Machine Learning Optimization of P-Type Transparent Conducting Films

Lingfei Wei,[†,‡] Xiaojie Xu,[§] Gurudayal,[‡,⊥] James Bullock,[‖] and Joel W. Ager,[,‡,⊥,∇]*

† School of Chemistry and Chemical Engineering, Southeast University, Nanjing 211189, P. R. China

‡ Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America

§ Materials Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, United States of America

⊥ Joint Center for Artificial Photosynthesis, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America

‖ Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Victoria, Australia

∇ Department of Materials Science and Engineering, University of California Berkeley, Berkeley, CA 94720, United States of America

ABSTRACT

P-type transparent conducting materials (p-TCMs) are important components of optoelectronic devices including solar cells, photodetectors, displays, and flexible sensors. Cu-Zn-S (CZS) thin films prepared by chemical bath deposition (CBD) can have both high transparency in the visible range (>80%) as well as excellent hole conductivity (>1000 S cm^{-1}). However, the interplay between the deposition parameters in the CBD process (metal and sulfur precursor concentrations, temperature, pH, complexing agents, etc.) creates a multi-dimensional parameter space such that optimization for a specific application is challenging and time consuming. Here, we show that strategic design of experiment (DOE) combined with machine learning (ML) allows for efficient optimization of p-TCM performance. The approach is guided by a figure of merit (FOM) calculated from the film conductivity and optical transmission in the desired spectral range. A specific example is shown using two steps of optimization using a selected subset of 4 experimental CBD factors. The machine learning model is based on support vector regression (SVR) employing a radial basis function (RBF) as the kernel function. 10-fold cross-validation was performed to mitigate overfitting. After the first round of optimization, predicted areas in the parameter space with maximal FOMs were selected for a second round of optimization. Films with optimal FOMs were incorporated into heterojunction solar cells and transparent photodiodes. The optimization approach shown here will be generally applicable to any materials synthesis process with multiple parameters.

INTRODUCTION

P-type transparent conducting materials (p-TCMs) have applications as components in a number of electronic devices such as touchscreens, LCD layers, LEDs, and photovoltaics.[1–4] Oxides such as delafossites (e.g., CuAlO$_2$, CuGaO$_2$, CuInO$_2$, etc.)[5], spinels (AB$_2$O$_4$, e.g., NiCo$_2$O$_4$)[6,7], and

mixed oxides (e.g., In_2O_3-Ag_2O)[8] have been investigated as p-TCMs in the past few years,[9] but it has been difficult to obtain the desired combination of high transparency and conductivity. Sulfides have also been investigated as TCMs.[10–16] In particular, the Cu-Zn-S system (CZS) has been the subject of investigation since the first report of p-type behavior in 1999.[17] Although a number of methods can be employed to make thin CZS films including pulsed laser deposition (PLD) and sputtering,[15,18–21] solution based methods such as chemical bath deposition (CBD) are attractive from a scalability point of view.[14,22–28] In our previous work, $(CuS)_x(ZnS)_{1-x}$ nanocomposites were deposited by CBD and found to possess hole conductivity as high as 1000 S cm^{-1} and optical transmission above 70% in the visible range of the spectrum.[29]

In contrast to the vapor-based deposition methods discussed above, it is less straightforward to tune CBD deposition parameters to produce continuous TCMs of the desired conductivity and transparency. As shown schematically in Figure 1a, heterogeneous nucleation on the substrate and homogeneous nucleation in the solution happen simultaneously during CBD. Additionally, subsequent film growth can proceed by either ion-by-ion or by a cluster-by-cluster process in which the clusters formed by homogeneous nucleation deposit on the substrate.[30] The latter case generates films with a nonuniform surface and large grain sizes, leading to poor substrate adhesion and lower optical transmission (lower part of Figure 1a). To form adherent films, nucleation should happen directly on substrate followed by a moderate rate of ion-by-ion growth (upper part of Figure 1a).[30,31]
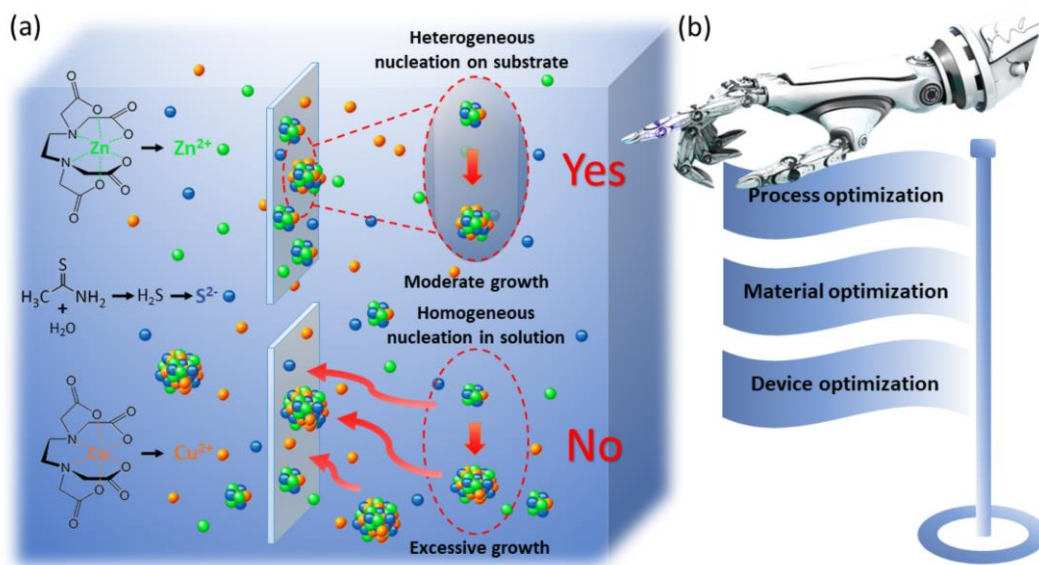
Figure 1. (a) Schematic illustration of the growth mechanisms of $(CuS)_x$:$(ZnS)_{1-x}$ films via chemical bath deposition. Orange, green and blue balls represent Cu^{2+}, Zn^{2+} and S^{2-}, respectively. The upper case depicts heterogeneous nucleation and growth on the substrate. (b) Aims of machine learning optimization in this work. Royalty-free graphic of the robotic hand was obtained from pngtree.com.

There are a number of interrelated process parameters in CBD which will influence the growth mode of $(CuS)_x(ZnS)_{1-x}$ nanocomposite thin films. The first is the relative ratios of the Cu and Zn precursors. Bulk CuS is p-type semiconductor material with a direct band gap of ~2.5 eV,[32] and ZnS is n-type semiconductor material with a direct band gap of ~3.6 eV.[33] The hole conductivity of $(CuS)_x(ZnS)_{1-x}$ nanocomposite films is attributed to the conducting network formed by covellite CuS nanocrystals while the transparency is dependent on the transparent fillers (ZnS) and average nanocrystal sizes.[29] Thus the Cu and Zn precursor ratio will influence the overall optical band gap of the composite, with higher Zn ratios increasing the effective band gap. Higher Cu ratios will

lower the energy band gap but also increase the proportion of the hole-conducting covellite CuS phase. Secondly, the concentration of the complexing agent (Na_2EDTA in our case) affects the crystallite size in two ways. It reduces the aggregation of metal ions and forms a concentration profile of metal ions which is more favorable to nucleation rather than crystal growth. [34,35] At the same time, a high concentration of complexing agent can be undesirable; additional Na_2EDTA can accelerate the growth rate and promote the cluster-by-cluster process, which results in poor quality deposition and nonuniformity. [36] Thirdly, the rate of both nucleation and growth are sensitive to temperature. [37] Finally, the reaction time influences the thickness and quality of CZS thin films. CBD is a self-limiting process with the growth rate declining as the concentration of precursors decreases. [38] In the case of TCM films, the sheet resistance would decrease with increasing film thickness at fixed film conductivity. But if lower quality material (less transparent or less conductive) would be deposited late in the growth cycle, extended reaction times would not be beneficial.

In spite of this complicated reaction environment and multi-dimensional parameter space, we note that most prior studies of this materials system, including our own, have employed a low-efficiency "one factor at a time" approach to optimizing p-TCM properties. We thus hypothesized that a more efficient approach would be to use a more strategic design of experiment guided by a regression model. We note recent examples of machine learning (ML) in materials sciences, [39–42] with a number of studies using a combination of first principles theory and experiment for materials discovery. [43–48] For navigation of a multi-dimensional materials synthesis space, we were inspired by recent studies of machine learning regression models applied to organic electronics. [49,50]

Here, we present a machine learning approach to efficiently optimize p-TCMs within the multi-dimensional parameter space for the CBD process. We use a strategic design of experiment (DOE)

to reduce the number of required experiments. We chose an ML model, support vector regression (SVR), which yields reliable predictions over our parameter space after cross validation. After two rounds of optimization using this approach, an optimal process for transparent (84% at 600 nm) and conductive (430 S cm^{-1}) CZS films was identified. We then compared the ML-optimized and non-optimized thin films as components of heterojunction solar cells and transparent photodiodes, with the optimized films producing superior behavior in both types of devices as expected. The approach illustrated here is generally applicable to any material synthesis process or materials/devices optimization process and can significantly advance the rate of materials/devices development.

EXPERIMENTAL METHODS

Chemical bath deposition. CZS thin films were synthesized via CBD at solution temperatures between 60 and 90 °C. Substrates (soda lime glass, silicon and FTO glass) were cleaned sequentially in an ultrasonic bath with detergent, methanol, and deionized water (DIW). The metal precursor solutions were prepared with $CuSO_4$ (purity 99%, Alfa Aesar) and $Zn(CH_3COO)_2$ (99.9%, Alfa Aesar) at a total metal ion concentration of 0.1 M (100 ml). The complexing agent Na_2EDTA (99.9%, Alfa Aesar) solution (50 ml) was adding into the above mixture and thoroughly mixed in an ultrasonic bath for 20 min. Prior to CBD, the growth solution was transferred to a hot plate under constant stirring and heated to, and maintained at, the desired temperature. The growth substrate was suspended vertically in the solution, and deposition was initiated by mixing in 20 ml of the sulfur precursor (0.1 M C_2H_5NS (99.9%, Alfa Aesar). During the reaction, the beaker was covered to limit water evaporation. After the desired growth time (30-120 minutes), the substrate was removed and rinsed with DIW. The obtained samples were cut into desired size for further characterization.

P-TCM characterization. Film thickness was measured with a step profiler (Veeco-Dektak 150). Atomic force microscopy (AFM) was used to determine surface roughness and morphology using a Veeco-DI instrument equipped with a Nanoscope IV Controller and silicon tip from MikroMasch. UV-vis transmittance and reflectance spectra of samples were obtained using a Lambda 950 UV-vis spectrometer. Room temperature conductivity, mobility, and carrier concentration were measured using the van der Pauw contact configuration with an Ecopia HMS-3000 Hall Measurement System equipped with a 0.55 T magnet.[19] The room temperature Seebeck coefficients were measured by a home-built thermopower system equipped with Keithley 2000-20 multimeter and Lakeshore 331 temperature controllers. All CZS films were p-type as evidenced by positive Hall and Seebeck coefficients.

Design of experiments and p-TCM figure of merit. In the first round of experiments, we chose two levels for the Cu content, Cu/(Cu+Zn), and 4 levels for the other three parameters: reaction temperature, reaction time, and concentration of Na_2EDTA. The parameter levels were based on values which produced higher performing films in our previous study.[29] Varying all the parameters independently would necessitate 128 experimental conditions. As repeat experiments are required to establish reproducibility, the full experimental design is clearly too time consuming.[51] Instead, we employed an approximate 1/4 fractional factorial design (16 conditions) plus 3 additional conditions for a total of 19 sets, as shown in Table 1. Either 3 or 4 independent depositions were done at each condition to ensure reproducibility, leading to a total of 62 independent experiments in the first round.

Figures of Merits (FOMs), e.g. *ZT* for thermoelectrics, are commonly used in engineering to evaluate the applicability of a materials with different underlying properties.[20,52,53] In this study, we applied an FOM for transparent conductors proposed by Haacke,[54]

$$\Phi_{TC}=T^{10}/R_s \qquad (1)$$

where T is the transmission at a given wavelength (600 nm in this work) or over a wavelength range and R_s ($\Omega/\square$) is sheet resistance.[20,54] The exponent of T in Eq. (1) weights the importance of optical transmission vs. electrical conduction for a given application; an exponent of 10 targets a transmission of greater than 90%.[54] For comparison to the values obtained in this study for p-type films, FOMs calculated with Eq (1) for commercial n-type SnO_2-coated glass (PPG) and In_2O_3-coated glass (PPG) are 900 and 7000 µS, respectively.[54] Figures of merit for p-TCM films compiled from the literature (0.01-225 µS) are listed in Table S1.

Table 1. Experimental conditions and results in the first round design of experiment (62 total experiments).

Set#	Cu/(Cu+Zn)	Temperature (°C)	Time (min)	Concentration of Na_2EDTA (M)	FOM (μS)[a]	Number of samples
1	0.65	80	60	0.03	195	3
2	0.65	80	90	0.03	301	3
3	0.65	80	120	0.03	144	3
4	0.65	90	60	0.03	240	3
5	0.65	90	120	0.03	144	3
6	0.85	80	60	0.03	98.4	3
7	0.85	80	120	0.03	65.1	3
8	0.85	90	60	0.03	114	3
9	0.85	80	60	0.1	31.5	3
10	0.85	80	90	0.1	25.9	3
11	0.85	80	120	0.1	25.8	3
12	0.85	80	90	0.02	97.1	3
13	0.85	80	60	0.02	188	3
14	0.85	90	60	0.02	205	4
15	0.65	80	90	0.07	119	4
16	0.65	60	60	0.1	123	4
17	0.65	90	30	0.03	166	4
18	0.85	70	30	0.07	140	4
19	0.85	60	90	0.03	153	4

[a] FOM values are shown as the median within samples from the same set.

Machine learning training model. There are a number of choices for a machine learning regression modeling including Linear Regression, Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, etc.[55,56] However, the so-called "no free lunch" theorem states that no algorithm is suitable for every problem and every dataset,[57] such that the choice applicable here to a multi-parameter synthesis space merits discussion. As might we expect correlations and non-

linear dependencies, a linear regression model would not be appropriate. Similarly, logistic regression assumes that factors are independent from each other thus is also inapplicable.[58] KNN makes predictions by searching through the entire training set for the most similar neighbors and is most often used for estimating continuous variables. Examination of Table 1 shows that the FOMs in our case hardly follow a simple "continuous" pattern, thus making KNN regression unsuitable in this work.[59] Random forest, one of the most popular regression algorithms in ML, uses bootstrap aggregation and a random subspace method.[60,61] However, there is a risk of overfitting, and this approach typically does not do well when extrapolation outside of the range of variables is required.

In this work, we employed support vector regression (SVR) implemented with Scikit-learn, a popular ML library for the Python programming language.[62] There are a number of different kernel functions which can be employed including linear, nonlinear, polynomial and sigmoid. We selected the radial basis function (RBF) kernel, as it would be expected to fit the Gaussian-shaped features that would normally be found in cases of optimization.[63–65] As discussed in detail below, a 10-fold cross validation scheme was applied to mitigate against overfitting and artificially high statistics (unreliable statistic results from sampling).[66] More details can be found in the Supporting Information, along with the Python codes we employed.

Device fabrication and measurement. Proof-of-concept p-CZS/n-Si, p-CZS/np$^+$-Si heterojunction solar cells and p-CZS/n-ZnO photodiodes were fabricated to demonstrate the improved device performance using optimized films. For the solar cells,[29] dot contacts were made by electron-beam evaporating Cu/Au (4 nm/60 nm) on the CZS film, while liquid InGa was smeared on the scratched rear surface on n-Si and np$^+$-Si as the back contact. For the photodiodes,[67] CZS films were deposited on FTO glasses via two steps of CBD processes, followed by six layers

of n-ZnO deposited by sol-gel spin coating. Each coating of ZnO was followed by an annealing process (300 °C for 2 min) on a hot plate to evaporate the organic solvent. Dot contacts were made by electron-beam evaporating 60 nm of Au on ZnO surface. The performance of the devices was analyzed by light and dark current density-voltage (*J-V*) measurements using a standard solar simulator (Xe lamp, Solar Co.) under AM 1.5G irradiation (100 mW cm^{-2}) and in the dark.

RESULTS & DISCUSSION

First round statistics. Observed FOMs for the CBD p-TCM films ranged from 25 to 304 µS and were reproducible at a given experimental condition, as shown in the box plot in Figure 2a. Analysis of variance (ANOVA) was employed to determine the relative influence of each of the 4 factors on the observed FOM. The p-values for all factors were small, showing that all the factors influence the FOM of the films (Table S2). Figure 2b compares the fraction of the total variance associated with each factor (the residual variance, see Table S2, is not included such that the contributions add to 1). It can be seen that the Cu/(Cu+Zn) and the Na$_2$EDTA concentration are the most influential factors while the process temperature contributes the least variance. For this reason, we decided to exclude the temperature from the ML model.
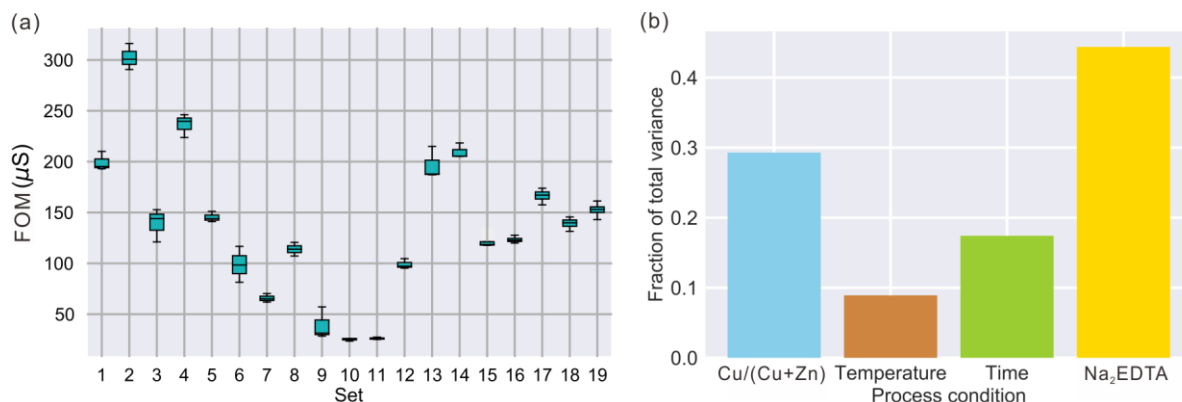

Figure 2. (a) Boxplot of FOMs grouped by set and (b) analysis of variance and factor evaluation from the first round of optimization.

Cross validation and first round optimization. Support vector regression (SVR) models with a radial basis function (RBF) employs a kernel function of the form: [68–71]

$$K_{RBF}(x,x') = \exp[-\| x-x' \|^2 /2\sigma^2]. \tag{2}$$

In Eq. (2), the independent variables x are the three factors, Cu/(Cu+Zn), reaction time and concentration of Na_2EDTA, while the dependent variable (K_{RBF}) is corresponding FOM. It is important to consider the values of several hyperparameters: ε specifies the epsilon-tube within which no penalty is associated in the training loss function and thus determines support vectors (details can be found in SI); C, the so-called penalty coefficient of SVM, determines tolerance level for errors; finally, γ ($\gamma =1/2\sigma^2$) defines how far the influence of a single training example reaches. Large C and γ values result in overfitting while small C and γ values lead to underfitting.[72,73] Overfitting happens when a ML model overemphasizes the data on which it was trained and therefore loses its applicability to other datasets. Underfitting happens when a ML model fails to accurately capture relationships between a dataset's parameters and its target variable.

We thus employed 10-fold cross validation to choose optimal values for C and γ. The data were randomly partitioned into 10 equal sized subsamples: 9 subsamples were used as training data, the remaining one as validation data for testing model. The cross validation then repeat for 10 times, with each subsample is used exactly once as the validation data.[74] Testing results are evaluated by mean-squared-error (MSE)[75] in the regression problem, where the smaller MSE is expected for better fitting. Both the input variables (parameters) and output variables (FOMs) were preprocessed to have a mean of zero and a standard deviation of one; [62] This step is necessary when the input variables have different scales. Cross validation learning curves for C and γ, scored by the MSE as mentioned before,[76] are shown in Figure 3. In Figure 3a, the red line corresponding to

the training data decrease as γ is increased from 0 to 0.1 and then remains constant for $\gamma > 0.1$.

The minimum loss for the cross validation data (green line) occurs near $\gamma = 0.3$, indicating possibilities of overfitting with too small or too high γ. In Figure 3b, a similar tendency can be found with training data, with low loss being obtained for $C > 23$. A clear minimum in the cross validation loss was not found, but $C = 40$ (and $\gamma = 0.3$) were ultimately selected as a good balance between under and overfitting.[77] Cross validation was also employed to select the value for ε (0.1) as shown in Figure S3.

Figure 3. Learning curve for training and cross validation process based on (a) γ value (when C=40) and (b) C value (when γ=0.3) using normalized data. Loss is defined as mean squared error (MSE).

Using parameters discussed above, the first round fitting results are showing in Figure 4a. The dashed line represents equal measured (x-axis) and predicted (y-axis) FOMs. There is good agreement between the SVR model and the experimental data; Figure 4b shows that low and high FOMs are approximately equally well predicted with most values lying within a range of $\pm 20\%$.
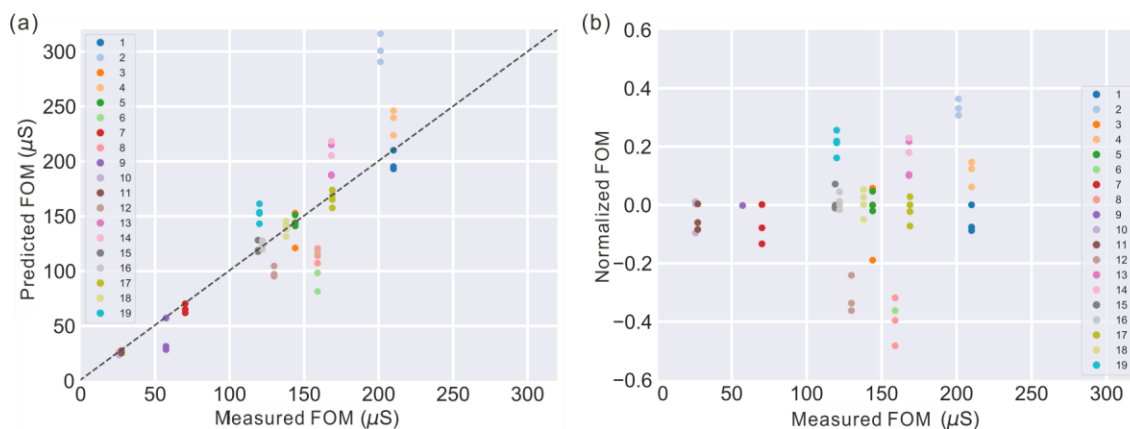
Figure 4. Support vector regression applied to the first round of experiments (a) Predicted versus measured FOM and (b) Normalized FOM versus measured FOM.

The SVR model can be used to create a three-dimensional map to predict the FOM at any point within the sampled synthesis space. To aid visualization, Cu/(Cu+Zn) was separated into two different levels as subplot figures, thus two-dimensional contour plots are shown in Figure 5, with reaction time as the x axis and the concentration of Na_2EDTA as the y axis. The color gradient, scaled as indicated by the color bar on the right, and the contour lines map out how FOM changes with different reaction time and concentration of Na_2EDTA.[49] For Cu/(Cu+Zn)=0.65, an area of interest can be found in the range 60-100 min of reaction time and 0.02-0.04 M of Na_2EDTA concentration, where FOMs more than 200 µS are predicted obtained. For Cu/(Cu+Zn)=0.85, situation becomes more complicated, as we anticipated from the complexity of the deposition process, but still it can be predicted that high FOM films can be obtained around 60 min and with low Na_2EDTA concentration.

Relating the results to the CBD process, higher Cu/(Cu+Zn) produces better conductivity but decreases the transparency. The definition of FOM (Eq. (1) emphasizes transparency, thus films with Cu/(Cu+Zn)=0.65 yield higher FOMs in general. It is also interesting that high FOMs

14

obtained in the range 60-100 min of reaction time. In our previous work, film thickness tends to be increased smoothly after 60 min[29], we thus assumed that longer reaction time could cause poor quality growth on surface, which influences both film conductivity and transparency. Further, FOMs decrease with higher Na_2EDTA concentration, which may because of poor film quality and nonuniformity,[36] as will be revealed in our case by AFM, discussed below.
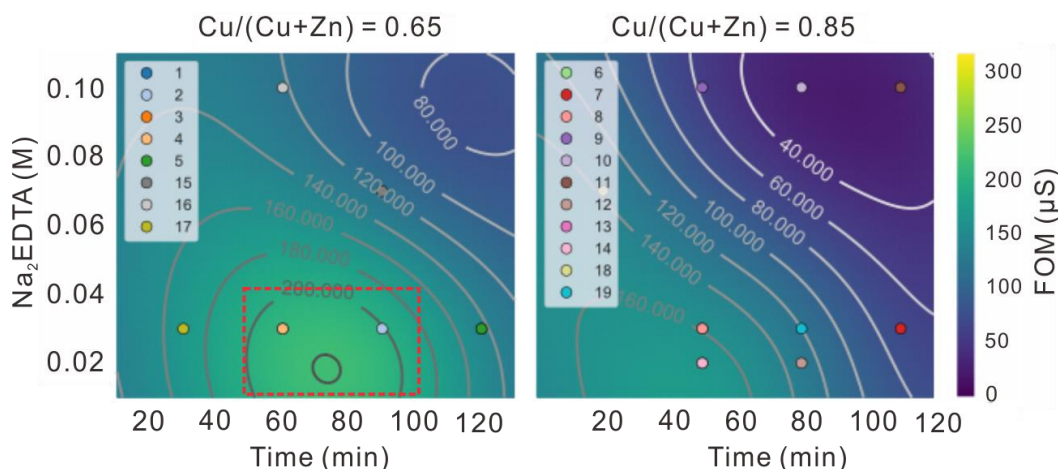

Figure 5. Contour maps of predicted FOMs as a function of deposition time and Na_2EDTA concentration from the first round of experiments. Dashed red line shows the area of interest selected for the second round of optimization. FOM scale and contours were chosen to emphasize the full range of the predicted FOMs.

Data evaluation and second round optimization. Based on results from the first round of optimization, the range of factors can be further narrowed. For the second round of optimization, the reaction temperature was kept at 80 °C as it has comparatively little influence on FOM. The experiments were thus designed with only three factors: Cu/(Cu+Zn), reaction time, concentration of Na_2EDTA. The experimental design is shown in Table 2: the range of Cu/(Cu+Zn) was examined between 0.6-0.7, reaction time between 60-80 min and Na_2EDTA concentration between 0.02-0.04 M. A full factorial design with three factors and three levels would require a minimum

of 27 experiments, exclusive of repeats. Here, a 1/3 fractional factorial DOE allows us to use only 9 conditions and an additional 3 with 4 repeats, yielding 48 experiments in total. As with the first round, reproducibility at a given condition was very good, as illustrated by the box plot in Figure S1a. Over this range of parameters, ANOVA showed that the Na_2EDTA concentration was responsible for the most variance, followed by deposition time and Cu/(Cu+Zn), Figure S1b.

Table 2. Experimental conditions and results in the second round design of experiment.

Set#	Cu/(Cu+Zn)	Temperature (ºC)	Time (min)	Concentration of Na_2EDTA (M)	FOM (µS)[a]	Number of samples
1	0.6	80	70	0.04	197	4
2	0.6	80	80	0.03	236	4
3	0.6	80	70	0.02	180	4
4	0.6	80	60	0.03	188	4
5	0.65	80	70	0.03	245	4
6	0.65	80	60	0.04	237	4
7	0.65	80	80	0.02	179	4
8	0.65	80	80	0.04	198	4
9	0.7	80	80	0.02	168	4
10	0.7	80	60	0.03	218	4
11	0.7	80	70	0.04	186	4
12	0.7	80	60	0.02	217	4

[a] FOM values are shown as the median within samples from the same set.

Considering that the second round data were generated from the first found optimization, the same RBF values of $C = 40$ and $\gamma = 0.3$ and $\varepsilon = 0.1$ were applied here. As shown in Figure S2, the model fits the data well and without apparent bias, with most values predicted to within a range of $\pm 10\%$. Figure 6 shows contour maps for the three levels of Cu/(Cu+Zn): 0.6, 0.65 and 0.7. The analysis allows a few observations to be made. (1) FOMs greater than 180 µS can be obtained for

each of the Cu/(Cu+Zn) levels, albeit with different deposition times and EDTA concentrations. This result is consistent with the region of interest selected after the first optimization round, Figure 5a. (2) Reaction times between 60 and 80 minutes produce higher FOMs compared to shorter or longer time for all three Cu/(Cu+Zn) levels. If the films resistivity did not vary with deposition time, longer reaction would decrease sheet resistance and transparency. However, the CBD process is self-limiting as the precursor concentrations decline in time, such that prolonged reaction times may decrease the overall quality of film, leading to a further decrease in the FOM. (3) Finally, at all Cu/(Cu+Zn) levels, a concentration of Na_2EDTA around 0.03 M produces the highest FOMs, which is consistent with a more uniform (and small) crystals, as shown below.
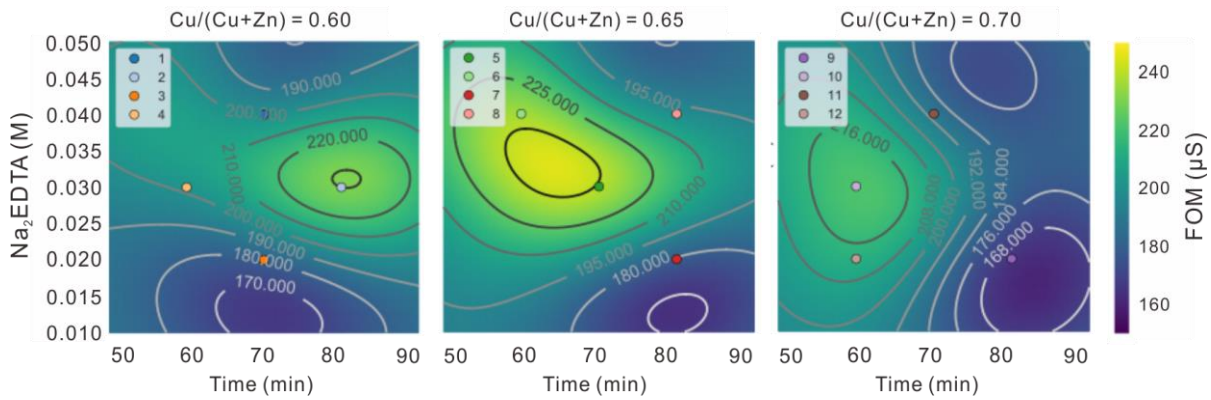
Figure 6. Contour maps of predicted FOMs as a function of deposition time and Na_2EDTA concentration from the second round of experiments. Contours chosen to emphasize maxima.

Comparison to literature reports. Table S1 tabulates calculated FOMs (using Eq (1) and $\lambda = 600$ nm) of p-TCMs deposited via different methods including pulsed laser deposition, sputtering, spray pyrolysis, and CBD. Values obtained here from two rounds of optimization compare well to reports of Cu-Zn-S and Cu-Al-S made by CBD and sputtering.

Optical properties and surface imaging. Two conditions were selected from SVR optimization as optimized and non-optimized thin films for further analysis. Table 3 gives the reaction conditions and film performance for them. A higher transmittance (Figure 7a) and lower sheet resistance are exhibited by the optimized film, thus yielding an excellent FOM. The Tauc plot in Figure 7b indicates that the band gaps of the two film are 2.7 and 2.5 eV, respectively. AFM images in Figure 7c and d show a uniform surface with homogeneous distributed small particles for the optimized film (R_q=7.9, indicating a small surface roughness), while the non-optimized film has a nonuniform surface with scattered large particles (R_q=13.8).

Table 3. Comparison of optical and electrical properties for optimized and non-optimized films.

Film	Cu/(Cu+Zn)	Temperature (°C)	Time (min)	Con. of Na$_2$EDTA (M)	Transmittance (600 nm, %)	Sheet resistance (Ω/□)	FOM (µS)[a]
Non-optimized	0.85	80	60	0.1	0.67	554	31.5
Optimized	0.65	80	70	0.03	0.78	360	245

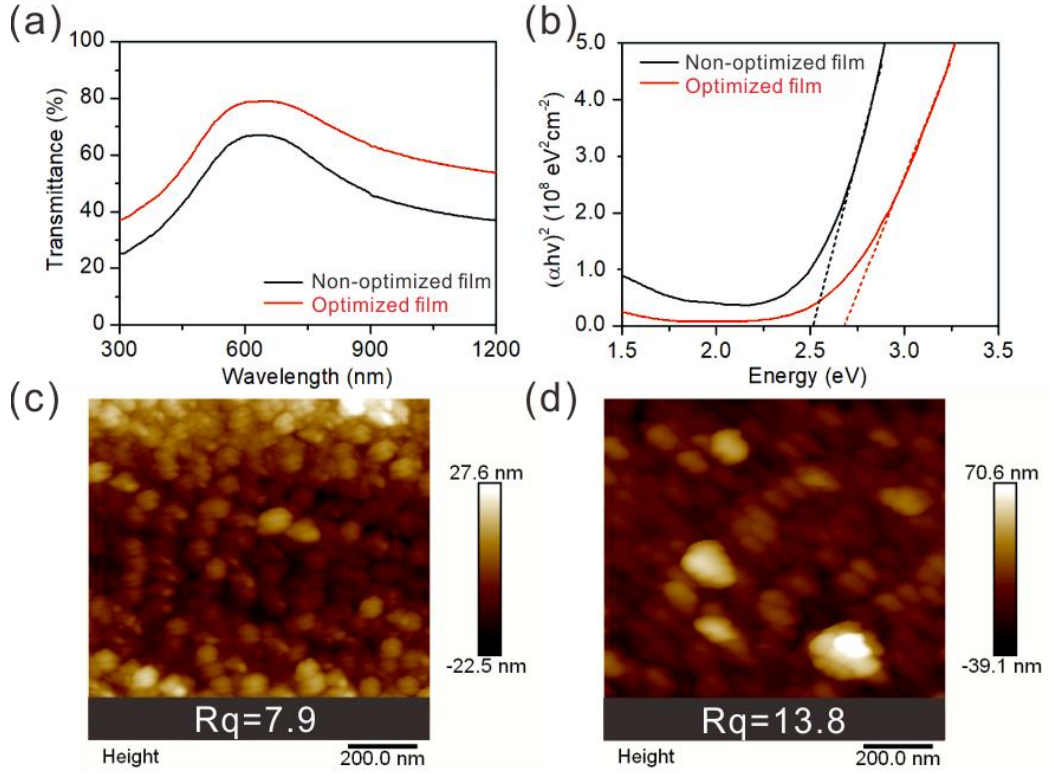[a] FOM values are shown as the median within samples from the same set.

Figure 7. (a) Transmission spectra and (b) Tauc plot of optimized and non-optimized CZS films; AFM images of (c) optimized and (d) non-optimized CZS films.

Devices. To assess the performance of optimized and non-optimized films in optoelectronic devices, p-CZS was employed as the hole selective contact in Si heterojunction solar cells and hole transport layer in p-CZS/n-ZnO photodiodes. CZS thin films in the devices were controlled to possess almost the same thickness and reflective index. Schematic structures and current density-voltage curves of devices are shown in Figure 8. For the p-CZS/n-Si and p-CZS/p$^+$n-Si solar cells shown in Figures 8a and b a clear enhancement in device efficiency can be observed in spite of issues with series and shunt resistance in the devices (Figure 8d and e). For photodiodes in Figure 8c, the device with optimized CZS film generates a much higher rectification ratio of 200 at ± 1 V bias, 50 times the device with non-optimized CZS film (Figure 8f).
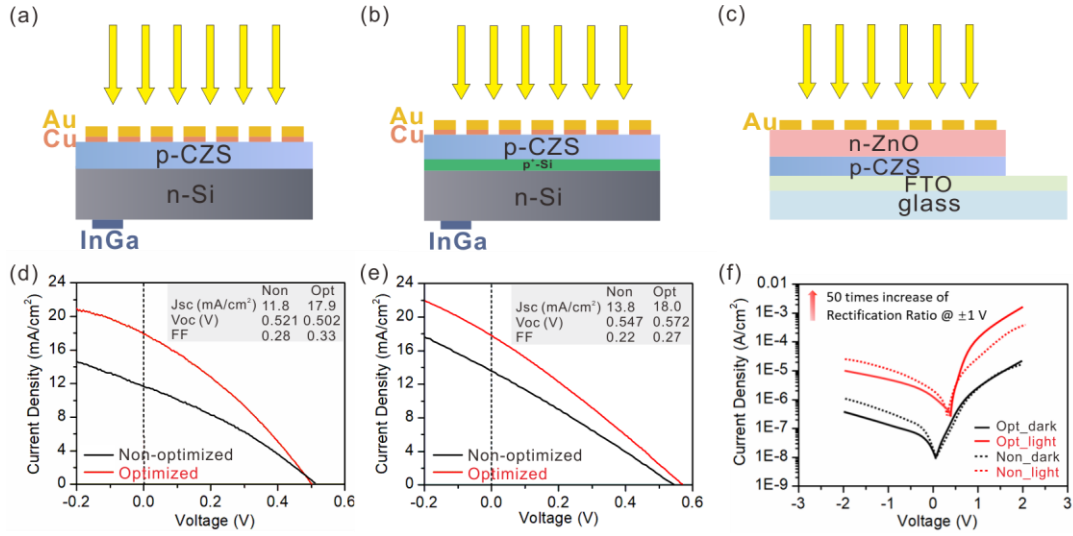
19

Figure 8. Schematic structure of (a) p-CZS/n-Si, (b) p-CZS/ p$^+$n-Si and (c) p-CZS/n-ZnO photodiodes; Photovoltaic performance of (d) p-CZS/n-Si, (e) p-CZS/ p$^+$n-Si and (f) p-CZS/n-ZnO photodiodes. ('Opt' represents optimized film while 'Non' represents non-optimized film)

CONCLUSIONS

Machine learning optimization in combination with design of experiment is shown to be effective in finding experimental conditions which produce p-type transparent films with high conductivity. Use of support vector regression produces a map of the chemical bath deposition synthesis space; conditions which produce peak performance can be predicted and, importantly, insights into the interactions between the deposition parameters can be obtained. Optimized p-TCM films perform well in devices such as heterojunction solar cells and photodiodes.

While the approach here could have general applicability in materials synthesis, we do need to acknowledge some limitations and challenges. We emphasize the "no free lunch" concept that no single machine learning approach is applicable to every regression problem. However, an advantage of the SVR approach taken here is that we obtain insight both from "successful" (high FOM) and "failed" (low FOM) experiments through examination of the model's predictions. It

may be possible to generalize this concept, as suggested by the proposal of Raccuglia *et al.*[78] to construct an on-line "dark reactions project" documenting experiments (failed and successful) and analyzing the outcomes using a series of regression models.

ASSOCIATED CONTENT

Supporting Information. FOMs from literature for p-TCMS; ANOVA results from first round of experiments; support vector regression; Python code. This material is available free of charge via the Internet at http://pubs.acs.org. Python codes and datasets also available on GitHub.

AUTHOR INFORMATION

Corresponding Author

*E-mail: JWAger@lbl.gov.

Author Contributions

LFW and JWA designed the study and LFW performed synthesis and characterization and developed the SVR regression model. XJX provided key expertise on the CBD process and G and JB contributed expertise with fabrication and testing of solar cells and photodiodes. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ORCID

Lingfei Wei: 0000-0002-4758-1426

Xiaojie Xu: 0000-0002-2158-6958

Gurudayal: 0000-0002-9245-2199

James Bullock: 0000-0001-7903-9642

Joel W. Ager: 0000-0001-9334-9751

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMENT

REFERENCES

(1)   Ginley, D. S.; Perkins, J. D. Transparent Conductors. In *Handbook of Transparent Conductors*; 2011.

(2)   Ellmer, K. Past Achievements and Future Challenges in the Development of Optically Transparent Electrodes. *Nat. Photonics* **2012**, *6*, 809–817.

(3)   Varley, J. B.; Miglio, A.; Ha, V.-A.; van Setten, M. J.; Rignanese, G.-M.; Hautier, G. High-Throughput Design of Non-Oxide p-Type Transparent Conducting Materials: Data Mining, Search Strategy, and Identification of Boron Phosphide. *Chem. Mater.* **2017**, *29*, 2568–2573.

(4)   Granqvist, C. G. Transparent Conductors as Solar Energy Materials: A Panoramic Review. *Sol. Energy Mater. Sol. Cells* **2007**, *91*, 1529–1598.

(5)   Nagarajan, R.; Duan, N.; Jayaraj, M. .; Li, J.; Vanaja, K. .; Yokochi, A.; Draeseke, A.; Tate, J.; Sleight, A. . P-Type Conductivity in the Delafossite Structure. *Int. J. Inorg. Mater.* **2001**, *3*, 265–270.

(6)   Fleischer, K.; Norton, E.; Mullarkey, D.; Caffrey, D.; Shvets, I. V. Quantifying the

Performance of P-Type Transparent Conducting Oxides by Experimental Methods. *Materials (Basel).* **2017**, *10*, 1019.

(7) Stoica, M.; S Lo, C. P-Type Zinc Oxide Spinels: Application to Transparent Conductors and Spintronics. *New J. Phys.* **2014**, *16*, 055011.

(8) Asbalter, J.; Subrahmanyam, A. P -Type Transparent Conducting In2O3–Ag2O Thin Films Prepared by Reactive Electron Beam Evaporation Technique. *J. Vac. Sci. Technol. A Vacuum, Surfaces, Film.* **2000**, *18*, 1672–1676.

(9) Zhang, K. H. L.; Xi, K.; Blamire, M. G.; Egdell, R. G. P -Type Transparent Conducting Oxides. *J. Phys. Condens. Matter* **2016**, *28*, 383002.

(10) Park, S.; Keszler, D. a.; Valencia, M. M.; Hoffman, R. L.; Bender, J. P.; Wager, J. F. Transparent P-Type Conducting BaCu[Sub 2]S[Sub 2] Films. *Appl. Phys. Lett.* **2002**, *80*, 4393–4394.

(11) Hiramatsu, H.; Ueda, K.; Ohta, H.; Hirano, M.; Kamiya, T.; Hosono, H. Degenerate P-Type Conductivity in Wide-Gap LaCuOS1−xSex (X=0–1) Epitaxial Films. *Appl. Phys. Lett.* **2003**, *82*, 1048–1050.

(12) Liu, M.-L.; Huang, F.-Q.; Chen, L.-D.; Wang, Y.-M.; Wang, Y.-H.; Li, G.-F.; Zhang, Q. P-Type Transparent Conductor: Zn-Doped CuAlS2. *Appl. Phys. Lett.* **2007**, *90*, 072109.

(13) Liu, M.-L.; Huang, F.-Q.; Chen, L.-D. P-Type Electrical Conduction and Wide Optical Band Gap in Mg-Doped CuAlS2. *Scr. Mater.* **2008**, *58*, 1002–1005.

(14) Yang, K.; Ichimura, M. Fabrication of Transparent P-Type Cu x Zn y S Thin Films by the Electrochemical Deposition Method. *Jpn. J. Appl. Phys.* **2011**, *50*, 40202.

(15) Parreira, P.; Lavareda, G.; Amaral, A.; Botelho do Rego, A. M.; Conde, O.; Valente, J.; Nunes, F.; Nunes de Carvalho, C. Transparent P-Type CuxS Thin Films. *J. Alloys Compd.* **2011**, *509*, 5099–5104.

(16) Chaudhary, N.; Chaudhary, R.; Kesari, J. P.; Patra, A.; Chand, S. Copper Thiocyanate (CuSCN): An Efficient Solution-Processable Hole Transporting Layer in Organic Solar Cells. *J. Mater. Chem. C* **2015**, *3*, 11886–11892.

(17) Kudo, A.; Sekizawa, M. Photocatalytic H2 Evolution under Visible Light Irradiation on Zn1-x Cu x S Solid Solution. *Catal. Letters* **1999**, *58*, 241–243.

(18) Diamond, A. M.; Corbellini, L.; Balasubramaniam, K. R.; Chen, S.; Wang, S.; Matthews, T. S.; Wang, L.-W.; Ramesh, R.; Ager, J. W. Copper-Alloyed ZnS as a p-Type Transparent Conducting Material. *Phys. status solidi* **2012**, *209*, 2101–2107.

(19) Woods-Robinson, R.; Cooper, J. K.; Xu, X.; Schelhas, L. T.; Pool, V. L.; Faghaninia, A.; Lo, C. S.; Toney, M. F.; Sharp, I. D.; Ager, J. W. P-Type Transparent Cu-Alloyed ZnS Deposited at Room Temperature. *Adv. Electron. Mater.* **2016**, *2*, 1500396.

(20) Maurya, S. K.; Liu, Y.; Xu, X.; Woods-Robinson, R.; Das, C.; Ager, J. W.; Balasubramaniam, K. R. High Figure-of-Merit p -Type Transparent Conductor, Cu Alloyed ZnS via Radio Frequency Magnetron Sputtering. *J. Phys. D. Appl. Phys.* **2017**, *50*, 505107.

(21) Parreira, P.; Lavareda, G.; Valente, J.; Nunes, F. T.; Amaral, A.; de Carvalho, C. N.

Optoelectronic Properties of Transparent P-Type Semiconductor CuxS Thin Films. *Phys. status solidi* **2010**, *207*, 1652–1654.

(22) Kang, S. R.; Shin, S. W.; Choi, D. S.; Moholkar, A. V.; Moon, J.-H.; Kim, J. H. Effect of PH on the Characteristics of Nanocrystalline ZnS Thin Films Prepared by CBD Method in Acidic Medium. *Curr. Appl. Phys.* **2010**, *10*, S473–S477.

(23) Dula, M.; Yang, K.; Ichimura, M. Photochemical Deposition of a P-Type Transparent Alloy Semiconductor Cu x Zn y S. *Semicond. Sci. Technol.* **2012**, *27*, 125007.

(24) Yang, K.; Nakashima, Y.; Ichimura, M. Electrochemical Deposition of CuxS and CuxZnyS Thin Films with P-Type Conduction and Photosensitivity. *J. Electrochem. Soc.* **2012**, *159*, H250–H254.

(25) Shin, S. W.; Agawane, G. L.; Gang, M. G.; Moholkar, A. V.; Moon, J.-H.; Kim, J. H.; Lee, J. Y. Preparation and Characteristics of Chemical Bath Deposited ZnS Thin Films: Effects of Different Complexing Agents. *J. Alloys Compd.* **2012**, *526*, 25–30.

(26) Ortíz-Ramos, D. E.; González, L. A.; Ramirez-Bon, R. P-Type Transparent Cu Doped ZnS Thin Films by the Chemical Bath Deposition Method. *Mater. Lett.* **2014**, *124*, 267–270.

(27) Ichimura, M.; Maeda, Y. Heterojunctions Based on Photochemically Deposited CuxZnyS and Electrochemically Deposited ZnO. *Solid. State. Electron.* **2015**, *107*, 8–10.

(28) Hong, J.; Lim, D.; Eo, Y.-J.; Choi, C. Chemical Bath Deposited ZnS Buffer Layer for Cu(In,Ga)Se 2 Thin Film Solar Cell. *Appl. Surf. Sci.* **2018**, *432*, 250–254.

(29) Xu, X.; Bullock, J.; Schelhas, L. T.; Stutz, E. Z.; Fonseca, J. J.; Hettick, M.; Pool, V. L.; Tai, K. F.; Toney, M. F.; Fang, X.; Javey, A.; Wong, L. H.; Ager, J. W. Chemical Bath Deposition of P-Type Transparent, Highly Conducting (CuS)x:(ZnS)1-x Nanocomposite Thin Films and Fabrication of Si Heterojunction Solar Cells. *Nano Lett.* **2016**, *16*, 1925–1932.

(30) Schneller, T.; Waser, R. *Chemical Solution Deposition of Functional Oxide Thin Films*; Schneller, T., Waser, R., Kosec, M., Payne, D., Eds.; Springer Vienna: Vienna, 2013.

(31) Zhang, H.; Ma, X.; Yang, D. Effects of Complexing Agent on CdS Thin Films Prepared by Chemical Bath Deposition. *Mater. Lett.* **2004**, *58*, 5–9.

(32) Nemade, K. R.; Waghuley, S. A. Band Gap Engineering of CuS Nanoparticles for Artificial Photosynthesis. *Mater. Sci. Semicond. Process.* **2015**, *39*, 781–785.

(33) Kumbhojkar, N.; Nikesh, V. V.; Kshirsagar, A.; Mahamuni, S. Photophysical Properties of ZnS Nanoclusters. *J. Appl. Phys.* **2000**, *88*, 6260–6264.

(34) Prem kumar, T.; Sankaranarayanan, K. Effect of EDTA Concentration on the Physical and Optical Properties of Cds Thin Films. *Can. J. Chem. Eng.* **2013**, *91*, 27–33.

(35) Yi, G.; Lu, H.; Zhao, S.; Ge, Y.; Yang, W.; Chen, D.; Guo, L.-H. Synthesis, Characterization, and Biological Application of Size-Controlled Nanocrystalline NaYF 4 :Yb,Er Infrared-to-Visible Up-Conversion Phosphors. *Nano Lett.* **2004**, *4*, 2191–2196.

(36) Park, S. Y.; Park, J. E.; Eom, T.; Park, J. H.; Bweupe, J.; Lim, D. Characterization of ZnS Thin Films Grown Using Chemical Bath Deposition with Three Different Complexing

Agents. *J. Nanosci. Nanotechnol.* **2018**, *18*, 6294–6299.

(37) Mane, R. S.; Lokhande, C. D. Chemical Deposition Method for Metal Chalcogenide Thin Films. *Mater. Chem. Phys.* **2000**, *65*, 1–31.

(38) Ortega-López, M. Improved Efficiency of the Chemical Bath Deposition Method during Growth of ZnO Thin Films. *Mater. Res. Bull.* **2003**, *38*, 1241–1248.

(39) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547–555.

(40) Pilania, G.; Whittle, K. R.; Jiang, C.; Grimes, R. W.; Stanek, C. R.; Sickafus, K. E.; Uberuaga, B. P. Using Machine Learning to Identify Factors That Govern Amorphization of Irradiated Pyrochlores. *Chem. Mater.* **2017**.

(41) Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K.-R.; Singh, A. K. Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXene. *Chem. Mater.* **2018**, *30*, 4031–4038.

(42) Turcani, L.; Greenaway, R. L.; Jelfs, K. E. Machine Learning for Organic Cage Property Prediction. *Chem. Mater.* **2019**, *31*, 714–727.

(43) Oliynyk, A. O.; Adutwum, L. A.; Rudyk, B. W.; Pisavadia, H.; Lotfi, S.; Hlukhyy, V.; Harynuk, 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**.

(44) 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.

(45) Bigun, I.; Steinberg, S.; Smetana, V.; Mudryk, Y.; Kalychak, Y.; Havela, L.; Pecharsky, V.; Mudring, A.-V. Magnetocaloric Behavior in Ternary Europium Indides EuT 5 In: Probing the Design Capability of First-Principles-Based Methods on the Multifaceted Magnetic Materials. *Chem. Mater.* **2017**, *29*, 2599–2614.

(46) Singh, A. K.; Zhou, L.; Shinde, A.; Suram, S. K.; Montoya, J. H.; Winston, D.; Gregoire, J. M.; Persson, K. A. Electrochemical Stability of Metastable Materials. *Chem. Mater.* **2017**, *29*, 10159–10167.

(47) He, Y.; Galli, G. Perovskites for Solar Thermoelectric Applications: A First Principle Study of CH 3 NH 3 AI 3 (A = Pb and Sn). *Chem. Mater.* **2014**, *26*, 5394–5400.

(48) Muñoz-García, A. B.; Pavone, M. First-Principles Design of New Electrodes for Proton-Conducting Solid-Oxide Electrochemical Cells: A-Site Doped Sr 2 Fe 1.5 Mo 0.5 O 6−δ Perovskite. *Chem. Mater.* **2016**, *28*, 490–500.

(49) Cao, B.; Adutwum, L. A.; Oliynyk, A. O.; Luber, E. J.; Olsen, B. C.; Mar, A.; Buriak, J. M. How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano* **2018**, *12*, 7434–7444.

(50) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos,

G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120–1127.

(51) Wulff, S. S. A First Course in Design and Analysis of Experiments. *Am. Stat.* **2003**, *57*, 66–67.

(52) Lu, X.; Morelli, D. T.; Xia, Y.; Ozolins, V. Increasing the Thermoelectric Figure of Merit of Tetrahedrites by Co-Doping with Nickel and Zinc. *Chem. Mater.* **2015**, *27*, 408–413.

(53) Du, B.; Li, H.; Xu, J.; Tang, X.; Uher, C. Enhanced Figure-of-Merit in Se-Doped p-Type AgSbTe 2 Thermoelectric Compound. *Chem. Mater.* **2010**, *22*, 5521–5527.

(54) Haacke, G. New Figure of Merit for Transparent Conductors. *J. Appl. Phys.* **1976**, *47*, 4086–4089.

(55) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260.

(56) Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning*; Cambridge University Press: Cambridge, 2014.

(57) Wolpert, D. H.; Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82.

(58) Scott, A. J.; Hosmer, D. W.; Lemeshow, S. Applied Logistic Regression. *Biometrics* **1991**, *47*, 1632.

(59) Peterson, L. K-Nearest Neighbor. *Scholarpedia* **2009**, *4*, 1883.

(60) Ho, T. K. A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors. *Pattern Anal. Appl.* **2002**, *5*, 102–112.

(61) Tin Kam Ho. The Random Subspace Method for Constructing Decision Forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.

(62) 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 {P}ython. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(63) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140*, 9844–9853.

(64) Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222.

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

(66) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; Routledge, 2018.

(67) Xu, X.; Shukla, S.; Liu, Y.; Yue, B.; Bullock, J.; Su, L.; Li, Y.; Javey, A.; Fang, X.; Ager, J. W. Solution-Processed Transparent Self-Powered p-CuS-ZnS/n-ZnO UV Photodiode. *Phys. status solidi - Rapid Res. Lett.* **2018**, *12*, 1700381.

(68) Scholkopf, B.; Kah-Kay Sung; Burges, C. J. C.; Girosi, F.; Niyogi, P.; Poggio, T.; Vapnik, V. Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers. *IEEE Trans. Signal Process.* **1997**, *45*, 2758–2765.

(69) Montegranario, H.; Espinosa, J. Radial Basis Functions. In *SpringerBriefs in Computer Science*; 2014; pp 69–81.

(70) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**.

(71) Clarke, S. M.; Griebsch, J. H.; Simpson, T. W. Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses. *J. Mech. Des.* **2005**, *127*, 1077.

(72) Huang, C.-L.; Wang, C.-J. A GA-Based Feature Selection and Parameters Optimizationfor Support Vector Machines. *Expert Syst. Appl.* **2006**, *31*, 231–240.

(73) J. Weston and S. Mukherjee and O. Chapelle and M. Pontil and T. Poggio and V. Vapnik. Feature Selection for SVMs. In *Advances in Neural Information Processing Systems 13*; 2000.

(74) Golub, G. H.; Heath, M.; Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **1979**, *21*, 215–223.

(75) Efron, B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *J. Am. Stat. Assoc.* **1983**, *78*, 316–331.

(76) Cherkassky, V.; Ma, Y. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks* **2004**, *17*, 113–126.

(77) Huang, C.-L.; Chen, M.-C.; Wang, C.-J. Credit Scoring with a Data Mining Approach Based on Support Vector Machines. *Expert Syst. Appl.* **2007**, *33*, 847–856.

(78) 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.

SYNOPSIS AND TOC GRAPHIC
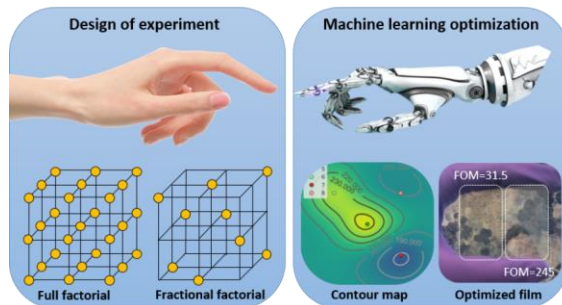
Strategic design of experiments combined with machine learning method is used to optimize the synthesis of p-type transparent conductors.