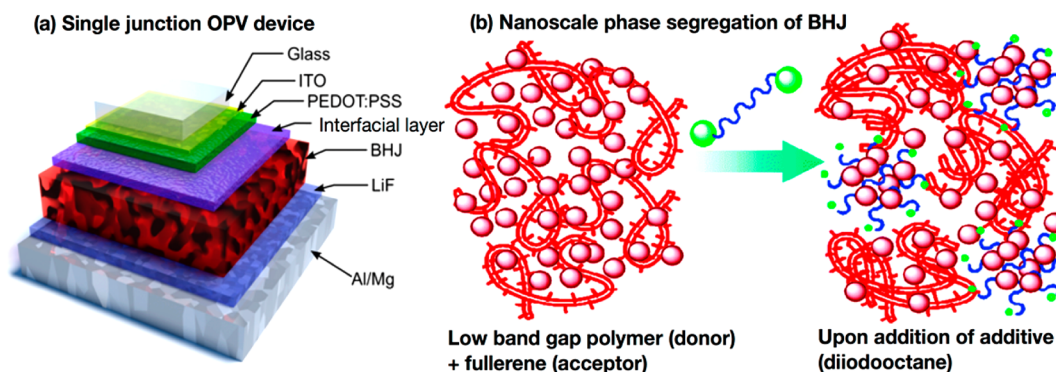
When moving away from OFAT experimental designs, one must think about how to sample the parameter spacing effectively when using DoE. The most naive method would be to implement a full factorial design (Figure 3a), where all



**Figure 3.** Examples of factorial sampling in 2-factor 4-value system showing (a) full factorial, (b) fractional factorial, and (c) Latin square.

possible combinations of parameters are tested. Although this approach would yield the highest fidelity approximation of the behavior of the desired output parameter, it would typically require an unfeasibly large number of experiments, which scales exponentially with the number of parameters [number of experiments = levels<sup>(number of parameters or factors)</sup>]. As such, an alternative approach would be a fractional factorial design, where the parameter space is sampled in a checkerboard-like manner (Figure 3b), reducing the number of experiments by a factor of 2 (this concept can be extended to an arbitrary number of dimensions). If experimental constraints require even sparser sampling, fractional factorial design can be used to reduce the number of experiments by a factor of 3 or 4. Given the exponential scaling of factorial designs (full or partial), these methods can be overly expensive during the first round of parameter testing but become useful implementations upon secondary rounds of optimization when a smaller parameter space is being sampled. In the first round of optimization, which often consists of a large number of parameters and levels being tested, an efficient sampling method to use is the Latin square (Figure 3c). This sampling technique tests every value from each parameter only once, enabling one to see if there are particular correlations between the variables and the individual effects on the experimental results.

**Example of Design of Experiments Applied to Organic Photovoltaics.** Organic photovoltaic (OPV) technologies are of great interest because of the potential to mass manufacture these “plastic” solar cells through ambient processing, such as roll-to-roll printing, inkjet printing, and spray coating.<sup>36,37</sup> Unlike silicon-based solar cells that have an energy payback period of several years, organic photovoltaics could theoretically have an energy payback as short as 24 h due to the low energy costs associated with their production.<sup>38</sup> Organic photovoltaics also have the advantages of being lightweight and flexible<sup>39</sup> with tunable color and the potential for recycling. Power conversion efficiencies (PCEs) of OPV devices have now reached over 14%, but these solar cells are not yet commercialized.<sup>40</sup> Organic photovoltaic devices suffer from problems of instability (thermal, photochemical, and a combination thereof), batch-to-batch variability of the polymer components, and subtle but important variations of solution-processing parameters.<sup>41</sup>



**Figure 4.** (a) Schematic of single junction organic photovoltaic (OPV) devices, showing the bulk heterojunction (BHJ; in red), and the multiple interfacial layers in the device. (b) Schematic of BHJ morphology: in this case, a low band gap polymer donor and a fullerene acceptor undergoing nanoscale phase segregation into discrete nanoscale domains of donor and acceptor. The use of an additive is often purported to assist in nanodomain formation, as shown here. Images reprinted from refs 42 and 43. Copyrights 2016 and 2008, respectively, American Chemical Society.

Organic photovoltaic devices have a complex sandwich architecture (Figure 4). The “meat” of the sandwich is called the bulk heterojunction (BHJ), a mixture of two or more phase-segregated materials that comprise polymers or small molecules.<sup>44</sup> The BHJ absorbs light and results in an exciton that separates into an electron–hole pair. Many hundreds, and almost certainly thousands, of BHJs have been tested and published for OPV performance, and the efficiencies of these devices depend precisely upon the morphology of the BHJ that results from the processing parameters chosen.<sup>36,45–47</sup> The two charge carriers formed in the BHJ, the electron and the hole, must then travel to their respective electrodes, a process that is assisted by interfacial layers on the electrodes that may specifically transport (or block) one of the charge carriers or produce beneficial compositional gradients in the BHJ, among a host of other effects. Many interfacial layers have been tested in both forward and reverse configurations;<sup>48</sup> the decision tree to arrive at the/an ideal layer(s) and configurations is not obvious,<sup>42,49</sup> meaning empirical screening is typically the primary route. The electrodes themselves have traditionally been a transparent conducting oxide such as indium tin oxide and a thin reflective metal back contact, but now there is a large and growing number of electrodes based upon metal nanowires and thin metal films with plasmonic and light trapping properties, and others.<sup>50</sup>

The array of choice with respect to potential components in a particular OPV device is vast, and it is physically impossible to screen every combination, or even a small subset, for every possible architecture due to a lack of time and resources. The number of combinations increases even further in the case of tandem architectures.<sup>51–53</sup> In addition, the role of the morphology of the BHJ is now well-established to be critical for device performance and is exquisitely dependent upon the processing parameters, which, if carried out from solution, implies further decisions regarding concentration, ratio of donor to acceptor, choice and quantity of additives, annealing procedures (thermal, solvent, or both), and choice of solvent(s).<sup>54–57</sup> Empirical screening is the typical route to a determination of what the experimentalist establishes to be optimum conditions, but as shown visually in Figure 2 for a simple, two-parameter optimization, missing the maximum is possible, if not probable, due to the vast space of choice regarding materials and conditions. Worse, as mentioned earlier, with an OFAT/OVAT approach to optimization, one

cannot know if an obtained optimum is indeed the best possible performance of the device. A many-dimensional analysis is needed as there are so many variables to consider. Thus, a DoE approach is ideal, if not essential, to determine which variables are important and, then, using those important variables, to find the optima.

The use of Design of Experiments, with its intentional exclusion of preconceived bias or notions, enables us to “see the big picture”—before, the experimentalist was walking through a narrow valley of one-dimensional data, but these maps are the equivalent of viewing the landscape from a mountain ridgetop.

Here, we outline the process of applying a DoE approach toward the optimization of a BHJ, the core of an OPV device, which comprises a low band gap polymer, poly[*N*-9'-heptadecanyl-2,7-carbazole-*alt*-5,5'-(4',7'-di-2-thienyl-2',1',3'-benzothiadiazole)] (PCDTBT). This polymer is a well-established low band gap polymer with a reported PCE range from 3.0–6.0% in the literature in the standard, forward solar-cell structure (ITO/PEDOT:PSS/BHJ/LiF/Al).<sup>58–60</sup> In this example, we started with the “standard” PEDOT:PSS/ITO electrodes and focused on applying a DoE approach to optimize the PCE of these BHJ devices. The characteristics of the BHJ examined included the ratio of donor to acceptor, the thickness of the BHJ layer, and the concentration of a common additive, diiodooctane; these seemingly simple parameters play critical roles in the nanoscale phase segregation of the layer and, thus, the resulting device efficiency. The thickness of the BHJ depends upon both the solution concentration and spin-casting speed, and thus, the four factors selected for the first round of optimization are summarized in Table 1. There are additional parameters that also could, and should, be considered if time and resources permit expansion of the DoE parametrization to higher dimensions, such as post-spin-coating annealing parameters, choice of solvent for spin-coating, and temperature of the solution before spin-coating of the donor–acceptor solution, among many others.<sup>61–63</sup> These

**Table 1. Factor Selection for the First Round of Design of Experiments for the Optimization of PCDTBT:PCBM Solar Cells**

parameters/factors	parameter range	levels
donor weight percentage (wt %)	10–55	4
total solution concentration (mg/mL)	10–25	4
bulk heterojunction spin-cast speed (rpm)	600–3000	4
processing additive (vol %)	0–12	4

parameters are almost certainly correlated to some degree and thus make optimization of the BHJ system extremely difficult if the experimentalist is changing only one parameter at a time. These principles would also, ideally, be applied to determine ideal interfacial layers and other aspects of the device.

**First Round of Bulk Heterojunction Optimization by Design of Experiments.** The three main physical processing parameters that can be adjusted during the BHJ processing step of solar-cell production are the donor–acceptor ratio, the solution concentration, and the spin speed. As will be shown here, the effects of these variables on device performance, as well as that of additives, are highly interdependent and complex. The thickness of a spun film is determined by the spin speed, solvent vapor pressure and solution viscosity. As both the ratio of the donor–acceptor components and the solution concentration can affect the solution viscosity, all three of these parameters affect the film thickness. The thickness of the BHJ is directly correlated with absorption, which manifests itself on the short-circuit current ( $J_{sc}$ ) of the device. The BHJ morphology—which plays a critical role in OPV performance, manifested through changes in the fill factor (FF) and  $J_{sc}$ —is directly influenced by spin speed and film thickness, both of which affect drying times.<sup>64</sup> Drying time is important because nanoscale phase segregation occurs when the film is wet by solvent or additives. Total concentration and the donor–acceptor ratio are also involved in the resulting morphology of the BHJ film. Another popular approach to affect the BHJ morphology is the use of a low-vapor-pressure additive such as diiodooctane in the BHJ solution to increase drying times. This additive, however, also affects solution viscosity, meaning that its actual role is multifaceted.<sup>54,56</sup> These

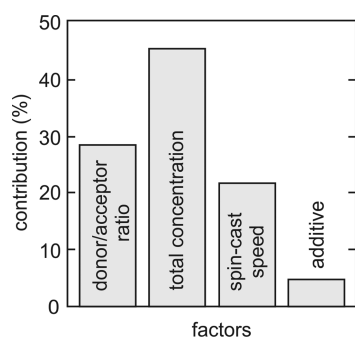
highly interdependent parameters render this system extremely difficult to optimize through the usual one-variable-at-a-time (OFAT/OVAT) approach.

For the four initial parameters chosen (Table 1) for DoE, it is essential to select a wide range of values with as little prejudice as possible. The values should ideally result in continuous BHJ films to enable testing of the resultant devices, but one only anticipates good device performance from a small subset of experiments; we knew from prior literature that some of the combinations would almost certainly lead to low efficiencies, but inclusion of the fullest range of values leads to more robust outcomes. This design, with four variables, is a fraction factorial design representing 1/16 of the total experiments for a full factorial design—with four factors (*i.e.*, parameters) and four levels for each factor, a full factorial design would require  $4^4 = 256$  experiments (not including repetition for statistical significance). The factor permutations were based on a Latin square sampling technique, enabling us to spread our 16 experiments judiciously over the chosen parameter space to approximate the functional dependence of the PCE on these input parameters.

Table 2 shows the outcomes of the 16 experiments using our factorial design and the resulting PCEs. Only experiment 1-16 failed due to the limited solubility of PCDTBT in dichlorobenzene—spin-casting the 25 mg/mL PCDTBT:PCBM solution with 55% PCDTBT was not possible at room temperature. The analysis of the data from this first round, summarized in Table 2, involved two steps. Analysis of variance (ANOVA) was used to evaluate the relative importance of each of the four parameters given by their percent contribution to the PCE of the OPV devices. An ANOVA compares the variance of the output, in this case PCE, for each of the input parameters. If the variance in the output is high for a particular parameter, then that parameter has a high contribution. These contributions can be seen in Figure 5. For more details on how to perform an ANOVA, please see the Supporting Information. The total solution concentration and donor weight percentage contribute about 45 and 28%, respectively, to the resulting PCE, whereas additives only account for a contribution of less than 5%. The

**Table 2. Summary of Parameters Used for PCDTBT:PC<sub>71</sub>BM Solar Cells in the First Round of Design of Experiments Optimization, Resulting Power Conversion Efficiency (Uncertainty Expressed as the Standard Deviation), and Number of Devices Prepared**

experiment #	donor % (wt %)	total concentration (mg/mL)	spin speed (rpm)	additive (vol %)	PCE (%)	number of devices
1-1	10	20	3000	2	0.05(5)	14
1-2	10	25	1000	8	3.24(11)	10
1-3	10	10	600	0	0.016(16)	14
1-4	10	15	2000	12	0.0004(4)	10
1-5	25	20	600	12	7.14(13)	8
1-6	25	15	1000	2	3.22(32)	8
1-7	25	10	3000	8	0.00033(7)	14
1-8	25	25	2000	0	7.21(17)	11
1-9	40	10	1000	12	1.85(5)	3
1-10	40	20	2000	8	6.16(28)	12
1-11	40	25	600	2	3.90(8)	11
1-12	40	15	3000	0	2.27(35)	9
1-13	55	10	2000	2	1.16(4)	3
1-14	55	15	600	8	3.18(12)	10
1-15	55	20	1000	0	3.89(10)	13
1-16	55	25	3000	12	n/a	n/a



**Figure 5.** Analysis of variance and factor evaluation of the first round of optimization for PCDTBT:PC<sub>71</sub>BM bulk heterojunction solar cells.

low contribution of additives enabled us to drop this parameter from the next round of testing and, hence, fit the current data with one less dimension (three dimensions instead of four) to find areas of interest for further optimization. The data were fit using Scikit-learn,<sup>9</sup> with a support vector machine (SVM) using a radial basis function (RBF) kernel, a popular machine-learning algorithm.<sup>65,66</sup> With an RBF kernel, the algorithm will fit best to Gaussian-shaped features that would normally be found in cases of optimization. For more information on this machine analysis, please see the [Supporting Information](#) to find the code to reproduce these fits.

Now that the data have been fit using three parameters, we can generate a three-dimensional map that represents an approximation of the PCE at any point in this space. In order to visualize this space, we generate two-dimensional value maps from the three-dimensional space in the following manner. Slices are taken through the three-dimensional space at certain intervals along one dimension, making a series of two-dimensional maps of the PCE. [Figure 6](#) shows these slices taken at the four donor concentrations (10, 25, 40, and 55 wt % of donor), with *x*- and *y*-axes showing spin speed and total concentration, respectively. The color gradient, scaled as indicated by the color bar on the right, and the contour lines map out the PCE fit from the data in the first round of experiments. The points plotted on the map correspond to the experimental results. Even with the sparse number of data points in this space, an area of interest (higher PCEs) around the 25% donor concentration ([Figure 6b](#)) can be seen in the higher total concentration range (higher values on *y*-axis) and

lower spin speeds (lower values on *x*-axis). This area then served as the basis for planning the next range of parameters to be tested in the second round of optimization.

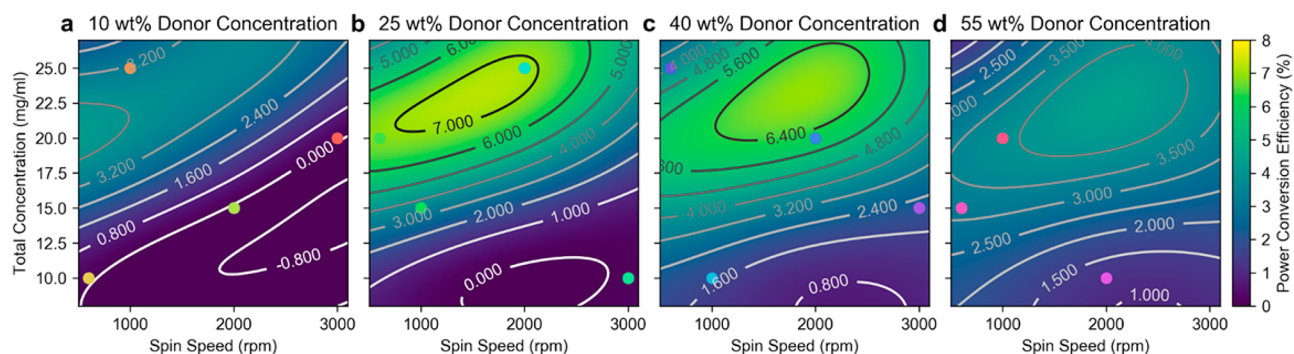
**Second (Subsequent) Rounds of Parameter Evaluation.** Based on the results from the first round of optimization, the parameters were further refined before a second round of experiments was carried out. First, the additive parameter was dropped as it contributed little to the PCE, as per the ANOVA analysis ([Figure 5](#) and [Supporting Information](#)). Second, the ranges of each of the factors were narrowed. For instance, the range of the donor–acceptor ratio was examined in the range of 20–27%, the total solution concentration to 20–25 mg/mL, and spin speed to 1000–2000 rpm from the wider ranges shown in [Table 1](#). As summarized in [Tables 3](#) and [4](#), a partial

**Table 3.** Three Factors Considered in the Second Round of Design of Experiments Parameter Selection for the Optimization of PCDTBT:PC<sub>71</sub>BM Solar Cells

parameters	level 1	level 2	level 3
donor % (wt %)	20	25	27
total concentration (mg/mL)	20	23	25
spin speed (rpm)	1000	1500	2000

factorial design, this time with only three factors and a smaller range, provided the bounds of our area as well as an even distribution filling the space. [Table 4](#) and [Figure 7b](#) reveal that all experiments resulted in PCEs ranging from 6.3 to 7.8%, with the highest PCE of 7.8% obtained from experiment 2-5.

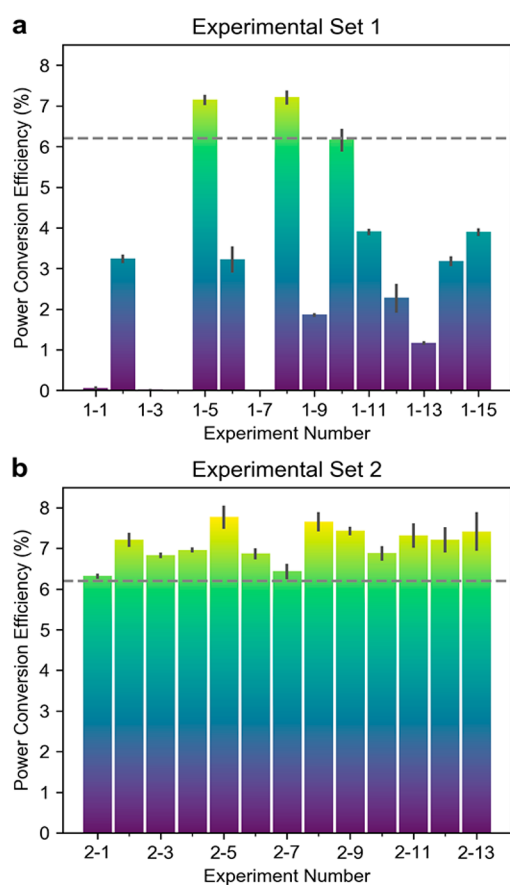
**Data Analysis and Visualization.** As before, in order to visualize the results of the experimental data from [Table 4](#), SVM with RBF fitting was applied to several of the measured device parameters shown in [Figure 8](#). The rows of [Figure 8](#) show the fitting for three measured parameters—power conversion efficiency, short-circuit current, and open-circuit voltage—as color maps where values of these parameters are indicated by the corresponding vertical color bars. Each row contains the same 13 points (the actual results) plotted on the three axes. The *x*- and *y*-axes correspond to spin speed and total concentration, respectively. The donor concentration is shown in the three plots in each row as a slice of the RBF at 20, 25, and 27 wt %. It can be seen that the 20 and 27 wt % donor concentrations have more variability within the test range for all the measured parameters than those with 25 wt %. The 20 and 27 wt % donor concentrations also have their maxima on



**Figure 6.** Support vector machine/radial basis function fits of the power conversion efficiency measured from solar cells produced using the 15 different parameter combinations (seen here as the dots) in the first round of optimization. They are plotted as value maps of slices through the three-dimensional fit, with spin speed on the *x*-axis, total concentration on the *y*-axis, and slices through the donor concentration axis plotted in (a) 10 wt % donor, (b) 25 wt % donor, (c) 40 wt % donor, and (d) 55 wt % donor.

**Table 4. Summary of Parameters Used for PDCTBT:PC<sub>71</sub>BM Solar Cells in the Second Round of Design of Experiments Optimization, Resulting Power Conversion Efficiency (Uncertainty Expressed as the Standard Deviation), and Number of Devices Prepared**

experiment #	donor % (wt %)	total concentration (mg/mL)	spin speed (rpm)	PCE (%)	thickness (nm)	number of devices
2-1	20	20	1500	6.32(6)	73	5
2-2	27	20	1500	7.21(17)	77	11
2-3	20	25	1500	6.83(7)	126	6
2-4	27	25	1500	6.96(6)	131	6
2-5	20	23	1000	7.77(29)	109	4
2-6	27	23	1000	6.87(14)	136	4
2-7	20	23	2000	6.43(19)	76	8
2-8	27	23	2000	7.65(24)	88	7
2-9	25	20	1000	7.43(11)	115	4
2-10	25	25	1000	6.88(18)	135	8
2-11	25	20	2000	7.32(30)	104	7
2-12	25	25	2000	7.21(31)	126	8
2-13	25	23	1500	7.4(5)	129	7



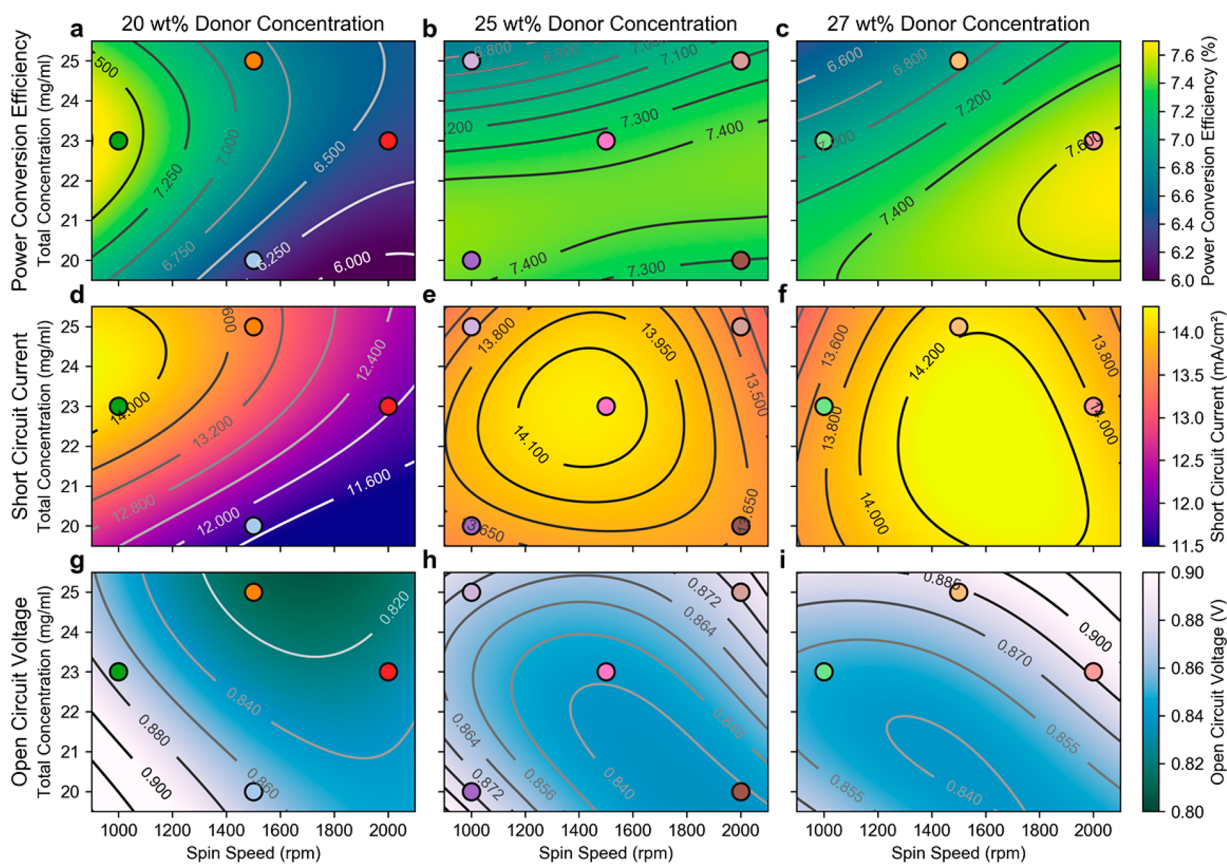
**Figure 7. Mean power conversion efficiency (PCE) of the cells from the first (a) and second (b) rounds of optimization. Error bars represent the standard deviation of the cells' PCEs. The dashed line at 6.2% represents the published values for the PCE of PCDTBT:PC<sub>71</sub>BM after traditional optimization in a standard forward structure, without addition of a third component, special electrode, or optical spacer.<sup>58–60</sup>**

the outer edge of the test range. These results indicate that further testing could reveal a larger area of even higher performing OPV devices.

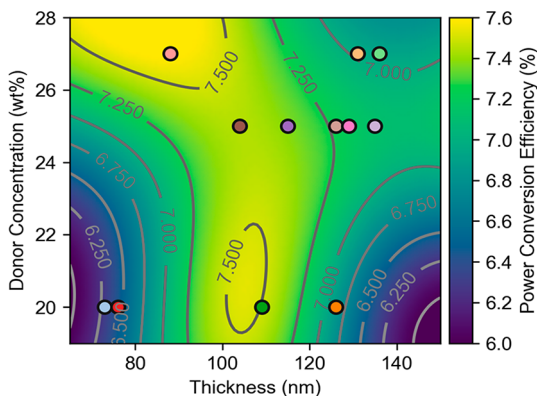
To simplify these results further, the thicknesses of the samples were measured using contact profilometry. As discussed earlier, all three of the test parameters affect the

sample thickness, but the thickness and the donor concentration have the greatest effects on visible light absorption. Figure 9 shows the RBF plot of thickness and donor concentration and the relationship to PCE. The 13 points plotted on this map are the same experimental points shown in Figure 8 and can be matched by their colors. What the experimentalist can learn from this figure is that there are two promising areas that are both surprising and worthy of further exploration. This map suggests that a wider range of donor concentrations can lead to high PCEs than had previously been “believed” or assumed.<sup>59,67</sup> The use of DoE, with its intentional exclusion of preconceived bias or notions, enables us to “see the big picture”—before, the experimentalist was walking through a narrow valley of one-dimensional data, but these maps are the equivalent of viewing the landscape from a mountain ridgetop. New routes and features that were previously unimagined become visible, which could open up new possibilities for research.

**Word of Caution: Here Be Dragons.** Caution should be taken when applying DoE and machine-learning statistics to any set of parameters (Figure 10). When designing partial factorial experiments, one must consider the correlation between variables, as well as the variance in the range tested. Organic photovoltaics tend to be good examples of devices with parameter sets comprising smooth transitions with respect to PCE and related parameters ( $J_{sc}$ , FF,  $V_{oc}$ ). Many systems in chemistry would be expected to be similar, but if the area of interest is very small within the range being tested, or the onset quite sharp, then the risk is that the area of interest will be missed by a partial factorial design. Even in Figure 9, we observed that the area of high PCE in the 20 wt % donor area in the lower part of the figure is quite small, with the sampling set almost missing this area entirely. Another note of caution is that many of the machine-learning fitting methods, such as RBF, were designed for interpolation only and not for extrapolation, as these data fits would have little meaning outside of the family of experimental data points. In our second set of OPV solar optimization experiments, the parameter space was not sufficiently large to encompass the entire area of interest, thus warranting further experiments to expand the boundaries as mentioned previously. Given the current fit (Figure 9), we have little reason to believe that the observed PCE will drop as the devices with 25 wt % BHJ layer decrease in thickness (the middle left area) as the fit shows because this



**Figure 8.** Support vector machine/radial basis function fits of the three experimental parameters of the second round of optimization: spin speed on the  $x$ -axis, total concentration on the  $y$ -axis, and donor concentration plotted in the columns labeled as (a,d,g) 20 wt % donor, (b,e,h) 25 wt % donor, and (c,f,i) 27 wt % donor. The rows represent three different parameters, (a–c) power conversion efficiency, (d–f) short-circuit current, and (g–i) open-circuit voltage, measured from solar cells produced using the 13 different parameter combinations (seen here as the dots).



**Figure 9.** Radial basis function visualization of measured cell power conversion efficiency versus thickness and donor concentration of the solar devices from Figure 8.

region is outside the experimental parameters. With a small number of experiments, we have explored a wide range of space as well as mapped areas of interest, but there is still further exploration that could be done both inside and outside the bounds of the current parameters.

### CONCLUSIONS: LESS HOPING AND MORE KNOWING

The DoE approach to materials optimization combined with machine-learning analysis enables the experimentalist to use



**Figure 10.** Carta marina 1st ed. by Olaus Magnus, 1572. A good map will show new areas to discover (high performing devices/materials), but there are pitfalls and shortcomings to avoid (dragons). Source: National Library of Sweden (<https://www.kb.se/hitta-och-bestall/om-samlingar-och-material/kartor.html>).

scarce resources more effectively, time being one of the scarcest, and to have a higher probability of arriving at a true optimum. With the traditional one-factor-at-a-time (OFAT or OVAT) linear approach to optimization, one cannot know if a true optimum was reached. Imagine the situation of a graduate student spending several years in the laboratory designing,

synthesizing, and characterizing a new family of low band gap polymers or small molecules. With what little time is left of their tenure in the laboratory, devices are prepared, typically following published protocols that were optimized for different compositions and materials. Application of a DoE approach could enable that student to optimize at least one or two rounds of OPV devices efficiently, or any other type of device, with less guesswork. The DoE approach enables less hoping and more knowing in terms of finding a productive and constructive direction forward. The subsequent application of machine-learning methods enables visualization of the data in a profoundly meaningful way, taking the experimentalist to a mountain top with a 360° view of the landscape, as opposed to seeing it as a one-dimensional string of individual experiments.

The ultimate key to progressing from the DoE approach of laboratory experimentation to machine learning is the generation of large data sets, which is particularly challenging in the case of devices (of all kinds, and not just OPV). Continuing with the example of OPV, many tens of thousands of OPV devices have been reported in the literature but scraping these data with relevant experimental parameters is a daunting data-mining problem. Each paper is written in a different format by experimentalists in different laboratories using different parameters, and the experimental sections themselves are written using different terms and with different levels of completeness; therefore, it is unclear that meaningful data can be easily extracted. In addition, if thousands of OPV devices have been reported, we can assume that the results of many low-performing devices were discarded and not published—these negative results would be extremely useful in the context of DoE. The DoE approach could therefore be a step toward formalizing data collection in a rational and standardized manner, if only starting small, within a single group or institution or through a collaboration. An organized approach to experimental design and data collection and reporting could serve as a starting point for real machine learning applied to devices such as OPVs. Such an approach to making academic materials science more efficient could have enormous implications, enabling us to solve problems that threaten humanity's very survival more rapidly, such as the optimization of technologies that underpin the transition to a low-carbon world.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.8b04726.

Experimental details, materials, OPV device fabrication and characterization, details of Design of Experiments (DoE), machine-learning algorithms and code, and photovoltaic test results for all devices (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: adutwum@ualberta.ca.

\*E-mail: bcolsen@ualberta.ca.

\*E-mail: arthur.mar@ualberta.ca.

\*E-mail: jhuriak@ualberta.ca.

### ORCID

Bing Cao: 0000-0002-3825-4669

Lawrence A. Adutwum: 0000-0001-6912-1001

Anton O. Oliynyk: 0000-0003-0732-7340

Erik J. Luber: 0000-0003-1623-0102

Brian C. Olsen: 0000-0001-9758-3641

Arthur Mar: 0000-0003-0474-5918

Jillian M. Buriak: 0000-0002-9567-4328

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by Future Energy Systems of the University of Alberta (<https://futureenergysystems.ca>; Grant Nos. T12-P04 and T12-P01), the Natural Sciences and Engineering Research Council (NSERC, Grant Nos. RGPIN-2014-05195 and RGPIN-2018-04294), Alberta Innovates Technology Futures (Grant No. AITF iCORE IC50-T1 G2013000198), and the Canada Research Chairs program (CRC 207142). We thank Dr. Mario Leclerc for synthesizing the PCDTBT polymer used in this study.

## REFERENCES

- (1) Le, T. C.; Winkler, D. A. Discovery and Optimization of Materials Using Evolutionary Approaches. *Chem. Rev.* **2016**, *116*, 6107–6132.
- (2) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwaraknath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; Amador-Bedolla, C.; Brabec, C. J.; Maruyama, B.; Persson, K. A.; Aspuru-Guzik, A. Accelerating the Discovery of Materials for Clean Energy in the Era of Smart Automation. *Nat. Rev. Mater.* **2018**, *3*, 5–20.
- (3) De Luna, P.; Wei, J.; Bengio, Y.; Aspuru-Guzik, A.; Sargent, E. Use Machine Learning To Find Energy Materials. *Nature* **2017**, *552*, 23–27.
- (4) Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29*, 9436–9444.
- (5) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73–76.
- (6) O'Mara, J.; Meredig, B.; Michel, K. Materials Data Infrastructure: A Case Study of the Citrination Platform To Examine Data Import, Storage, and Access. *JOM* **2016**, *68*, 2031–2034.
- (7) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60–69.
- (8) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016**, *2*, 16028.
- (9) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (10) Ueno, T.; Rhone, T. D.; Hou, Z.; Mizoguchi, T.; Tsuda, K. COMBO: An Efficient Bayesian Optimization Library for Materials Science. *Material. Discovery* **2016**, *4*, 18–21.
- (11) O'Mara, J.; Meredig, B.; Michel, K. Materials Data Infrastructure: A Case Study of the Citrination Platform to Examine Data Import, Storage, and Access. *JOM* **2016**, *68*, 2031–2034.
- (12) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput. Mater.* **2017**, *3*, 54.

- (13) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat. Commun.* **2016**, *7*, 11241.
- (14) Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental Search for High-Temperature Ferroelectric Perovskites Guided by Two-Step Machine Learning. *Nat. Commun.* **2018**, *9*, 1668.
- (15) Gaultois, M. W.; Oliynyk, A. O.; Mar, A.; Sparks, T. D.; Mulholland, G. J.; Meredig, B. Perspective: Web-Based Machine Learning Models for Real-Time Screening of Thermoelectric Materials Properties. *APL Mater.* **2016**, *4*, 053213.
- (16) Oliynyk, A. O.; Sparks, T. D.; Gaultois, M. W.; Ghadbeigi, L.; Mar, A. Gd<sub>12</sub>Co<sub>5.3</sub>Bi and Gd<sub>12</sub>Co<sub>5</sub>Bi, Crystalline Doppelgänger with Low Thermal Conductivities. *Inorg. Chem.* **2016**, *55*, 6625–6633.
- (17) Sparks, T. D.; Gaultois, M. W.; Oliynyk, A.; Brgoch, J.; Meredig, B. Data Mining Our Way to the Next Generation of Thermoelectrics. *Scr. Mater.* **2016**, *111*, 10–15.
- (18) Maier, W. F.; Stöwe, K.; Sieg, S. Combinatorial and High-Throughput Materials Science. *Angew. Chem., Int. Ed.* **2007**, *46*, 6016–6067.
- (19) Oliynyk, A. O.; Antono, E.; Sparks, T. D.; Ghadbeigi, L.; Gaultois, M. W.; Meredig, B.; Mar, A. High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chem. Mater.* **2016**, *28*, 7324–7331.
- (20) Oliynyk, A. O.; Adutwum, L. A.; Harynyuk, J. J.; Mar, A. Classifying Crystal Structures of Binary Compounds AB Through Cluster Resolution Feature Selection and Support Vector Machine Analysis. *Chem. Mater.* **2016**, *28*, 6672–6681.
- (21) Oliynyk, A. O.; Adutwum, L. A.; Rudyk, B. W.; Pisavadia, H.; Lotfi, S.; Hlukhyy, V.; Harynyuk, J. J.; Mar, A.; Brgoch, J. Disentangling Structural Confusion Through Machine Learning: Structure Prediction and Polymorphism of Equiatomic Ternary Phases ABC. *J. Am. Chem. Soc.* **2017**, *139*, 17870–17881.
- (22) Oliynyk, A. O.; Mar, A. Discovery of Intermetallic Compounds from Traditional to Machine-Learning Approaches. *Acc. Chem. Res.* **2018**, *51*, 59–68.
- (23) Pels, K.; Dickson, P.; An, H.; Kodadek, T. DNA-Compatible Solid-Phase Combinatorial Synthesis of  $\beta$ -Cyanoacrylamides and Related Electrophiles. *ACS Comb. Sci.* **2018**, *20*, 61–69.
- (24) Trinh, T. B.; Upadhyaya, P.; Qian, Z.; Pei, D. Discovery of a Direct Ras Inhibitor by Screening a Combinatorial Library of Cell-Permeable Bicyclic Peptides. *ACS Comb. Sci.* **2016**, *18*, 75–85.
- (25) Weissman, S. A.; Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **2015**, *19*, 1605–1633.
- (26) Fisher, R. A. *The Design of Experiments*, 9th ed.; Macmillan Publishing Co., 1971.
- (27) Box, G. E. P.; Wilson, K. B. On the Experimental Attainment of Optimum Conditions. In *Breakthroughs in Statistics*; Springer Series in Statistics; Springer: New York, 1992; pp 270–310.
- (28) Janacek, G. Time Series Analysis Forecasting and Control. *J. Time. Ser. Anal.* **2009**, *31*, 303–303.
- (29) Box, G. E. P.; Hunter, S.; Hunter, W. G. *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; Wiley-Interscience, 2005.
- (30) Box, G. E. P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799.
- (31) DeGroot, M. H. A Conversation with George Box. *Statist. Sci.* **1987**, *2*, 239–258.
- (32) Smith, A. F. M. George Edward Pelham Box. 10 October 1919–28 March 2013. *Biogr. Mem. Fellows R. Soc.* **2015**, *61*, 23–37.
- (33) Ilzarbe, L.; Álvarez, M. J.; Viles, E.; Tanco, M. Practical Applications of Design of Experiments in the Field of Engineering: A Bibliographical Review. *Qual. Reliab. Eng. Int.* **2008**, *24*, 417–428.
- (34) Leardi, R. Experimental Design in Chemistry: A Tutorial. *Anal. Chim. Acta* **2009**, *652*, 161–172.
- (35) Vera Candiotti, L.; De Zan, M. M.; Cámara, M. S.; Goicoechea, H. C. Experimental Design and Multiple Response Optimization. Using the Desirability Function in Analytical Methods Development. *Talanta* **2014**, *124*, 123–138.
- (36) Lu, L.; Zheng, T.; Wu, Q.; Schneider, A. M.; Zhao, D.; Yu, L. Recent Advances in Bulk Heterojunction Polymer Solar Cells. *Chem. Rev.* **2015**, *115*, 12666–12731.
- (37) Zhang, G.; Zhao, J.; Chow, P. C. Y.; Jiang, K.; Zhang, J.; Zhu, Z.; Zhang, J.; Huang, F.; Yan, H. Nonfullerene Acceptor Molecules for Bulk Heterojunction Organic Solar Cells. *Chem. Rev.* **2018**, *118*, 3447–3507.
- (38) Espinosa, N.; Hösel, M.; Angmo, D.; Krebs, F. C. Solar Cells with One-Day Energy Payback for the Factories of the Future. *Energy Environ. Sci.* **2012**, *5*, 5117–5132.
- (39) Li, Y.; Xu, G.; Cui, C.; Li, Y. Flexible and Semitransparent Organic Solar Cells. *Adv. Energy Mater.* **2018**, *8*, 1701791.
- (40) Li, S.; Ye, L.; Zhao, W.; Yan, H.; Yang, B.; Liu, D.; Li, W.; Ade, H.; Hou, J. A Wide Band Gap Polymer with a Deep Highest Occupied Molecular Orbital Level Enables 14.2% Efficiency in Polymer Solar Cells. *J. Am. Chem. Soc.* **2018**, *140*, 7159–7167.
- (41) Lee, E. K.; Lee, M. Y.; Park, C. H.; Lee, H. R.; Oh, J. H. Toward Environmentally Robust Organic Electronics: Approaches and Applications. *Adv. Mater.* **2017**, *29*, 1703638.
- (42) Cao, B.; He, X.; Fetterly, C. R.; Olsen, B. C.; Lubner, E. J.; Buriak, J. M. Role of Interfacial Layers in Organic Solar Cells: Energy Level Pinning Versus Phase Segregation. *ACS Appl. Mater. Interfaces* **2016**, *8*, 18238.
- (43) Lee, J. K.; Ma, W. L.; Brabec, C. J.; Yuen, J.; Moon, J. S.; Kim, J. Y.; Lee, K.; Bazan, G. C.; Heeger, A. J. Processing Additives for Improved Efficiency from Bulk Heterojunction Solar Cells. *J. Am. Chem. Soc.* **2008**, *130*, 3619–3623.
- (44) Cheng, P.; Li, G.; Zhan, X.; Yang, Y. Next-Generation Organic Photovoltaics Based on Non-Fullerene Acceptors. *Nat. Photonics* **2018**, *12*, 131–142.
- (45) Song, J.; Zhang, M.; Yuan, M.; Qian, Y.; Sun, Y.; Liu, F. Morphology Characterization of Bulk Heterojunction Solar Cells. *Small Methods* **2018**, *2*, 1700229.
- (46) Gedefaw, D.; Prosa, M.; Bolognesi, M.; Seri, M.; Andersson, M. R. Recent Development of Quinoxaline Based Polymers/Small Molecules for Organic Photovoltaics. *Adv. Energy Mater.* **2017**, *7*, 1700575.
- (47) Yan, C.; Barlow, S.; Wang, Z.; Yan, H.; Jen, A. K.-Y.; Marder, S. R.; Zhan, X. Non-Fullerene Acceptors for Organic Solar Cells. *Nat. Rev. Mater.* **2018**, *3*, 18003.
- (48) Yin, Z.; Wei, J.; Zheng, Q. Interfacial Materials for Organic Solar Cells: Recent Advances and Perspectives. *Adv. Sci.* **2016**, *3*, 1500362.
- (49) Huang, L.; Wang, G.; Zhou, W.; Fu, B.; Cheng, X.; Zhang, L.; Yuan, Z.; Xiong, S.; Zhang, L.; Xie, Y.; Zhang, A.; Zhang, Y.; Ma, W.; Li, W.; Zhou, Y.; Reichmanis, E.; Chen, Y. Vertical Stratification Engineering for Organic Bulk-Heterojunction Devices. *ACS Nano* **2018**, *12*, 4440–4452.
- (50) Lu, H.; Ren, X.; Ouyang, D.; Choy, W. C. H. Emerging Novel Metal Electrodes for Photovoltaic Applications. *Small* **2018**, *14*, 1703140.
- (51) Andersen, T. R.; Dam, H. F.; Hösel, M.; Helgesen, M.; Carlé, J. E.; Larsen-Olsen, T. T.; Gevorgyan, S. A.; Andreasen, J. W.; Adams, J.; Li, N.; Machui, F.; Spyropoulos, G. D.; Ameri, T.; Lemaître, N.; Legros, M.; Scheel, A.; Gaiser, D.; Kreul, K.; Berny, S.; Lozman, O. R.; et al. Scalable, Ambient Atmosphere Roll-to-Roll Manufacture of Encapsulated Large Area, Flexible Organic Tandem Solar Cell Modules. *Energy Environ. Sci.* **2014**, *7*, 2925–2933.
- (52) Lu, S.; Ouyang, D.; Choy, W. C. H. Recent Progress of Interconnecting Layer for Tandem Organic Solar Cells. *Sci. China: Chem.* **2017**, *60*, 460–471.
- (53) Li, M.; Gao, K.; Wan, X.; Zhang, Q.; Kan, B.; Xia, R.; Liu, F.; Yang, X.; Feng, H.; Ni, W.; Wang, Y.; Peng, J.; Zhang, H.; Liang, Z.; Yip, H.-L.; Peng, X.; Cao, Y.; Chen, Y. Solution-Processed Organic Tandem Solar Cells with Power Conversion Efficiencies > 12%. *Nat. Photonics* **2017**, *11*, 85–90.

(54) Pivrikas, A.; Neugebauer, H.; Sariciftci, N. S. Influence of Processing Additives to Nano-Morphology and Efficiency of Bulk-Heterojunction Solar Cells: A Comparative Review. *Sol. Energy* **2011**, *85*, 1226–1237.

(55) Ma, W.; Yang, G.; Jiang, K.; Carpenter, J. H.; Wu, Y.; Meng, X.; McAfee, T.; Zhao, J.; Zhu, C.; Wang, C.; Ade, H.; Yan, H. Influence of Processing Parameters and Molecular Weight on the Morphology and Properties of High-Performance PffBT4T-2OD:PC71BM Organic Solar Cells. *Adv. Energy Mater.* **2015**, *5*, 1501400.

(56) Zhao, J.; Zhao, S.; Xu, Z.; Qiao, B.; Huang, D.; Zhao, L.; Li, Y.; Zhu, Y.; Wang, P. Revealing the Effect of Additives with Different Solubility on the Morphology and the Donor Crystalline Structures of Organic Solar Cells. *ACS Appl. Mater. Interfaces* **2016**, *8*, 18231–18237.

(57) Babics, M.; Liang, R.-Z.; Wang, K.; Cruciani, F.; Kan, Z.; Wohlfahrt, M.; Tang, M.-C.; Laquai, F.; Beaujuge, P. M. Solvent Vapor Annealing-Mediated Crystallization Directs Charge Generation, Recombination and Extraction in BHJ Solar Cells. *Chem. Mater.* **2018**, *30*, 789–798.

(58) Blouin, N.; Michaud, A.; Leclerc, M. A Low-Bandgap Poly(2,7-Carbazole) Derivative for Use in High-Performance Solar Cells. *Adv. Mater.* **2007**, *19*, 2295–2300.

(59) Park, S. H.; Roy, A.; Beaupré, S.; Cho, S.; Coates, N.; Moon, J. S.; Moses, D.; Leclerc, M.; Lee, K.; Heeger, A. J. Bulk Heterojunction Solar Cells with Internal Quantum Efficiency Approaching 100%. *Nat. Photonics* **2009**, *3*, 297–302.

(60) Alem, S.; Chu, T.-Y.; Tse, S. C.; Wakim, S.; Lu, J.; Movileanu, R.; Tao, Y.; Bélanger, F.; Désilets, D.; Beaupré, S.; Leclerc, M.; Rodman, S.; Waller, D.; Gaudiana, R. Effect of Mixed Solvents on PCDTBT:PC70BM Based Solar Cells. *Org. Electron.* **2011**, *12*, 1788–1793.

(61) Yi, Z.; Ni, W.; Zhang, Q.; Li, M.; Kan, B.; Wan, X.; Chen, Y. Effect of Thermal Annealing on Active Layer Morphology and Performance for Small Molecule Bulk Heterojunction Organic Solar Cells. *J. Mater. Chem. C* **2014**, *2*, 7247–7255.

(62) Wang, J.; Liang, Z. Synergetic Solvent Engineering of Film Nanomorphology To Enhance Planar Perylene Diimide-Based Organic Photovoltaics. *ACS Appl. Mater. Interfaces* **2016**, *8*, 22418–22424.

(63) Engmann, S.; Ro, H. W.; Herzing, A.; Snyder, C. R.; Richter, L. J.; Geraghty, P. B.; Jones, D. J. Film Morphology Evolution During Solvent Vapor Annealing of Highly Efficient Small Molecule Donor/Acceptor Blends. *J. Mater. Chem. A* **2016**, *4*, 15511–15521.

(64) Huang, Y.; Kramer, E. J.; Heeger, A. J.; Bazan, G. C. Bulk Heterojunction Solar Cells: Morphology and Performance Relationships. *Chem. Rev.* **2014**, *114*, 7006–7043.

(65) Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Statistics and Computing* **2004**, *14*, 199–222.

(66) Chang, Y.-W.; Hsieh, C.-J.; Chang, K.-W.; Ringgaard, M.; Lin, C.-J. Training and Testing Low-Degree Polynomial Data Mappings via Linear SVM. *J. Mach. Learn. Res.* **2010**, *11*, 1471–1490.

(67) Sun, Y.; Seo, J. H.; Takacs, C. J.; Seifert, J.; Heeger, A. J. Inverted Polymer Solar Cells Integrated with a Low-Temperature-Annealed Sol-Gel-Derived ZnO Film as an Electron Transport Layer. *Adv. Mater.* **2011**, *23*, 1679–1683.