

UC Berkeley

UC Berkeley Previously Published Works

Title

Rapid Photovoltaic Device Characterization through Bayesian Parameter Estimation

Permalink

<https://escholarship.org/uc/item/0vd5t4bx>

Journal

Joule, 1(4)

ISSN

2542-4785

Authors

Brandt, Riley E
Kurchin, Rachel C
Steinmann, Vera
et al.

Publication Date

2017-12-01

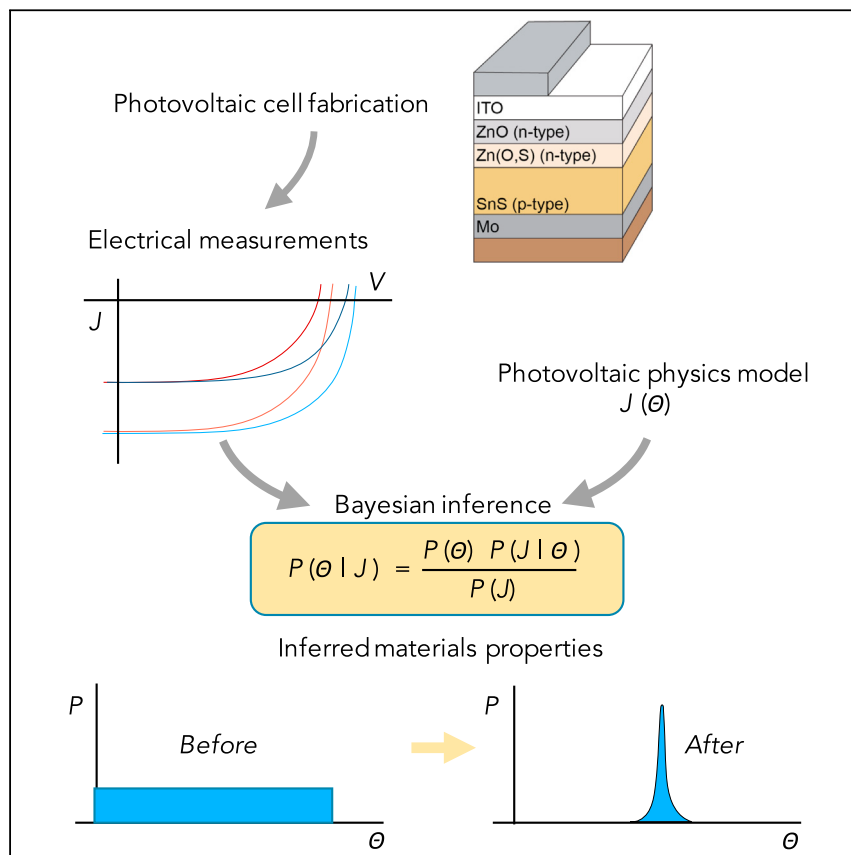
DOI

10.1016/j.joule.2017.10.001

Peer reviewed

Article

Rapid Photovoltaic Device Characterization through Bayesian Parameter Estimation



High-performance computing can greatly improve the workflow of experimentalists in energy materials, through the use of Bayesian inference. This allows us to solve the inverse problem of extracting underlying materials properties through the measurement of the electrical behavior of completed devices. Cheaper, faster measurements can be substituted for longer direct measurements of individual properties, without sacrificing accuracy or precision. We provide a general framework to apply this to other materials systems and devices.

Riley E. Brandt, Rachel C. Kurchin, Vera Steinmann, ..., Gerbrand Ceder, Thomas Unold, Tonio Buonassisi

rbrandt@alum.mit.edu

HIGHLIGHTS

Bayesian inference is applied to simple photovoltaic electrical measurements

This inference is able to uncover underlying materials properties in a PV cell

Properties are inferred significantly faster than they would be measured separately

This fast inference is enabled by high-performance computing

Article

Rapid Photovoltaic Device Characterization through Bayesian Parameter Estimation

Riley E. Brandt,^{1,5,*} Rachel C. Kurchin,¹ Vera Steinmann,¹ Daniil Kitchaev,¹ Chris Roat,² Sergiu Levenco,³ Gerbrand Ceder,^{1,4} Thomas Unold,³ and Tonio Buonassisi¹

SUMMARY

In photovoltaic (PV) materials development, the complex relationship between device performance and underlying materials parameters obfuscates experimental feedback from current-voltage (J-V) characteristics alone. Here, we address this complexity by adding temperature and injection dependence and applying a Bayesian inference approach to extract multiple device-relevant materials parameters simultaneously. Our approach is an order of magnitude faster than the cumulative time of multiple individual spectroscopy techniques, with added advantages of using device-relevant materials stacks and interface conditions. We posit that this approach could be broadly applied to other semiconductor- and energy-device problems of similar complexity, accelerating the pace of experimental research.

INTRODUCTION

Historically, new photovoltaic (PV) materials have taken several years (halide perovskites) or decades (silicon or cadmium telluride) to advance from single-digit efficiencies to industrially relevant values in excess of 10%.^{1,2} This slow progress is typical of novel materials development in many other fields, where the average time from discovery to commercial success is approximately a decade.³ These long development times are in direct conflict with more urgent, competing timelines: the rapid pace of climate change, the short investment time horizon of venture capital, and the increasingly high speed of computational materials discovery. With the long materials development timeline as the weakest link, it is clear that experimentalists must take steps to improve the rate at which we discover and develop promising candidate materials.

There are several reasons for this protracted development timeline. The first is that experimental cycles of fabrication and characterization of devices are often slow and resource intensive, which is illustrated well by PVs. These difficulties are in large part due to the complex relationships between observable PV device performance and the underlying materials parameters. Experimental observations of PV device behavior may be consistent with many underlying causes, which results in uncertainty, necessitates extensive experimentation, and increases the importance of scientists' intuition and experience. At the same time, the market pull for new materials is very weak until they have demonstrated promising performance, which means resources are limited for early-stage development.⁴ These limited resources compound the difficulty of executing a large number of experimental cycles and training enough experts to guide experiments. The result is a long development period for most semiconductors and poor return on investment.⁵

Context & Scale

Photovoltaic (PV) research has historically taken decades to bring new materials to market, as the pace of development is often limited by our ability to identify the causes of underperformance. There are many ways to make an underperforming PV cell, and their signatures on device efficiency alone are not unique. This uncertainty plagues other energy-storage or -conversion devices with similarly complex combinations of materials but may be used to our advantage. By probing or biasing a PV cell in different operating conditions, the signature of underlying material properties becomes more unique. This connection between the operating conditions, the material properties, and the PV cell output may be solved through Bayesian inference algorithms. While computationally expensive, high-performance computing can enable such inference as a tool for experimentalists, a tool that could become increasingly valuable for accelerating the pace of materials research in PV and related fields.

In the present work, we demonstrate how a Bayesian parameter estimation (BPE) or inference approach can be used to address this challenge directly, by both decreasing the cost of experiments and more accurately identifying root causes of PV efficiency losses. We apply existing BPE frameworks^{6–8} to pinpoint underperforming device-relevant materials parameters in thin-film solar-cell materials on a timescale of <24 hr in device-relevant geometries and fabrication conditions. This approach of imbuing the experimental intuition into an inference algorithm can help dramatically shorten a typical experiment cycle and free scientists to develop PV materials more quickly.

We discuss the theoretical framework for BPE to solve this problem over a discretized parameter space. We validate the approach with a numerical device model of gallium arsenide (GaAs), a well-known PV material. We then apply the approach to an earlier-stage PV material, tin monosulfide (SnS), to fit materials and interface properties, including surface recombination velocity (S_{eff}), mobility (μ), minority carrier lifetime (τ), and conduction band offset (ΔE_C), using automated cell electrical measurements. Measurement of these properties would typically require fabrication of multiple separate samples and many person-hours of effort on specialized equipment, as illustrated in Figure 1B. We validate the approach through parallel spectroscopic measurements of each property and demonstrate comparable accuracy and precision with an order of magnitude faster approach. We note that several spectroscopic techniques have recently been developed to extract more *in situ* or *in operando* materials properties,^{9–11} but a BPE approach may be considered complementary to these techniques and can be applied to any measurements of device performance.

To extract device-relevant materials parameters from experimental results, we simulate and analyze large experimental datasets using high-performance computing (HPC). Moore's Law, and the resulting rise of HPC, has revolutionized the scope and accuracy of predictions that theorists can make about the natural world through computational simulation. However, despite the exponentially decreasing costs of computation (as well as increasing costs to conduct increasingly sophisticated experiments), experimental scientists have been slower to embrace HPC. This BPE framework applied to empirical materials development offers an opportunity for experimentalists to leverage the growing benefits of HPC.¹² It also offers a route for the tremendous advances in machine learning and inference to be applied to a practical problem in a field that has not yet fully embraced these breakthroughs.¹³

Given the advantages of speed, accuracy, and device-relevant form factor, a Bayesian inference approach such as the one presented here is broadly applicable to other complex optimization problems in semiconductor and energy-conversion materials and may help to accelerate the development of these materials as well.

Photovoltaic Device Fabrication and Characterization

A PV device consists of a stack of thin-film layers, each of which serves an important purpose in the process of using photons to excite, separate, and collect electrical charge that can do useful work. Several dozen materials properties in each layer, as well as the interfaces between them, determine the performance of the overall device. This interplay is illustrated by a typical early-stage PV material, tin monosulfide (SnS). Over a decade of research has led to PV conversion efficiencies in excess of 4%,¹⁴ due to advances in SnS fabrication techniques,^{15,16} control of SnS doping

¹Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Google, Inc., Mountain View, CA 94043, USA

³Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin 14109, Germany

⁴Department of Materials Science and Engineering, University of California Berkeley, Berkeley, CA 94720, USA

⁵Lead Contact

*Correspondence: rbrandt@alum.mit.edu

<https://doi.org/10.1016/j.joule.2017.10.001>

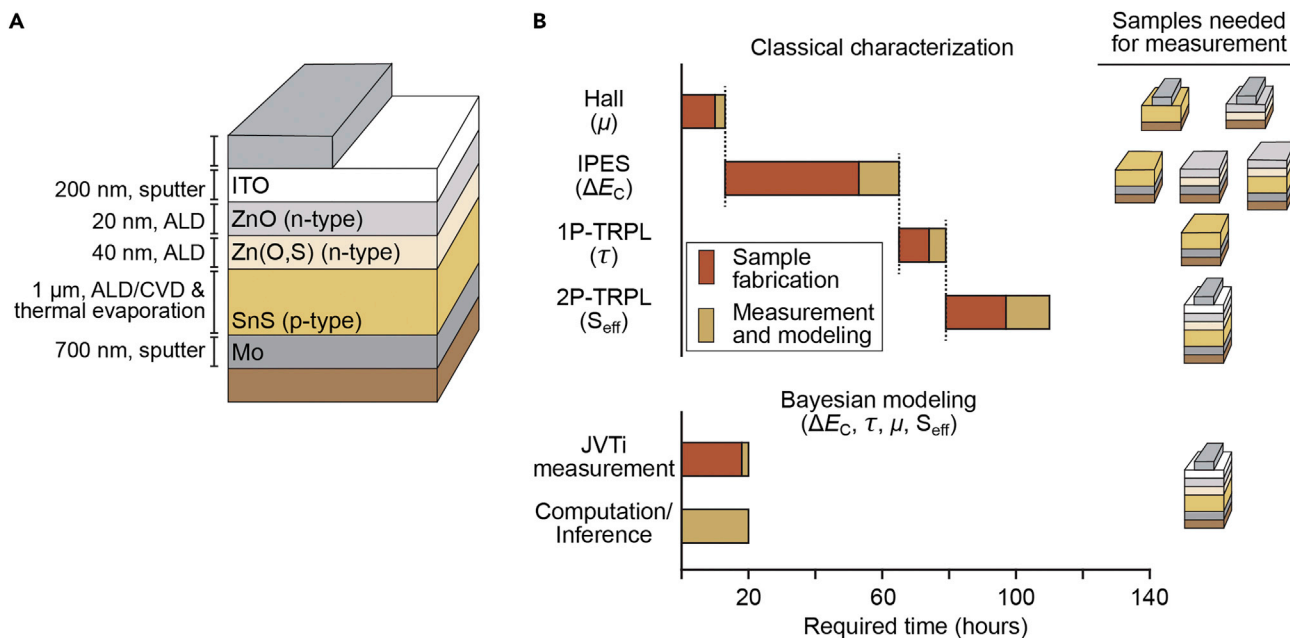


Figure 1. SnS PV Device and Sample Construction

(A) PV device stack for SnS, showing the composition of each layer and their relative thicknesses. (B) Flowchart of different workflows comparing (top) the sample layer stack(s) and time required to extract a variety of different materials properties, and (bottom) the device stack and hours required in a BPE approach. Times are specific to the tool setup herein and vary depending upon the tools and materials.

and transport properties,^{17,18} control of interfaces,^{19,20} optimization of contact materials,^{19,21} and a tremendous investment in fundamental understanding of other related PV materials.

Figure 1A shows the device stack of a state-of-the-art SnS heterojunction, thin-film solar cell measured in this study, including materials, thicknesses, and deposition techniques for each layer. First, a molybdenum back contact is sputtered onto a Si/SiO₂ substrate. Next, the active semiconducting absorber layer is formed through thermal evaporation of SnS powder (Sigma-Aldrich, >99.99% purity), which is subsequently annealed in an H₂S atmosphere to improve the film morphology.¹⁵ The heterojunction is formed by the deposition of n-type Zn(O,S) and ZnO layers through atomic layer deposition.^{14,19} Indium tin oxide, a transparent conductor, is sputtered to form the front contact, and finally, silver grid fingers are deposited via evaporative deposition. Further details on each step are found in prior work.^{15–17} This cell architecture holds record efficiencies for both thermally evaporated and chemical vapor-deposited SnS cells.^{14,16}

With six active layers in this device stack, and five corresponding interfaces between them, there are dozens of potential materials and interface properties that may influence electrical performance. In an exhaustive effort to identify which of these properties may be the limiting factor, many of these properties must be independently measured and modeled.^{14,18,19,21}

Typically, characterization of each property involves a study of a sample of one or a few of these layers in isolation. This requires that separate samples be fabricated for each measurement, taking additional time (see Figure 1B), and as these materials are not measured in the context of the device in which they will operate,

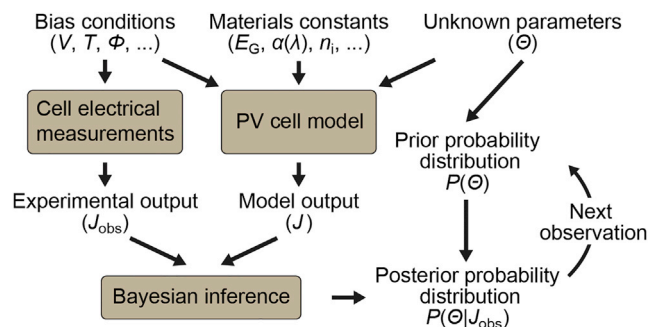


Figure 2. Flowchart of the Bayesian Parameter Estimation Approach

Combining measurements and a forward model of the PV device produces a posterior probability distribution of parameters. The posterior distribution can be iteratively refined with additional observed data by serving as the new prior distribution in subsequent cycles of the method. Bias conditions include voltage (V), temperature (T), and illumination intensity (Φ), while materials constants include properties such as bandgaps (E_G), absorption coefficients (α), and intrinsic carrier concentrations (n_i). Unknown parameters (which will be fit) are contained in Θ . Current density is the output of both experiment (J_{obs}) and the PV cell model (J).

the measurements may not be truly representative of how they will perform after complete processing and during operation. In the example shown in Figure 1B, four device-relevant materials parameters are extracted from four discrete measurements: inverse photoemission spectroscopy determines conduction band offset (ΔE_C) with three or more custom test structures; single-photon time-resolved photoluminescence (1P-TRPL) determines minority carrier lifetime (τ); Hall-effect measurements estimate majority carrier mobility (μ_{maj}) with two custom test structures; and two-photon TRPL (2P-TRPL) determines surface recombination velocity (S_{eff}). The test structures are not always representative of realistic device conditions; for example, a thin film grown on an electrically insulating substrate to perform Hall measurements will have a different grain structure, and hence electrical properties, than a thin film grown on a conductive metal back contact in a PV device.²² These Hall measurements assess majority carrier mobility and are commonly dominated by grain boundaries as carriers travel laterally through the film, while solar cells are limited by minority carriers, traveling out of the film.

In contrast to this complexity, a Bayesian inference approach attempts to simultaneously extract multiple underlying materials and interface properties in a single device measurement, which allows for a significantly faster experimental feedback from more representative device-relevant conditions. To solve such an inverse problem requires construction of an appropriate model of the device, development of a Bayesian framework for applying it, and access to HPC resources to perform the simulations and inference calculations.

Device Model and Bayes Theorem

As shown in Figure 2, BPE requires three steps: cell electrical measurements, an optoelectronic model connecting materials properties to cell behavior, and a Bayesian inference algorithm.

First, cell electrical measurements require the ability to probe a solar-cell output current over a range of bias conditions, including voltage (V), temperature (T), and illumination (Φ). The measurement suite for performing these measurements is often

referred to as a *JVTi* measurement, and has been described in prior work.^{23–26} Our specific experimental setup is further described in Chakraborty et al.²⁷

Next, the PV cell optoelectronic model is accomplished by finite-element solutions to a set of partial differential equations, which describe the transport and recombination of electrons and holes in the semiconductor device. These equations include (1) transport of carriers under electrochemical potential gradients (Poisson and continuity equations); (2) generation of carriers by thermal and/or optical means; and (3) recombination of free carriers through radiative, Auger, and defect-assisted recombination events.

These partial differential equations (PDEs) are solved by knowing: (1) materials constants of each material in the cell stack, such as densities of states, bandgaps, and absorption coefficients; (2) the bias conditions such as voltage, temperature, and illumination intensity; and (3) cell-specific materials properties such as μ , N_A (acceptor density), and N_D (donor density), and the recombination rate constants. Of the several dozen input parameters into this model, 5–10 may be unknown and depend upon fabrication conditions. This feature vector of unknown parameters, Θ , is of interest for the experimentalist to fit in order to understand the device's performance.

The software SCAPS-1D is used here to solve this system of differential equations using a finite-difference approach,²⁸ as it is particularly accurate for thin-film, heterojunction solar cells. It is important to choose a model and PDE solver that captures the physics relevant to the problem at hand. The [Supplemental Information](#) includes code for controlling SCAPS through Python or MATLAB, as well as more information on the cell model used.

Finally, now that the forward problem is well defined, we may solve the “inverse problem” using either classical statistics algorithms or Bayesian inference. As the forward problem requires a numerical solution to a set of PDEs, there is no way to directly invert the problem as with an analytical model. For numerical solutions, classical parameter estimation/optimization algorithms generally assume that there is one true value for the underlying parameter, and to find it they seek the model parameter(s) that minimize a statistical estimator such as the root-mean-squared error. However, where many equivalently good fits exist, this approach is insufficient.

A discretized Bayesian framework is instead able to assign a posterior probability to every combination of parameters. This probability represents the belief that a given set of discrete parameter values is the true set of material properties and is represented by a multivariate probability distribution over all possible combinations of model fit parameters. This probability is modified every time one observes new data, based on belief in the new data and its uncertainty. In that way, a BPE routine answers the question: “what is the probability that a hypothesis is true, given some new evidence that has been discovered?”

A non-parametric, discretized BPE approach performs well when multiple sets of parameters provide an equivalently good fit to the data or when the solution space is a complex multivariate distribution of parameters. BPE can also continuously accommodate new observations, allowing beliefs to change with every subsequent measurement.²⁹ BPE can also handle different uncertainty in the observed data, with the resulting probability distributions, including the noise distribution of the data,

handling both random noise related to the precision of an instrument and systematic uncertainty arising from uncertain calibrations of tools.

In our BPE framework, every set of possible model parameters, θ , may be considered as a unique hypothesis, H . Each new piece of evidence may be referred to as E . Bayes' Rule, the underpinning of all Bayesian methods, states

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} \quad (\text{Equation 1})$$

$P(H)$ is the prior probability distribution, or how strongly one believes in each hypothesis in θ before doing any experiments. $P(H|E)$, our goal, is the posterior probability distribution that follows after each piece of evidence, or the updated probability distribution given the results of an experiment. $P(E|H)$ is the likelihood, specifically the likelihood that the observed evidence would have occurred if the parameters hypothesized were true (e.g., if the set of hypotheses would have produced a very different value from what was observed, the likelihood $P(E|H)$ would be small). Lastly, $P(E)$ is a normalizing constant; it is the sum of $P(H)P(E|H)$ over all the possible hypotheses. This normalization reflects the fact that certainty in any one hypothesis is not only related to the likelihood for that particular hypothesis it is dampened by the likelihoods of all competing hypotheses.

BPE is an iterative process, where Equation (1) is applied repeatedly after each observation. One may continue to make progressive observations and increasingly gain certainty by performing this inference step every time.

In the case of PV device parameter fitting, our evidence is the observation of an output current from our experimental setup. The hypotheses are a vector of discrete values of the underlying parameters (θ) that we would like to fit, for example, the minority carrier lifetime, τ . BPE poses the question: what is the credibility or probability distribution of the carrier lifetime, given that we observe a particular current value (J_{obs}) under certain bias conditions:

$$P(\theta|J_{\text{obs}}) = \frac{P(\theta)P(J_{\text{obs}}|\theta)}{P(J_{\text{obs}})}. \quad (\text{Equation 2})$$

These probabilities must typically be determined numerically, as closed-form analytical solutions for BPE only exist for particular probability distributions. The process for evaluating Equation 2 numerically requires determining prior knowledge about the problem, gathering data, and then computing the posterior probability distribution. HPC enables these calculations (and in particular the most expensive one, computing the likelihood) to be performed within hours, instead of days or months.

Specifying the prior probability distribution is a subjective step. The simplest assumption we make here is that the prior probability is uniformly distributed over a particular interval, e.g., the conduction band offset may be anywhere from -0.5 eV to $+0.5$ eV. This distribution assumes that values within the interval are equally likely, but values outside of it are impossible. These ranges are conservatively large bounds from the literature on the material in question. The uniform distribution over a limited set of values constrains the results but simplifies the discretized solution approach.

For N parameters in θ , and M discrete values for each parameter, the hypotheses may be represented as an $N \times M$ dimensional matrix. A uniform probability distribution over all M divisions in N dimensions results in a probability of $1/MN$ for each hypothesis.

Next, to compute the likelihood $P(J_{\text{obs}}|\Theta)$, one must know what current density to expect for every combination of discrete parameters in Θ and every discrete bias condition. This calculation requires evaluating the device electrical model at every hypothesis in Θ , and at every bias condition. The number of simulations required for M discrete divisions for all N model parameters scales with M^N , multiplied by the number of operating conditions. This scale highlights the central role that HPC plays in enabling BPE for materials science. For just two different temperatures and light intensities, with $M = 10$ discrete divisions over four dimensions ($N = 4$) in Θ , a grid BPE search would require 40,000 total simulations. Where this could take several days on a single computer, HPC offers the possibility of massively parallelizing the computation to make it tractable in timescales of only hours.

Computing the likelihood of observing an output current, given Θ , requires knowledge about the uncertainty of a measurement due to random noise or error in the measurement apparatus. We may assume that the measured current, J_{obs} , is distributed around the “true” value, J_{real} , with a normal distribution, with an SD of σ_J . Therefore, the likelihood of observing any current (J_{obs}) at a particular set of parameters Θ is simply its probability of occurring, or

$$P(J_{\text{obs}}|\Theta) = \frac{1}{\sigma_J \sqrt{2\pi}} \exp\left(-\frac{(J_{\text{obs}} - J(\Theta))^2}{2\sigma_J^2}\right). \quad (\text{Equation 3})$$

The uncertainty in current is determined through measuring uncertainties in T , V , and Φ , which propagate through as

$$\sigma_J = \sqrt{\left(\frac{\partial J}{\partial V}\right)^2 \sigma_V^2 + \left(\frac{\partial J}{\partial T}\right)^2 \sigma_T^2 + \left(\frac{\partial J}{\partial \Phi}\right)^2 \sigma_\Phi^2}. \quad (\text{Equation 4})$$

The voltage uncertainty σ_V is measured to be very small (10 μV). Uncertainty due to temperature σ_T may be relatively larger (1 K) due to the thermal resistance between the sensor and sample, and fluctuations in temperature due to ohmic heating during measurements. Note, however, that J is exponentially dependent upon V and T , so the partial derivative terms can become quite large, particularly at positive voltage biases. The final source of error, illumination variance σ_Φ , is by far the largest. With the class ABB solar simulator used here, the light intensity of the lamp is rated for a temporal stability range of $\pm 3\%$. For the typical current densities measured in the present work, the variation in illumination corresponds to a typical variability in current density σ_J of between 0.5 and 1.0 mA/cm^2 . To calculate the impact of these uncertainties on σ_J , we first compute the partial derivatives at all bias conditions. Fortunately, as J is exponential with V and T , the partial derivatives are simply proportional to J and do not need to be calculated for every bias condition. Note that this uncertainty is not a model uncertainty; we have assumed that our model certainty is high or that the general 1D partial differential equation model is an accurate description of the physics of thin-film solar cells.

The likelihood may now be computed element-wise for every set of parameters in Θ , and then multiplied by the prior probability distribution during each iterative observation step (as per [Equations 3 and 1](#)). Further details may be found in the [Supplemental Information](#).

Applying Bayesian Parameter Estimation to PV Devices

To validate our approach, we begin with a well-controlled test case of a simulated GaAs PV device. We model a high-performing GaAs solar cell with arbitrary layer thicknesses and doping densities. “Observation” data on this cell are produced

by setting all discrete input parameters and running the forward PDE model at several bias conditions. Then, the observation data are fed into the BPE algorithm and used to infer the underlying parameter values (θ) of the GaAs cell. We verify that the inferred parameters include the original forward simulation parameters to prove the accuracy of the approach. We can also observe where the precision of the inference is low, or where the device performance is not sensitive to the underlying material parameter. See [Supplemental Information](#) for more details on this process.

Next, the BPE approach is applied to real SnS devices. *JVTi* data were collected on a baseline SnS solar cell (room temperature data originally published in Steinmann et al.³⁰) using the method described in Chakraborty et al.²⁷ Then, particular *J-V* sweeps were selected from these data at 1.08-Sun and 0.31-Sun illumination intensity and temperatures of 280 K, 300 K, and 320 K (see full dataset in [Supplemental Information](#)).

The cell model was built for SnS, using parameters outlined in prior work²¹ with accurate layer thicknesses, doping levels, and optical constants (including front surface reflectance) based on prior measurements. Four parameters that govern efficiency are chosen for fitting: the minority carrier mobility in SnS, the minority carrier lifetime in SnS, the conduction band offset between SnS and Zn(O,S), and the effective surface recombination velocity at the SnS/Zn(O,S) interface. We assume that over our bias conditions, these parameters are constant. Uniform prior probability distributions were assigned to each of these parameters, and simulations were run over a 4D grid ($N = 4$) of discrete values ($M = 16\text{--}20$) in this parameter space at every bias condition, resulting in $96,000 \times 2$ illumination intensities $\times 3$ temperatures = 576,000 discrete device simulations. At 10 s per simulation, that would require 1,600 core-hours of HPC time, which is tractable in approximately 1 day with a small HPC cluster of 60 cores. Depending upon the operating system and simulation setup, SCAPS simulations may vary from 2 to 15 s each, which results in a different number of cores necessary to perform the simulation in <24 hr.

Each observation is then fed into the BPE algorithm as current density values at each bias condition. The likelihood of observing this output current is then computed for all possible hypotheses, as per [Equation 3](#), followed by computing the posterior using [Equation 1](#). This short-circuit condition constrains the joint distribution of minority carrier lifetime and mobility, but for this SnS device does not constrain interface properties; these retain fairly uninformed distributions, indicating that the information content for these parameters is low at applied voltages near 0 V. Moving further into forward bias, the nature of the interface recombination begins to appear; either the device must have a small cliff offset and a large S_{eff} or vice versa. This is reflected by a line of constant probability in θ space (see [Movie S1](#) for an animated version of [Figure 3](#)).

To constrain the distributions further, we vary illumination and temperature. Full *J-V* curves at 280 K, 300 K, and 320 K are fed in as observations, and the resulting posterior probability distribution is plotted in [Figure 3](#). While the minority carrier mobility still exhibits a relatively uniform probability, the minority carrier lifetime and interface properties are well constrained. The 95% confidence interval for τ is between 21 and 70 ps, for ΔE_C is -0.21 ± 0.03 eV, and for S_{eff} is 1,000–1,800 cm/s. While the confidence interval for μ is not well constrained by this series of measurements, the joint distribution of μ and τ is well constrained. This is a feature of BPE in its ability to uncover non-parametric co-varying distributions like these. In the case of PV materials, the $\mu\tau$

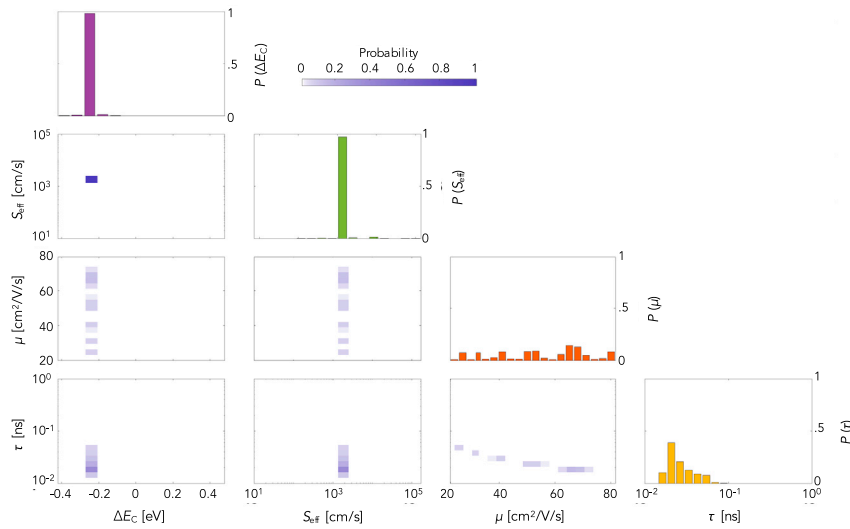


Figure 3. Probability Distribution of Fit Parameters after all Observations

For each pair of parameters in θ , single variable distributions are shown on the diagonal in four different colors, and multivariate distributions are shown off the diagonal. The scale bar for the multivariate distribution colormaps only is shown at top. These are the final probabilities after all observations, however the animated evolution may be seen in [Supplemental Information](#). The probability distributions are constrained to a narrow range of values for lifetime (yellow), band offset (purple), and surface recombination velocity (green) but not for mobility (red).

product is proportional to the characteristic transport length of free carriers; here, the $\mu\tau$ product is tightly constrained around $1.5 \times 10^{-9} \text{ cm}^2/\text{V}$. Note that the gaps or discontinuities visible in the distributions are largely artifacts of the gridded sampling and disappear with higher-resolution sampling.

Remarkably, these inferred values are consistent with lifetimes measured in the range of 20–40 ps in Jaramillo et al.,¹⁸ and conduction band offsets measured to be in the range of $-0.38 \pm 0.2 \text{ eV}$ in Mangan et al.²¹ extracted here with an all electrical measurement. Their precision is also promising, particularly for parameters that may vary over several orders of magnitude such as τ . In particular, the conduction band offset is inferred with higher precision than previous spectroscopic measurements (30 mV versus 200 mV). Furthermore, we are capable of extracting an estimate for S_{eff} , which was not previously possible to do for the completed heterojunction interface.

To further understand how each measurement shapes our beliefs about the underlying parameters, we may calculate the increase in information content (or decrease in entropy) of the distribution, normalized by the number of discrete hypotheses (M^N) as

$$s = \frac{P(\theta)\log(P(\theta))}{\log(\text{no. of hypotheses})}. \quad (\text{Equation 5})$$

In [Figure 4](#), the entropy of each parameter’s probability distribution is tracked as a function of the observation sequence, sweeping from 0 V (short circuit) to open-circuit conditions at each light intensity and each temperature. The entropies map well to what we might expect from device physics. For the bulk transport properties, the lifetime is not well constrained until we observe short-circuit conditions at multiple light intensities. Afterward, the entropy does not further decrease by additional observations, as the short-circuit current density is no longer providing any new

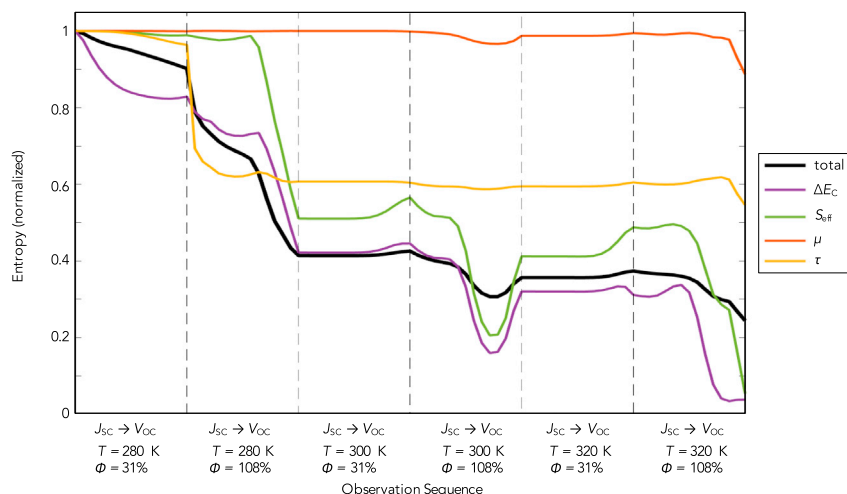


Figure 4. Entropy of the Probability Distributions Evolving with Each Observation

The axis labels represent measurement temperature (in Kelvin) and illumination intensity (% of 1 Sun). On the x axis, over each temperature and illumination, the bias voltage is varied between 0 V (short-circuit current or J_{SC}) and open-circuit voltage (V_{OC}) in discrete voltage steps.

information or sufficiently useful information. Across all observations, there is insufficient information to constrain the minority carrier mobility further. The interface properties are well constrained generally only when we move close to open-circuit conditions at higher light intensities. This is first observed strongly at the V_{OC} point at 280 K and again further at 300 K and 320 K. In addition, the entropy may rise locally, likely where observations carry noise or systemic deviations that would implicate a different parameter set. This lack of a monotonic decrease is a reflection that different observations offer slightly conflicting evidence. Tracking the entropy of the probability distribution suggests when measurements may be complete and highlights which parameters deserve more attention and further dedicated measurement.

DISCUSSION

In under 24 hr of device measurement and concurrent BPE computation, we show that it is possible to infer four underlying materials properties of a SnS device. These parameter fits revealed comparable accuracy, and in some cases higher precision, with what was measured through more expensive and time-intensive spectroscopic techniques. The significance of this result is that BPE enabled by HPC has offered a first example of how the empirical cycle of PV materials development may be greatly accelerated.

SnS and other low-efficiency materials offer a great opportunity with BPE, as the Bayesian approach works best in identifying underperforming material properties; the more impact a parameter has on device performance, the more information content we will have about that parameter. Once the parameter exceeds a critical threshold, wherein device performance is not affected by it, it is difficult to infer its precise value from device performance, as seen in the case of GaAs.

Where performance losses are uncertain, the BPE technique may be applied to many other PV materials, such as the lead halide perovskites. The [Supplemental Information](#) contains generalized code for plugging in any cell model and measurements to

begin inferring multiple unknown material or interface properties. For the lead halide perovskites in particular, this technique will require two things. The first is stable devices, as the properties of the device must remain constant over the course of measuring. The second requirement is a representative model that accurately captures the underlying physics. Lead halide perovskite performance, for example, is affected by ion diffusion to interfaces,³¹ and this effect is typically not modeled into common device solving software. To increase parameter certainty, model uncertainty must be removed.

Building upon this initial proof of concept, there are many ways in which the BPE approach may be improved upon and generalized more broadly.

Firstly, this mesh or grid approach to BPE can be improved. Not only does the current method sample discretely, and may therefore miss important regions of solution space if the mesh is chosen too sparsely, but it also spends significant time sampling and computing likelihoods in regions of θ space that have a very low probability. The simplest strategy to avoid this is to begin with sparse sampling of θ space and then use the initial results to sample the relevant parts of θ space more densely. A more robust and automated method to refine the sampling numerically is a Markov Chain Monte Carlo algorithm.³² In this approach, the choice of where to run simulations is combined with the process of computing the likelihood and posterior, meaning that sampling is much more efficient and produces higher-resolution data only in the regions of interest.

Another important improvement is increased automation of the observations themselves through active learning.³³ This class of machine learning techniques allows the BPE algorithm to actively advise on what data it would like to be fed, selecting bias conditions most likely to provide high information content and reduce the entropy of the probability distribution over θ more efficiently.

The reliability of BPE depends strongly on the choice of noise model, the choice of prior probabilities, and the quality of the device model used. Regarding the noise model, we have constructed this through direct measurement of the noise in bias conditions, and may therefore compute the expected noise in measured current based on a simple error propagation analysis. Regarding the choice of priors, here we selected uniform priors with upper and lower bounds outside of the values we have characteristically seen for early-stage PV absorbers. One may reduce the risk of error by allowing for the possibility of non-zero probabilities outside of this range, but this makes the ability to discretize our parameter space very difficult. To make BPE tractable, we make assumptions that must be rigorously tested later on; for example, checking to ensure that our inferred probability distribution does not sit at the boundary of our prior.

The fit accuracy and precision may also be influenced by assumptions made about other inputs into the model; where these are uncertain, they should be included as fit parameters. One relevant example of this is our decision to assume temperature- and injection-independence of parameters over the bias conditions. Bias-dependent parameters can be easily accommodated if the functional form of the dependence is included, but this increases the number of fit parameters accordingly. For example, series resistance is expected to vary with temperature. Here, we have fixed series and shunt resistances in our model and restricted measurements to a narrower temperature and voltage range, but this assumption may influence our estimated parameters and residual entropy of their distributions.

The BPE approach can in principle also account for systematic errors in the observation conditions, distinct from the random errors described earlier. For example, the device temperature may consistently be 2 K warmer than the temperature sensor reads due to a probe miscalibration, or the light intensity may consistently be 5% weaker than anticipated due to a damaged optic element. These “nuisance parameters” can also be accounted for in BPE²⁹ by assuming that they do exist as non-zero values and treating them as input parameters in Θ with their own prior probability distributions. After running BPE, the parameters of interest will have joint probability distributions with the nuisance parameters. In other words, the minority carrier lifetime will co-vary with the expected lamp-power offset, for example. Integrating over the nuisance parameter’s probability distribution marginalizes it and results in an adjusted distribution for the parameters of interest. This new distribution can now fully account for the possibility of systematic errors in the measurement.

Lastly, all parameter estimation is only as good as the model built to describe the PV cell. The present model has been verified for SnS thin-film PV cells through prior work,^{11,17} and its underlying parameters are described further in [Supplemental Information](#). It is difficult to provide an absolute quantification of how good the chosen model is, beyond qualitative observations of the goodness of fit (see [Supplemental Information](#)). However, one may quantify the relative accuracy of different competing models by treating each model as a different hypothesis and thereby comparing the probability of observing the evidence under each different model.

Conclusion

With increasing demands upon the field of materials science to discover and develop materials at a faster rate, it is important for materials scientists to identify areas where HPC may be leveraged to complement and improve upon classical experimental techniques. In complex optoelectronic and energy devices, materials and interface properties represent the largest unknown and the primary focus of scientific research.

In the present work, we have demonstrated that a BPE approach may help address the uncertainties of empirical materials characterization and development by greatly accelerating the experimental feedback loop compared with slower techniques such as spectroscopy. This BPE framework was used to pinpoint underperforming materials parameters in a solar-cell material, SnS, with 10× faster throughput than traditional spectroscopy techniques, measuring material and interface parameters in a device-relevant form factor. It also identified cross-correlated parameters and their joint probability distributions, which are difficult to extract using classical statistical techniques without prior knowledge of their functional forms.

The implications of this proof of concept are broad. Through a similar BPE approach, substituting the physical system model for those of batteries, thermoelectrics, transistors, and many other complex devices, we propose that experimentalists may be able to significantly improve their rate of learning and development. As the 21st century demands faster materials development, HPC tools such as BPE will enable this across a variety of fields.

EXPERIMENTAL PROCEDURES

Full experimental procedures are provided in the [Supplemental Information](#).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, two tables, and one movie and can be found with this article online at <https://doi.org/10.1016/j.joule.2017.10.001>.

AUTHOR CONTRIBUTIONS

R.E.B. and R.C.K. planned the experiment and Bayesian approach with assistance from T.B. and C.R. V.S. designed and fabricated the cell measured here, and R.E.B. performed measurements with the assistance of S.L. and T.U. R.E.B. developed the cell model and Python code, and R.E.B., R.C.K., and D.K. implemented and ran the algorithm together. G.C. and D.K. assisted with setting up HPC resources. R.E.B., R.C.K., and T.B. wrote the manuscript with assistance from all co-authors.

ACKNOWLEDGMENTS

The authors acknowledge Rupak Chakraborty, Chuanxi Yang, and Roy Gordon for assistance with cell design and fabrication, as well as Niall Mangan and Rafael Jaramillo for insightful conversations. The authors also acknowledge the assistance of several members of Google Accelerated Sciences for helpful discussions; and would like to recognize Mark Fickett, Break Yang, Terran Lane, Kate Steele, and Arielle Bertman for facilitating initial introductions and guidance. This work was primarily supported by a Google Faculty Research Award to develop and access HPC capabilities and expertise, as well as the Center for Next Generation Materials by Design (CNGMD), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, which supported film synthesis, theory, and basic materials characterization. Advanced materials characterization was enabled by a TOTAL research grant. A DAAD fellowship provided R.E.B. with travel funds to visit S.L. and T.U. in Germany to conduct additional *JVTi* measurements. A portion of the computational work was performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory.

Received: June 17, 2017

Revised: September 12, 2017

Accepted: September 28, 2017

Published: December 20, 2017

REFERENCES

- Green, M.A., Emery, K., Hishikawa, Y., Warta, W., and Dunlop, E.D. (2015). Solar cell efficiency tables (version 46). *Prog. Photovolt. Res. Appl.* **23**, 805–812.
- Unold, T., and Schock, H.W. (2011). Nonconventional (non-silicon-based) photovoltaic materials. *Annu. Rev. Mater. Res.* **41**, 297–321.
- Curtarolo, S. (2013). Fueling discovery by sharing. *Nat. Mater.* **12**, 173.
- Musso, C.S. (2005). *Beating the System: Accelerating Commercialization of New Materials* (PhD thesis (Massachusetts Institute of Technology)).
- Gaddy, B.E., Sivaram, V., Jones, T.B., and Wayman, L. (2017). Venture capital and cleantech: the wrong model for energy innovation. *Energy Policy* **102**, 385–395.
- Bretthorst, G.L. (2013). *Bayesian Spectrum Analysis and Parameter Estimation* (Springer Science & Business Media).
- Christensen, N., Meyer, R., Knox, L., and Luey, B. (2001). Bayesian methods for cosmological parameter estimation from cosmic microwave background measurements. *Class. Quant. Gravity* **18**, 2677.
- Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.* **37**, 2521–2535.
- Sinton, R.A., and Cuevas, A. (1996). Contactless determination of current-voltage characteristics and minority-carrier lifetimes in semiconductors from quasi-steady-state photoconductance data. *Appl. Phys. Lett.* **69**, 2510–2512.
- Giridharagopal, R., Cox, P.A., and Ginger, D.S. (2016). Functional scanning probe imaging of nanostructured solar energy materials. *Acc. Chem. Res.* **49**, 1769–1776.
- Leite, M.S., Abashin, M., Lezec, H.J., Gianfrancesco, A., Talin, A.A., and Zhitenev, N.B. (2014). Nanoscale imaging of photocurrent and efficiency in CdTe solar cells. *ACS Nano* **8**, 11883–11890.
- Kalinin, S.V., Sumpster, B.G., and Archibald, R.K. (2015). Big-deep-smart data in imaging for guiding materials design. *Nat. Mater.* **14**, 973–980.

13. Wagstaff, K. (2012). Machine learning that matters. *arXiv*, arXiv:1206.4656.
14. Sinsermsuksakul, P., Sun, L., Lee, S.W., Park, H.H., Kim, S.B., Yang, C., and Gordon, R.G. (2014). Overcoming efficiency limitations of SnS-based solar cells. *Adv. Energy Mater.* **2014**, 1400496.
15. Jaramillo, R., Steinmann, V., Yang, C., Hartman, K., Chakraborty, R., Poindexter, J.R., Castillo, M.L., Gordon, R., and Buonassisi, T. (2015). Making record-efficiency SnS solar cells by thermal evaporation and atomic layer deposition. *J. Vis. Exp.* **99**, e52705.
16. Steinmann, V., Jaramillo, R., Hartman, K., Chakraborty, R., Brandt, R.E., Poindexter, J.R., Lee, Y.S., Sun, L., Polizzotti, A., Park, H.H., et al. (2014). 3.88% efficient tin sulfide solar cells using congruent thermal evaporation. *Adv. Mater.* **26**, 7488–7492.
17. Chakraborty, R., Steinmann, V., Mangan, N.M., Brandt, R.E., Poindexter, J.R., Jaramillo, R., Mailoa, J.P., Hartman, K., Polizzotti, A., Yang, C., et al. (2015). Non-monotonic effect of growth temperature on carrier collection in SnS solar cells. *Appl. Phys. Lett.* **106**, 203901.
18. Jaramillo, R., Sher, M.-J., Ofori-Okai, B.K., Steinmann, V., Yang, C., Hartman, K., Nelson, K.A., Lindenberg, A.M., Gordon, R.G., and Buonassisi, T. (2016). Transient terahertz photoconductivity measurements of minority-carrier lifetime in tin sulfide thin films: advanced metrology for an early stage photovoltaic material. *J. Appl. Phys.* **119**, 35101.
19. Park, H.H., Heasley, R., Sun, L., Steinmann, V., Jaramillo, R., Hartman, K., Chakraborty, R., Sinsermsuksakul, P., Chua, D., Buonassisi, T., et al. (2014). Co-optimization of SnS absorber and Zn (O, S) buffer materials for improved solar cells. *Prog. Photovolt. Res. Appl.* **23**, 901–908.
20. Schneikart, A., Schimper, H.-J., Klein, A., and Jaegermann, W. (2013). Efficiency limitations of thermally evaporated thin-film SnS solar cells. *J. Phys. Appl. Phys.* **46**, 305109.
21. Mangan, N.M., Brandt, R.E., Steinmann, V., Jaramillo, R., Yang, C., Poindexter, J.R., Chakraborty, R., Park, H.H., Zhao, X., Gordon, R.G., et al. (2015). Framework to predict optimal buffer layer pairing for thin film solar cell absorbers: a case study for tin sulfide/zinc oxysulfide. *J. Appl. Phys.* **118**, 115102.
22. Chakraborty, R. (2016). Structural Defect Engineering of Tin(II) Sulfide Thin Films for Photovoltaics (Thesis (Massachusetts Institute of Technology)).
23. Scheer, R. (2009). Activation energy of heterojunction diode currents in the limit of interface recombination. *J. Appl. Phys.* **105**, 104505.
24. Hages, C.J., Carter, N.J., Agrawal, R., and Unold, T. (2014). Generalized current-voltage analysis and efficiency limitations in non-ideal solar cells: case of $\text{Cu}_2\text{ZnSn}(\text{S}_x\text{Se}_{1-x})_4$ and $\text{Cu}_2\text{Zn}(\text{Sn}_y\text{Ge}_{1-y})(\text{S}_x\text{Se}_{1-x})_4$. *J. Appl. Phys.* **115**, 234504.
25. Grover, S., Li, J.V., Young, D.L., Stradins, P., and Branz, H.M. (2013). Reformulation of solar cell physics to facilitate experimental separation of recombination pathways. *Appl. Phys. Lett.* **103**, 93502.
26. Brandt, R.E., Mangan, N.M., Li, J.V., Lee, Y.S., and Buonassisi, T. (2017). Determining interface properties limiting open-circuit voltage in heterojunction solar cells. *J. Appl. Phys.* **121**, 185301.
27. Chakraborty, R., Brandt, R.E., Steinmann, V., Rekemeyer, P., Gradečak, S., and Buonassisi, T. (2017). Best practices for the application of temperature- and illumination-dependent current density-voltage $J(V, T, I)$ and electron-beam induced current EBIC to novel thin film solar cells. *arXiv*, arXiv:170305477. <https://arxiv.org/abs/1703.05477>.
28. Burgelman, M., Nollet, P., and Degraeve, S. (2000). Modelling polycrystalline semiconductor solar cells. *Thin Solid Films* **361–362**, 527–532.
29. Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104.
30. Steinmann, V., Chakraborty, R., Rekemeyer, P.H., Hartman, K., Brandt, R.E., Polizzotti, A., Yang, C., Moriarty, T., Gradečak, S., Gordon, R.G., et al. (2016). A two-step absorber deposition approach to overcome shunt losses in thin-film solar cells: using tin sulfide as a proof-of-concept material system. *ACS Appl. Mater. Interfaces* **8**, 22664–22670.
31. van Reenen, S., Kemerink, M., and Snaith, H.J. (2015). Modeling anomalous hysteresis in perovskite solar cells. *J. Phys. Chem. Lett.* **6**, 3808–3814.
32. Gamerman, D., and Lopes, H.F. (2006). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition (CRC Press).
33. Tong, S., and Koller, D. (2001). Active Learning for Structure in Bayesian Networks. In *International Joint Conference on Artificial Intelligence, Vol. 17* (Lawrence Erlbaum), pp. 863–869.