ADVANCED ENERGY MATERIALS
www.advenergymat.de

Machine Learning for Ultra High Throughput Screening of Organic Solar Cells: Solving the Needle in the Haystack Problem

*Markus Hußner, Richard Adam Pacalaj, Gerhard Olaf Müller-Dieckert, Chao Liu, Zhisheng Zhou, Nahdia Majeed, Steve Greedy, Ivan Ramirez, Ning Li, Seyed Mehrdad Hosseini, Christian Uhrich, Christoph Josef Brabec, James Robert Durrant, Carsten Deibel, and Roderick Charles Ian MacKenzie**

Over the last two decades the organic solar cell community has synthesized tens of thousands of novel polymers and small molecules in the search for an optimum light harvesting material. These materials are often crudely evaluated simply by measuring the current–voltage (JV) curves in the light to obtain power conversion efficiencies (PCEs). Materials with low PCEs are quickly disregarded in the search for higher efficiencies. More complex measurements such as frequency/time domain characterization that could explain why the material performed as it is often not performed as they are too time consuming/complex. This limited feedback forced the field to advance using a more or less random walk of material development and has significantly slowed progress. Herein, a simple technique based on machine learning that can quickly and accurately extract recombination time constants and charge carrier mobilities as a function of light intensity simply from light/dark JV curves alone. This technique reduces the time to fully analyze a working cell from weeks to seconds and opens up the possibility of not only fully characterizing new devices as they are fabricated, but also data mining historical data sets for promising materials the community has overlooked.

1. Introduction

Over the last 22 years, organic solar cell efficiencies have risen from 2.5% in 2001[1] to over 19%[2] today. Much of this increase in performance can be attributed to steady improvement in material systems.[3,4] The first reported cells relied on blends of MEH-PPV/P3HT and C60 fullerene derivatives.[1,5] Later in the late 2000s low bandgap polymers started to emerge with alternating copolymers of fluorene with Donor-Acceptor-Donor (D-A-D) segments such as PTPTB with efficiencies ≈10%.[6] In the late 2010s the community moved away from fullerene-based acceptors to small molecules, with this came efficiencies nearing 20%.[7-9] Although efficiencies are slowly increasing at a rate of ≈1% a year it takes tremendous effort from thousands of researchers across the world to achieve this. Furthermore,

M. Hußner, R. C. I. MacKenzie
Department of Engineering
Lower Mount Joy
Durham University
South Road, Durham DH1 3LE, UK
E-mail: roderick.mackenzie@durham.ac.uk

R. A. Pacalaj, J. R. Durrant
Department of Chemistry and Centre for Processable Electronics
Imperial College London
82 Wood Lane, London W12 0BZ, UK

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aenm.202303000

DOI: 10.1002/aenm.202303000

G. Olaf Müller-Dieckert, C. Deibel
Institut für Physik
Technische Universität Chemnitz
09126 Chemnitz, Germany

C. Liu
Department of Materials Science and Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg
91054 Erlangen, Germany

C. Liu, N. Li, C. J. Brabec
Helmholtz Institute Erlangen-Nürnberg for Renewable Energy (HI ERN)
Immerwahrstrasse 2, 91058 Erlangen, Germany

Z. Zhou
State Key Laboratory of Luminescent Materials and Devices
Institute of Polymer Optoelectronic Materials and Devices
School of Materials Science and Engineering
South China University of Technology
Guangzhou 510640, China

quantities such as device life time and efficiency still need to be significantly optimized before commercialization can be considered for polymer cells.[10,11] This points to another decade of slowly improving device performance that humanity can ill afford given the rapidly rising global temperatures.[12] Part of the reason for this slow progress in organic photo-voltaics (OPV) development is a lack of timely and detailed feedback to chemists from device engineers.[13,14] Typically a new material will be synthesized and then used to fabricate a few test devices using a handful of solvents and a few annealing temperatures. Simple current–voltage (JV) curve sweeps will be performed to determine PCE, Fill Factor (FF), Open Circuit Voltage (V_{oc}) and short-circuit current (J_{sc}). These measurements will take only seconds and allow the scientist to see if the material has good photovoltaic properties. However, JV measurements will not give information as to why the device/material works well or poorly and do not give hints as to how material form/function should be improved. To obtain this information one has to perform more time consuming measurements to extract key device parameters such as recombination rate, charge carrier mobility, and measures of disorder. Examples of techniques that can extract this information are, impedance spectroscopy (IS),[15,16] Impedance Modulated Photocurrent Spectroscopy (IMPS),[17] Impedance Modulated Photovoltage Spectroscopy (IMVS),[18,19] Transient Photocurrent (TPC),[20,21] Transient Photovoltage (TPV)[22-24] and charge extraction (CE) measurements.[25,26]

Although considerable efforts have gone into refining these methods they remain complex and require expertise and equipment that is often not found in the same lab as the people with knowledge in synthesis. Other approaches to get at fundamental device parameters such as fitting numerical models to experimental data can often take longer than the experiments themselves and also require expertise and models which are rarely found in the same place as where the material is fabricated.[27] Thus very often without detailed characterization the scientist is left guessing as to why one molecule performs better than another or why devices fabricated under given conditions perform as they do. This makes it very difficult to determine the next steps in material/device optimization.

Thus one can think of the development of OPV materials as a random walk, with chemists developing new materials and disregarding the majority of them as on first glance they do not perform. Some more highly performing materials are occasionally investigated with more comprehensive methods (such as P3HT:PCBM in the past and more recently PM6:Y6). This may well have led to promising materials being disregarded and skipped over as they did not perform well in the first batch or two of fabricated devices due to selecting the wrong solvents/annealing conditions or molecular weights. We are in effect searching for a needle in the haystack but in the dark.

Although this problem is serious in the academic setting where a researcher may make a new material every few weeks, it is much worse in high through-put labs where new materials are generated daily. Candidate materials are often only tested against a few standard combinations of donor/acceptor molecules, solvents and annealed at a few temperatures before the materials are disregarded. Thus there exists a huge back catalogue of JV curves both in the literature and in the industry for material that was never fully analyzed.

Our aim when writing this paper was to develop a method that can accurately extract charge carrier recombination time (τ) and mobility (μ) as a function of light intensity using the most simple, quickest and easy to perform set of experiments possible. We wanted a measurement technique that took seconds to apply, that anybody without expensive lasers/frequency domain equipment could use and enabled the feedback loop from device performance to material parameters to be efficiently closed for all in the community. We focused on the recombination time constant and charge carrier mobility because they can be used to identify if recombination or transport is the key bottleneck in device performance, which can in turn give hints as to how to tune the molecular packing and/or morphology. Furthermore, when combined in the $\mu \cdot \tau$ product they give a standard benchmark for material performance.[28–30]

Herein, we demonstrate that both charge carrier recombination time and mobility can be extracted from JV curves alone using a combination of machine learning (ML) models trained on physically accurate device models. We compare the values of recombination rate and charge carrier mobility extracted by our new method to values extracted by more traditional frequency domain/transient measurements from both spin-coated and evaporated cells. Thus we develop a high throughput tool that has the potential to close the feedback loop and accelerate device development.

2. Methods

2.1. Time Domain Measurements on Evaporated Devices

Two devices of layer structure Glass/ITO/nC_{60}/C_{60}/DCV-V-Fu-Ind-Fu-V:C_{60}/MoO_3/Ag were deposited by evaporation, in one device the substrate temperature was held at 50 °C during deposition of the active layer, while in the other device substrate

N. Majeed, S. Greedy
Faculty of Engineering
The University of Nottingham
University Park, Nottingham NG7 2RD, UK
I. Ramirez, S. M. Hosseini, C. Uhrich
Heliatek GmbH
Treidlerstraße 3, 01139 Dresden, Germany
N. Li
Department of Materials Science and Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg
91054 Erlangen, Germany
N. Li
State Key Laboratory of Luminescent Materials and Devices
Institute of Polymer Optoelectronic Materials and Devices
School of Materials Science and Engineering
South China University of Technology
Guangzhou 510640, China
C. J. Brabec
Department of Materials Science and Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg
91054 Erlangen, Germany
J. R. Durrant
SPECIFIC IKC
Department of Materials
University of Swansea
Bay Campus, Swansea SA1 8EN, UK

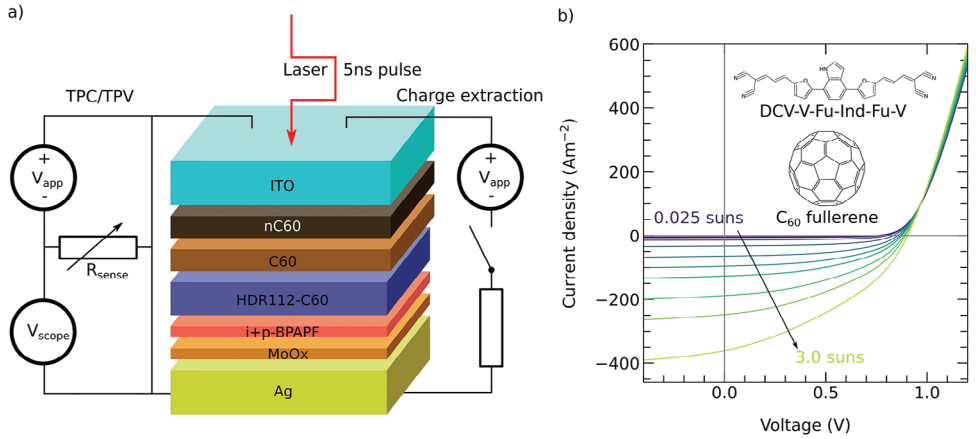

Figure 1. a) Device architecture and schematic depiction of transient techniques TPC/TPV and charge extraction, the nC60 layer is n-doped C60.; b) Measured JV-curves from 0.025 to 1 Suns for the device evaporated at room temperature. Inset: The molecular structures of DCV-V-Fu-Ind-Fu-V and C60.

temperature was allowed to float at room temperature.[31] The device structure is depicted in **Figure 1a** while the molecular structures and example JV curves can be seen in Figure 1b. The active layer was 50 nm thick and made by co-evaporating the small molecule donor DCV-V-Fu-Ind-Fu-V with C_{60}. TPV was performed at open circuit and charge extraction at short circuit to measure recombination times and effective charge carrier mobility respectively. A summary of these measurements can be seen in **Figure 2**.

Both JV curve and transient measurements were performed at light intensities ranging from 0.025 to 3 Suns. It could be seen that the charge carrier mobility measured at J_{sc} was a factor two higher for the 50 °C devices than for the room temperature device. This was attributed to slightly better transport properties caused by favorable morphology. Lifetimes at V_{oc} were almost identical for both devices, indeed it could be seen from the JV-curves in Figure 2d that V_{oc} was very close for both temperatures.

It is now anaim to see if using the JV curves alone (see Figure 1b) coupled with machine learning, this study can predict all the data extracted using transient measurements presented in Figure 2. JV curves were very quick and easy to measure. Thus if this study was able to extract μ and τ from these curves alone months of measurement work could be saved. To do this, the device structure was first set up in the drift-diffusion model OghmaNano.[27,32]

The model solves Poisson's equation to take account of electrostatic effects within the device, electron/hole charge carrier continuity and drift-diffusion equations to describe carrier transport. Finally to describe carrier trapping and recombination, the LUMO and HOMO Urbach tails were each split up into eight discrete trap levels and a Shockley-Read-Hall capture escape equation was solved for each energetic range. This approach allowed carries to be described both in energy and position space within the device. More details about the model can be found elsewhere.[33,27,34]

Using this base device structure, over 20,000 copies of the simulation file were made to form a sample set of virtual devices. Each virtual device had randomly assigned electron/hole mobilities, trap densities, Urbach tail slopes and other electrical paramters. From these virtual devices corresponding light and dark JV curves were generated. Furthermore, for each device the calculated recombination rate at V_{oc} and charge carrier mobility at J_{sc} were stored. This process is described in **Figure 3**.

Generating this data set took ≈2 h and provided the basis for training the machine learning algorithm. The advantage of training the machine learning algorithm on virtual data was that most machine learning algorithms were very data hungry requiring thousands of examples to learn. Furthermore, it enables us to know exactly what the recombination rate is at V_{oc} (mobility at J_{sc}) which would be hard to do experimentally.

The next task was to train the machine learning algorithm with the data. This is depicted in **Figure 4**. For each device in turn the light and dark JV curves were presented to the inputs of the neural network. The network was then asked to predict the values of charge carrier mobility and recombination rate as a function of light intensity on the outputs. At the start of training the model predicts these values quite poorly, however as training progresses and the network sees more examples, the predicted values of μ and τ for each JV curve become closer to the correct values (more details on the training can be found in the Supporting Information). Once the network had been trained on all devices, the order of the devices was shuffled and training begins again, this process repeats until the network can correctly predict μ and τ for any given JV curve in the data set. Once the error was sufficiently small, the weights were fixed and the model was ready to predict on experimental data. To test the ability of the network to extract μ and τ from as of yet unseen data, 20% of the training set was kept out of the training process, and used at the end of the training process to assess the performance of the network. Once the model was trained on virtual data to our satisfaction, the experimental JV curves for each device in Figure 1b were fed into the neural network in an attempt to predict the values in Figure 2.

The values of τ and μ predicted from the JV curves are shown in Figure 2 as solid triangles. It can be seen that the predicted values follow those of the directly measured values within one order of magnitude, accurately following the trend of the experimental data. This demonstrated that there was indeed enough information in the JV curves alone to determine τ and μ. As V_{oc}

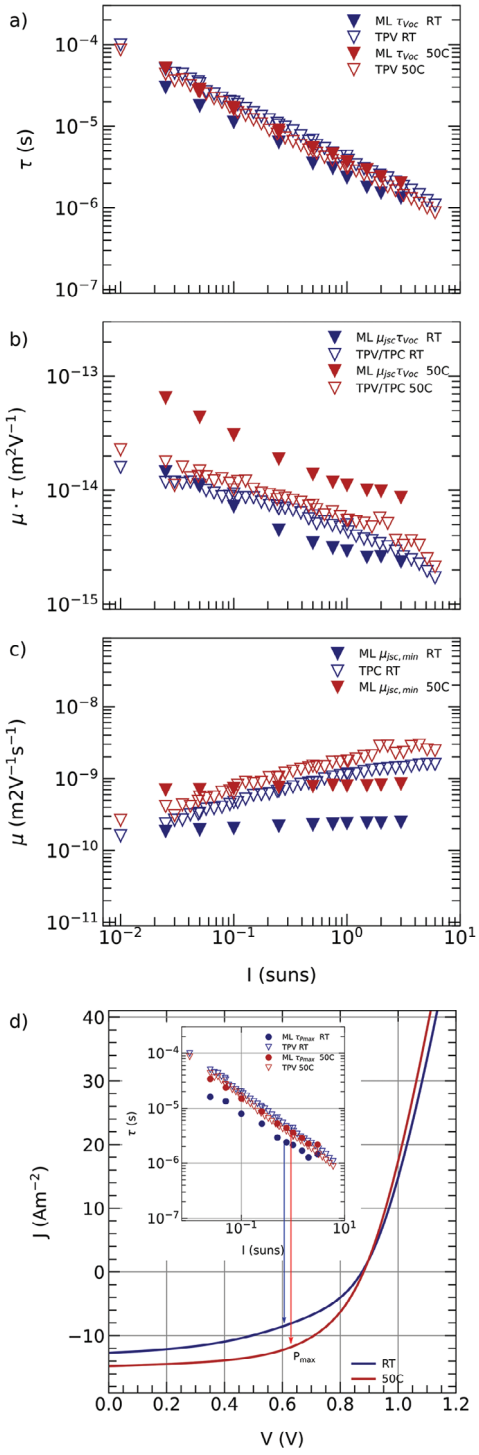

Figure 2. a) Light intensity dependent charge carrier lifetime measured using TPV for a device deposited at room temperature (blue)/50 °C (red); b) The $\mu_{jsc}\cdot\tau_{Voc}$ product c) Light intensity dependent charge carrier mobility measured using charge carrier extraction for a device deposited at room temperature (blue)/50 °C (red). In this figure the open triangles represent the experimental measurements and the solid triangles represent the results of the ML. d) JV-curves for devices deposited at room temperature (blue)/50 °C (red); inset shows charge carrier lifetime, closed circles show the predicted lifetime at maximum power point P_{max}. Confusion matrices and R2 scores can be found in Figures S3 and S4 (Supporting Information).

was almost the same for both devices, the information gained with TPV is limited in our case, however the machine learning model also enables us to also predict lifetimes away from Voc at the maximum power point P_{max}. The inset in 2b showed this prediction.

2.2. Frequency Domain Measurements on Spin Coated Devices

In the previous section, the ability of machine learning was compared to extract τ or μ from JV curves to the values τ or μ extracted from transient measurements. In this section, we demonstrate the general ability of the ML-approach by turning our attention to state-of-the-art PM6:DT-Y6 spin coated devices measured using frequency domain techniques. A series of glass/ITO/SnO_2/PM6:DT-Y6/MoO_3/Ag devices with varying DT-Y6 content were fabricated. The ratios chosen were 0:100, 15:85, 30:70, 45:55, 55:45, 70:30 and 85:15 of DT-Y6 to PM6 respectively. The molecular structure of these materials along with the device structure can be seen in **Figure** 5a,b. Current voltage curves were measured under AM1.5G illumination to obtain V_{oc}, PCE, J_{sc}, and FF, theseare plotted in Figure 5c as a function of blend ratio (See Supporting Information for full curves). It can be seen that as the DT-Y6 ratio increases so does the PCE with a maximum PCE observed at around 70:30.

To investigate the performance of these devices in terms of charge carrier transport and recombination rate IMVS was performed at open circuit and IMPS at short circuit to obtain charge carrier recombination rates and mobility as a function of light intensity. Example IMVS/IMPS curves are shown in **Figure** 6b,c from 300 μSuns to 1 Suns. The experimental charge carrier lifetimes were calculated from the real part of the IMVS-signal and charge carrier mobilities were inferred from the real part of the IMPS-signal. A summary of these measurements can be seen in the top two rows of **Figure** 7. Note all data extracted from experiment was plotted as open triangles, the closed triangles were the results of the machine learning and will be discussed later. If the $\mu\cdot\tau$ product was examined (bottom line of the figure) it could be seen that $\mu\cdot\tau$ was higher for the high performing DT-Y6 ratios (55:45, 70:30, and 85:15) mainly due to a higher effective charge carrier mobility.

The above results represent a base line against which to compare the machine learning. Before going further however, it is worth underlining some of the points made in the introduction about detailed characterization being the bottleneck to device development by noting that the above measurements took around 6 months to measure and analyze by hand.

Again the experimental JV curves for each device in Figure 6a were fed into the neural network in an attempt to predict the values in Figure 7. The predicted values were shown as solid triangles for mean values (geometric mean in case of charge carrier mobility), solid squares for electrons, and solid circles for holes. Taking the top row of graphs first, it can be seen that the model predicts electron mobility to be orders of magnitude higher than hole mobility. This was in accordance with literature.[35] Further the predicted geometric mobility was in good agreement with the experimental IMPS data. Examining the second line of graphs it could be seen that the Neural Network can predict the absolute value of the recombination time constant as a function of light

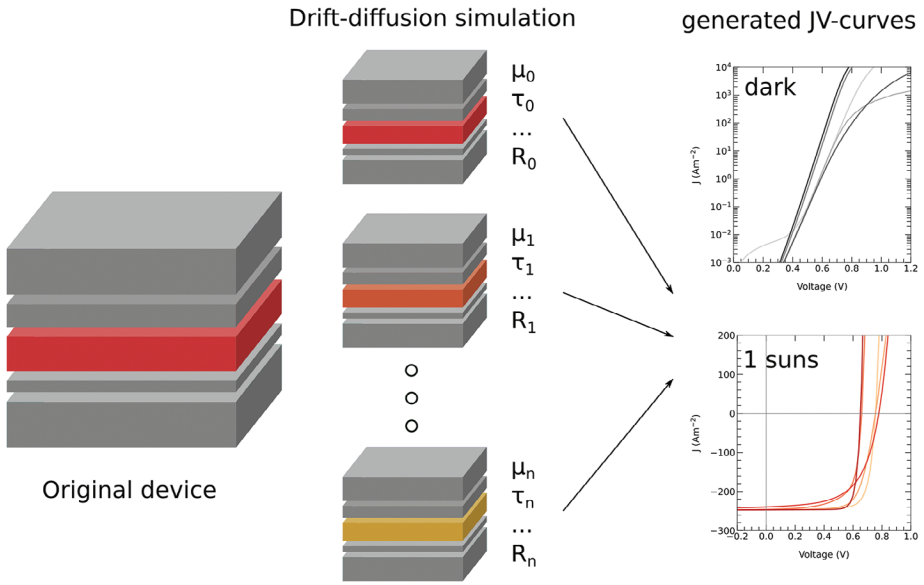

Figure 3. Creation of the training data set by artificially generating the device with randomly assigned parameters in a drift-diffusion simulation. The dark JV-curves and at 1 Sun as well as recombination rate at V_{oc} and mobility at J_{sc} are simulated and stored.

intensity very well with the error being slightly higher for the lower light intensities. Still the error stays well below one order of magnitude. Furthermore, the trend of the lifetime was also accurately reproduced. The bottom row of graphs compared the predicted $\mu_{jsc} \cdot \tau_{Voc}$ product to the measured values with these trends also agreeing well.

3. Discussion

Above we have demonstrated that using a combination of ML algorithms trained on simulated JV curves alone, one can build a tool to extract charge carrier mobility and recombination rate as a function of light intensity, thus removing the need for

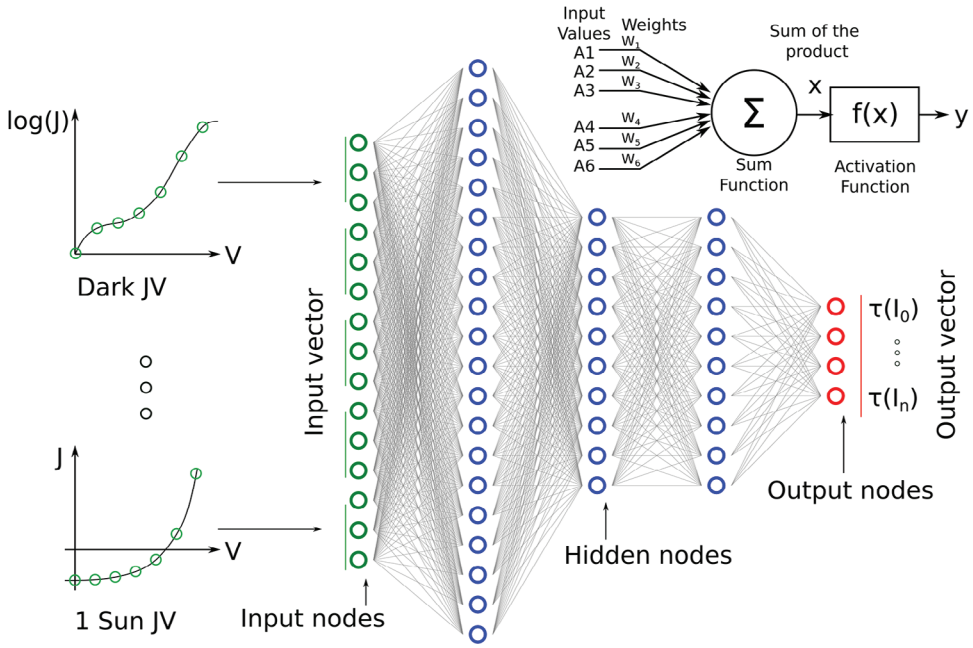

Figure 4. A diagram of the neural network used to extract material parameters from the data within this paper. Visible on the left-hand side of the image is the experimental (or simulated) data, with the green dots on the curves representing the points at which the curves were sampled to form input vectors for the neural network. The JV-curves are being sampled at discrete voltages to provide data points to the neural networks input nodes. Any number or combination of experimental measurements can be placed on the input to the network, one simply has to extend the number of input neurons, and retrain the network. The neural network itself has green input nodes, blue hidden layers, and red output nodes. Each output node corresponds to a device/material parameter such as charge carrier mobility or recombination rate. Inset: A single neuron.

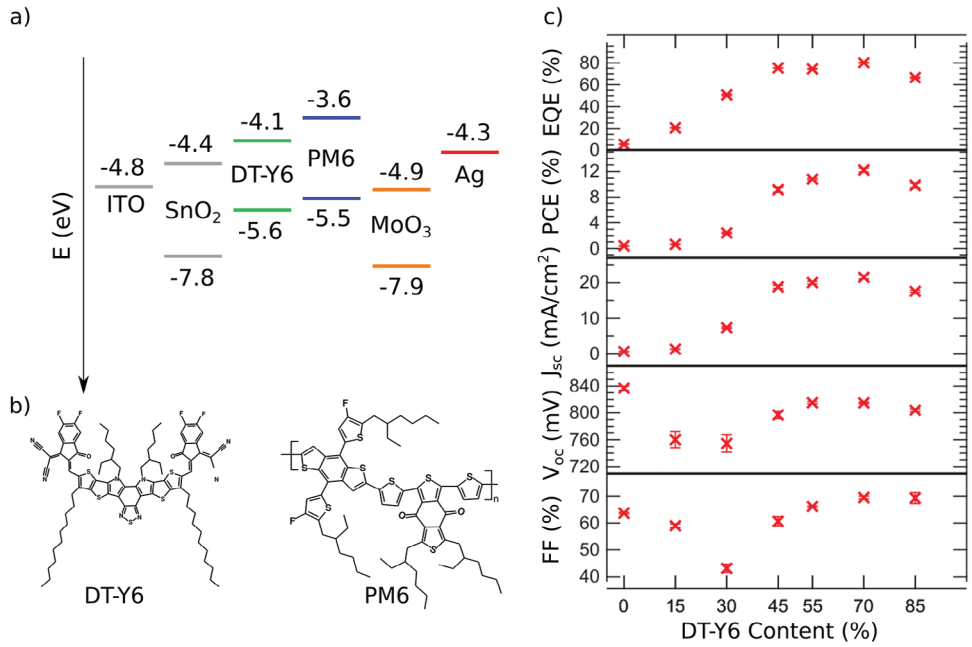

Figure 5. a) Device structure; b) Polymers of the active layer c) Device parameters depending on DT-Y6 content.

time-consuming and costly characterization. We anticipate this tool being used by the community to quickly screen new devices and materials and also as a tool to screen the vast historical data sets available in the literature and in industry. The method can also be thought of as a tool to democratize the characterization of OPV devices. Currently only well-funded labs can perform mobility and life-time measurements as they require relatively expensive lasers. This tool will allow more people to start extracting this data.

In some ways it is remarkable that using a simple drift-diffusion model and a machine learning algorithm we are able to extract carrier recombination time and charge carrier mobility as a function of light intensity. One would have thought that some type of transient measurement would be needed to extract this information. However, this preconception comes from a human centric view of solar cell measurements, in that one thinks measurements such as TPC and SCLC are needed to measure charge carrier mobility because that is what has been done in the past. However, we should approach the problem from the perspective of Shannon entropy. Entropy in information theory[36] is a measure of how much information is in a signal. For example a photograph of a perfectly clear blue sky contains low entropy (embodied information) as it simply tells you it is a sunny day. However, a picture of a clouded sky has higher entropy (embodied information), as it can tell you how high the clouds are, what type of clouds there are, likelihood of rain and likelihood of thunder. We should therefore think of electrical/optical measurements in the same context and ask how much embodied information does the measurement signal contain? In this case it is clear JV curves do encode information about τ and μ that the Neural Network can find and decode.

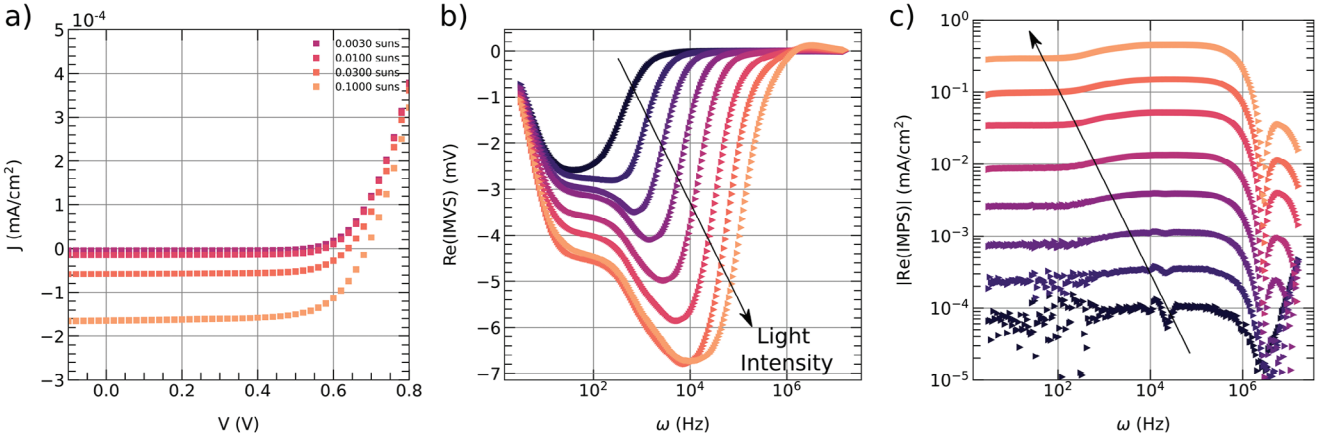

Figure 6. a) JV-curves for selected intensities; b) Intensity dependent experimental IMVS; and c) IMPS data for the 45% DT-Y6 device.

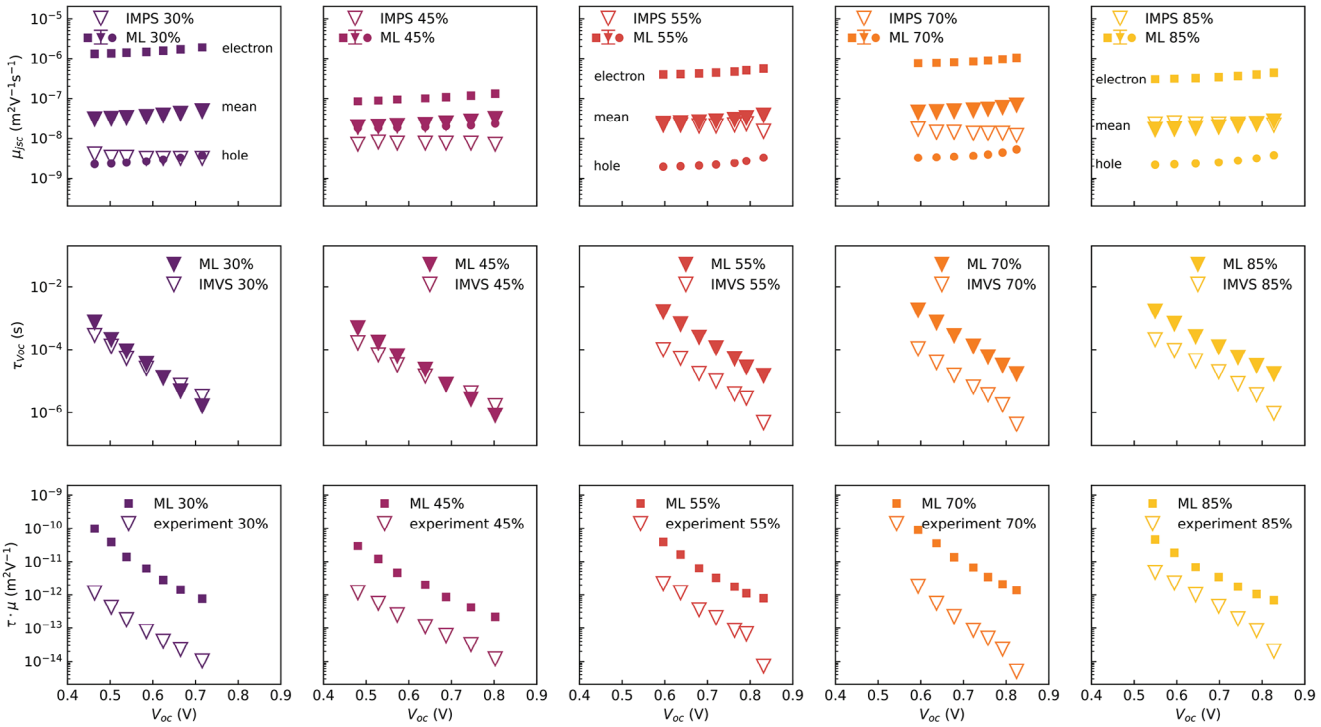

Figure 7. Top row) Open triangles represent charge carrier mobility as a function of light intensity measured at J_{sc} using IMPS for varying DT-Y6 content. Middle row) Open triangles represent recombination time constants measured at V_{oc} using IMVS plotted as a function of light intensity for varying DT-Y6 content. Bottom row) Open triangles represent calculated $\mu \cdot \tau$ products from the above two rows. Closed markers represent predicted values extracted from JV curves alone using machine learning. Closed squares relate to electrons, circles to holes and triangles to the (geometric) mean. Mean mobility was calculated using Equation (1). A discussion on differentiating between electron and hole mobilities can be found in the Supporting Information. Confusion matrices and R2 scores can be found in Figures S5 and S6 (Supporting Information).

Continuing this line of reasoning, there is no reason why we should focus our efforts on decoding JV curves or other standard measurements such as TPC alone. There may be another, as of yet unknown, measurement that may be as easy to obtain as a JV curve but contain more information that a machine learning algorithm can extract. In other words, an experiment designed for machine learning extraction rather than for human extraction. Indeed, it may be that the machine has to design its own perfect experiment to extract maximum possible information from a solar cell.

Now we comment on accuracy, although we demonstrated above that our method is accurate for the devices we chose. It should also be noted that it does not need to be completely accurate for all unusual classes of devices to be successful. Our method just needs to be good enough to show trends between devices and also flag up promising materials which are unusual. This first sift can then be used to flag devices to be investigated with more traditional experimental methods.

A general comment should be made about the measurement of τ and μ. It should be noted that the fundamentally difficult thing about measuring τ and μ in organic devices is that they are both a very strong function of carrier density due to the large number of trap states in the materials. Thus if applied voltage, photon flux, or contact materials are changed τ and μ will change. Therefore it is well known that different experiments that subject a device to different experimental conditions will produce different values of mobility/lifetime. For example both Charge Extraction by Lineally Increasing Voltage (CELIV) and TPC are commonly used to measure charge carrier mobility. In CELIV the device is held at under constant illumination and a negative voltage ramp is applied to study charge carrier mobility while in TPC the device is usually held at J_{sc} and the response of the device to a laser pulse is used to calculate mobility. Generally, such measurements will produce values of mobility within an order of magnitude to each other with trends that agree but will not be identical. Thus it should be noted that when we compare our simulated values to the experimental values we are not comparing identical quantities (as it always is the case in organics). Our simulated values of τ and μ are defined as:

$$\mu_{eff} = \frac{1}{d} \int_0^d \frac{\mu_{free} n_{free}(x)}{n_{free}(x) + n_{trap}(x)} dx \quad (1)$$

where μ_{free} is the charge carrier mobility of completely free carriers, n_{free} is the density of completely free carriers and n_{trap} is the density of trapped carriers. The effective mobility is calculated for each charge carrier specimen separately and an average mobility is calculated by taking the geometric mean:

$$\mu = \sqrt{\mu_e \cdot \mu_h} \quad (2)$$

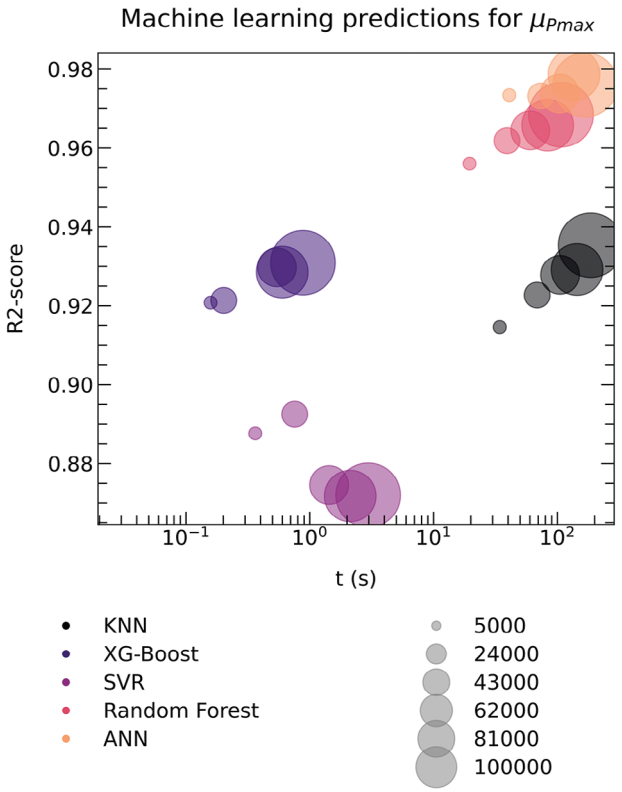

Figure 8. Comparison of accuracy and time taken to train Neural Networks, k-nearest neighbour regression (KNN), random-forest regression, extreme-boosted-gradient-descent regression (XG-Boost) and support-vector regression (SVR) on the SN21 data set. It can be seen the Neural Network performs best but is slowest to train.

The lifetime τ is calculated by:

$$\tau = \frac{\left(n_{total} - n_0\right)\left(p_{total} - p_0\right)}{R} \quad (3)$$

with n_{total}/p_{total} being the total charge carrier density in the device, n_0/p_0 the equilibrium free charge carrier density and R the total recombination rate.

Thus some of the error in the graphs may be down to slightly different definitions of mobility and time constant. Further it has been shown that charge carrier mobility results for the same device vary up to one order of magnitude when using different measurement techniques and up to a factor of three when different scientists analyze an identical dataset.[37] Difference between the ML predictions and experimental measurements is within the expected experimental error.

Finally, in the above examples we used Neural Networks for the machine learning, this is because we found their performance to be more accurate than other more traditional methods. Neural Networks do however require a lot of data and are also relatively slow to train. For comparison **Figure 8** plots the machine learning results from four other methods these include, k-nearest neighbour regression (KNN),[38] random-forest regression,[39] extreme-boosted-gradient-descent regression (XG-Boost)[40] and support-vector regression (SVR).[41] The figure plots *R2* score (accuracy) v.s. time is taken to train for the data set generated for the PM6:DT-Y6 device. The size of the bubble represents the size of the training data set. Data sets of between 5000 and 100 000 devices were used. It can be seen that the XG-Boost algorithm is the fastest but also the worst, SVRs and KNNs have the same level of performance while KNN is slower. The best performing method is the Neural Network, closely followed by the random forest. Each of these algorithms can be optimised, for example the number and size of layers in the Neural Network can be tuned to obtain best performance. However, these results represent our best efforts.

4. Predicting on Databases

The real strength of the machine learning approach is revealed when large sets of data have to be analysed, as it enables material parameters to be extracted that have not directly been measured. Indeed, the devices may have been made and discarded years ago. As a demonstration of our method the ML algorithm was used to predict mobility and trap state density from a set of over 10,000 historical JV curves held by Heliatek GmbH, the results can be seen in **Figure 9**. The original database only contained JV-curves at dark conditions and at 1 *Suns* light intensity. It can be seen that the model identifies a clear correlation between V_{oc} and charge carrier mobility, as well as a clear correlation between PCE and trap density.

This technique would allow one to data mine these historical data sets and identify devices with optimal charge carrier transport properties that were potentially overlooked in the past.

5. Conclusion

Above we demonstrated that one does not need complex time domain/frequency domain measurement techniques to access charge carrier mobilities and recombination time constants. This information is encoded within the far more simple to obtain current–voltage curves. One simply needs a relatively low-cost computer to extract this information. Furthermore, once trained the machine learning models take a fraction of a second to apply which means devices can be analyzed as they are produced. This is important in the academic setting but more important in an industrial setting where tens of devices are produced per day. Furthermore, this approach will allow researchers to scour historical materials for promising candidates that we have skipped over as a community. Finally, we emphasise that experimental data should be seen from an information theory point of view. Maximising entropy by conducting the right combination of experiments will be key to optimise the use of machine learning.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank Heliatek GmbH for funding MH's PhD through the EPSRC Centre for Doctoral Training in Renewable Energy Northeast Universities (ReNU). The authors also thank the Deutsche Forschungsgemeinschaft (DFG Research Unit FOR 5387 POPULAR, Project No. 461909888)

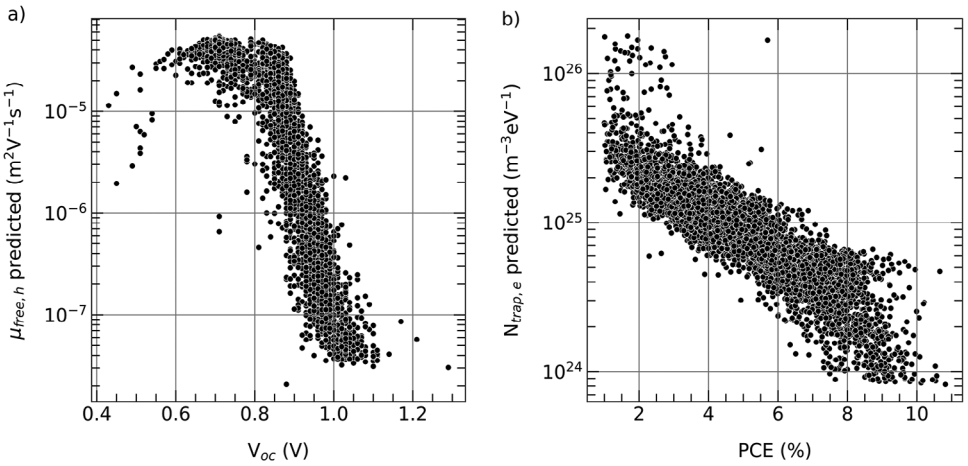

Figure 9. Predicted device parameters of a database containing around10 000 devices. The predictions are plotted over the experimentally determined V_{oc}/PCE. a) Experimental V_{oc} v.s. free hole mobility; b) Experimental PCE v.s. trap state density for electrons. Confusion matrices and R2 scores can be found in Figures S1 and S2 (Supporting Information). Note the value of mobility represents only free carriers, it is the same free carrier mobility as in Equation (1).

for their support. R.A.P. would further like to acknowledge the support through the Centre for Processable Electronics CDT program as well as the ATIP project (grant number EP/L016702/1 and EP/T028513/1). This work was supported by the Engineering and Physical Sciences Research Council (grant number EP/S023836/1).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Where possible data will be provided by the authors upon reasonable request. However, some data is commercially sensitive so can not be provided.

Keywords

drift diffusion, machine learning, organic photovoltaic, solar

Received: September 7, 2023
Revised: October 12, 2023
Published online: December 3, 2023

[1] S. E. Shaheen, C. J. Brabec, N. S. Sariciftci, F. Padinger, T. Fromherz, J. C. Hummelen, *Appl. Phys. Lett.* **2001**, *78*, 841.
[2] L. Zhu, M. Zhang, J. Xu, C. Li, J. Yan, G. Zhou, W. Zhong, T. Hao, J. Song, X. Xue, Z. Zhou, R. Zeng, H. Zhu, C.-C. Chen, R. C. I. Mackenzie, Y. Zou, J. Nelson, Y. Zhang, Y. Sun, F. Liu, *Nat. Mater.* **2022**, *21*, 656.
[3] G. Li, R. Zhu, Y. Yang, *Nat. Photonics* **2012**, *6*,153.
[4] Y. Liu, B. Liu, C.-Q. Ma, F. Huang, G. Feng, H. Chen, J. Hou, L. Yan, Q. Wei, Q. Luo, Q. Bao, W. Ma, W. Liu, W. Li, X. Wan, X. Hu, Y. Han, Y. Li, Y. Zhou, Y. Zou, Y. Chen, Y. Li, Y. Chen, Z. Tang, Z. Hu, Z.-G. Zhang, Z. Bo, *Sci. China: Chem.* **2022**, *65*, 224.
[5] G. Dennler, N. S. Sariciftci, *Proc. IEEE* **2005**, *93*, 1429.
[6] A. Dhanabalan, J. K. J. Van Duren, P. A. Van Hal, J. L. J. Van Dongen, R. A. J. Janssen, *Adv. Funct. Mater.* **2001**, *11*, 255.
[7] T. Liu, A. Troisi, *Adv. Mater.* **2013**, *25*, 1038.
[8] J. Zhang, H. S. Tan, X. Guo, A. Facchetti, H. Yan, *Nat. Energy* **2018**, *3*, 720.
[9] J. Hou, O. Inganäs, R. H. Friend, F. Gao, *Nat. Mater.* **2018**, *17*, 119.
[10] M. Riede, D. Spoltore, K. Leo, *Adv. Energy Mater.* **2021**, *11*, 2002653.
[11] L. Duan, A. Uddin, *Adv. Sci.* **2020**, *7*, 1903259.
[12] Intergovernmental Panel On Climate Change (Ipcc), *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.*, **2023**.
[13] L. Yang, L. Yan, W. You, *J. Phys. Chem. Lett.* **2013**, *4*, 1802.
[14] K. Kranthiraja, A. Saeki, *Adv. Funct. Mater.* **2021**, *31*, 2011168.
[15] J. R. Macdonald, *Ann. Biomed. Eng.* **1992**, *20*, 289.
[16] F. Fabregat-Santiago, G. Garcia-Belmonte, I. Mora-Seró, J. Bisquert, *Phys. Chem. Chem. Phys.* **2011**, *13*, 9083.
[17] P. M. DiCarmine, O. A. Semenikhin, *Electrochim. Acta* **2008**, *53*, 3744.
[18] Y. T. Set, B. Li, F. J. Lim, E. Birgersson, J. Luther, *Appl. Phys. Lett.* **2015**, *107*, 173301.
[19] J. C. Byers, S. Ballantyne, K. Rodionov, A. Mann, O. A. Semenikhin, *ACS Appl. Mater. Interfaces* **2011**, *3*, 392.
[20] R. A. Street, *Phys. Rev. B* **2011**, *84*, 075208.
[21] J. Vollbrecht, N. Tokmoldin, B. Sun, V. V. Brus, S. Shoaee, D. Neher, *J. Appl. Phys.* **2022**, *131*, 221101.
[22] A. Foertig, A. Wagenpfahl, T. Gerbich, D. Cheyns, V. Dyakonov, C. Deibel, *Adv. Energy Mater.* **2012**, *2*, 1483.
[23] J. Bisquert, M. Janssen, *J. Phys. Chem. Lett.* **2021**, *12*, 7964.
[24] K. Nakano, Y. Chen, K. Tajima, *AIP Adv.* **2019**, *9*, 125205.
[25] W. Tress, S. Corvers, K. Leo, M. Riede, *Adv. Energy Mater.* **2013**, *3*, 873.
[26] R. Hanfland, M. A. Fischer, W. Brütting, U. Würfel, R. C. I. Mackenzie, *Appl. Phys. Lett.* **2013**, *103*, 063904.
[27] R. C. I. Mackenzie, C. G. Shuttle, M. L. Chabinyc, J. Nelson, *Adv. Energy Mater.* **2012**, *2*, 662.
[28] J. R. Tumbleston, Y. Liu, E. T. Samulski, R. Lopez, *Adv. Energy Mater.* **2012**, *2*, 477.
[29] R. A. Street, A. Krakaris, S. R. Cowan, *Adv. Funct. Mater.* **2012**, *22*, 4608.
[30] T. Kirchartz, T. Agostinelli, M. Campoy-Quiles, W. Gong, J. Nelson, *J. Phys. Chem. Lett.* **2012**, *3*, 3470.

[31] M. Saladina, C. Wöpke, C. Göhler, I. Ramirez, O. Gerdes, C. Liu, N. Li, T. Heumüller, C. J. Brabec, K. Walzer, M. Pfeiffer, C. Deibel, *Phys. Rev. Lett.* **2023**, *130*, 236403.
[32] R. C. I. MacKenzie, "OghmaNano.", https://www.oghma-nano.com/. (accessed: September **2010**).
[33] R. C. I. Mackenzie, T. Kirchartz, G. F. A. Dibb, J. Nelson, *J. Phys. Chem. C* **2011**, *115*, 9806.
[34] J. A. Röhr, R. C. I. Mackenzie, *J. Appl. Phys.* **2020**, *128*, 165701.
[35] N. Tokmoldin, S. M. Hosseini, M. Raoufi, L. Q. Phuong, O. J. Sandberg, H. Guan, Y. Zou, D. Neher, S. Shoaee, *J. Mater. Chem. A* **2020**, *8*, 7854.
[36] C. E. Shannon, *The Bell Syst. Tech. J.* **1948**, *27*, 379.
[37] J. C. Blakesley, F. A. Castro, W. Kylberg, G. F. A. Dibb, C. Arantes, R. Valaski, M. Cremona, J. S. Kim, J.-S. Kim, *Org. Electron.* **2014**, *15*, 1263.
[38] T. Cover, P. Hart, *IEEE Trans. Inform. Theory* **1967**, *13*, 21.
[39] L. Breiman, *Mach. Learning* **2001**, *45*, 5.
[40] T. Chen, C. Guestrin, *KDD '16: Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, USA **2016**, 785.
[41] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, V. Vapnik, *Advances in Neural Information Processing Systems*, MIT Press, Massachusetts, USA **1996**, *9*.