Physiochemical Machine Learning Models Predict Operational Lifetimes of $CH_3NH_3PbI_3$ Perovskite Solar Cells

Wiley A. Dunlap-Shohl,[†] Yuhuan Meng,[†] Preetham P. Sunkari,[†] David A. C. Beck, [†,‡] Marina Meila,[‡,¶] Hugh W. Hillhouse[†,*]

[†] Department of Chemical Engineering, Clean Energy Institute, and Molecular Engineering and Science Institute, University of Washington; Seattle, USA.

[‡] eScience Institute, University of Washington; Seattle, USA.

[¶] Department of Statistics, University of Washington; Seattle, USA.

*Corresponding author. E-mail: h2@uw.edu.

Abstract: Halide perovskites are promising photovoltaic (PV) materials with the potential to lower the cost of electricity and greatly expand the penetration of PV if they can demonstrate long-term stability under illumination in the presence of moisture and oxygen. The solar cell service lifetime as quantified by the T_{80} (the time required for the power conversion efficiency to drop to 80% of its starting value) is a useful metric to assess stability. The T_{80} for utility, commercial, or residential PV systems needs to be several decades in order to yield low-cost electricity, and thus it is not practical to directly measure the T_{80}. It would be useful if T_{80} could be predicted from the initial dynamics of a solar cell's performance, but until now no models have been developed to forecast T_{80}. In this work, we report the development of machine learning models to predict T_{80} of ITO/NiO_x/$CH_3NH_3PbI_3$/C_{60}/BCP/Ag solar cells operating at maximum power point under 1-sun equivalent photon flux in air at varying temperatures and relative humidities. Efficiency losses are driven by short-circuit current and fill factor, indicating that chemical decomposition of the perovskite is a major contributor to degradation. Spatial patterns evident from in situ dark field optical microscopy suggest that the electric field gradient at device edges plays a significant role in perovskite decomposition, along with photochemical reactions with O_2 and H_2O. Models are trained using a menu of features from three distinct categories: (i) features based on measurements of the initial rates of change of device parameters, (ii) features based on the ambient conditions during operation (temperature, & partial pressure of H_2O), and (iii) features based on underlying physics and chemistry. We show that a theory-based physiochemical feature derived from a model of the chemical reaction kinetics of the rate of degradation of the $CH_3NH_3PbI_3$ is particularly valuable for prediction. This physiochemical feature was selected as the first or second most dominant feature in the best performing models. With a dataset consisting of 45 accelerated degradation experiments with T_{80} that range over a factor of 30, the model predicts T_{80} with an accuracy of about 40% (|predicted T_{80} - observed T_{80}| / observed T_{80}) on samples not used in training. This hybrid ML approach should be effective when applied to other compositions, device architectures, and advanced packaging schemes.

Introduction. Halide perovskites are on the cusp of breaking out as mainstream commercial photovoltaic materials, but there are some concerns about their potential for long-term stability due to the susceptibility of perovskite absorbers to decomposition in the presence of heat,[1] light,[2,3] oxygen,[2] moisture,[4–6] and electrical bias.[7–10] The success of perovskite solar cells (PSCs) and other perovskite-based technologies may rest on how accurately manufacturers are able to estimate the service lifetimes of devices. Accurate predictions will allow manufacturers to identify appropriate markets and provide warranties, while at the same time allowing customers to assess the economics of the purchase. There have been many promising advances toward extending the operational life of perovskite solar cells, with several reports of perovskite solar cells passing the standard IEC solar cell durability tests (e.g., damp heat and thermal cycling).[11–14] However, the mechanisms of degradation are quite different in perovskites than in conventional solar cells, and it is currently unclear if PSCs that can survive on the time scale of decades in the wide variety of potential operational environments. In the early years of perovskite photovoltaics research, the large number of possible degradation processes, coupled with a lack of information on how to assign a hierarchy of their effects on material and device degradation, led to many testing protocols (accelerated or otherwise) generally developed independently by individual research groups. This made comparing the stability data from lab to lab difficult. Even in cases where standardized accelerated testing regimes (such as the IEC protocols) were obeyed, the IEC tests were developed for devices based on conventional inorganic semiconductors, and do not necessarily probe the performance-limiting processes in PSCs. Although the perovskite research community has started to develop testing protocols that are better designed to capture the most important degradation mechanisms,[15] the field's knowledge of these mechanisms remains incomplete, complicating these efforts. Furthermore, the large number of perovskite compositions and device architectures introduces massive complexity to the space of possible chemical degradation pathways. A recent report[16] attempting to construct a comprehensive overview of perovskite photovoltaics research catalogs over 5,500 architectures as determined by the contact materials alone—i.e., not even accounting for the additional complexities posed by the compositional flexibility of the perovskite absorber itself—of which over 1,000 were reported to yield PSCs that have power conversion efficiency >18% and thus represent potentially attractive fabrication strategies for commercialization. Since the principal degradation mechanisms that limit device lifetimes, and the major environmental factors that dictate them, may vary considerably across perovskite absorbers and device architectures, the development of “one-size-fits-all” testing protocols is a risky strategy for lifetime validation.

Development of mathematical models that can predict operational lifetimes can significantly reduce the uncertainty associated with selecting the correct testing protocols. Prediction offers several crucial advantages over standardized protocol-focused testing. It can assimilate information gathered under non-ideal circumstances (e.g., negative results based on “bad” devices or those with short lifetimes, which are seldom reported in the literature) that can nevertheless furnish useful information for model training. Through the incorporation of physical variables such as temperature or humidity as predictive features, these models’ functional forms can also indicate which aspects of degradation are most important. Despite their advantages, sufficiently accurate models can be challenging to build. In the limit of perfect information about a device’s construction, packaging, and the conditions it is subjected to, an idealized model would be able to describe how the chemical interactions resulting from environmental exposure and operation lead to changes in its performance. Such a model may be envisioned in principle as a set of coupled partial differential equations that could be solved in a 3-dimensional multi-physics numerical simulation. However, this approach would require thousands of variables and immense computing resources. Even assuming resource-intensiveness is not an impediment, many of the relevant physical processes this model would need to describe are currently unknown, ultimately making this strategy prohibitively difficult to achieve in practice.

If a fully mechanistic model represents a “bottom-up” approach proceeding from elementary physical knowledge, we may consider models constructed via machine learning (ML) as providing an opposite, “top-down” approach that leverages empirically-determined relationships between the target behavior (i.e., how power conversion efficiency evolves over time) and easily measured variables of interest with a clear relationship with the target (e.g., environmental conditions, initial performance metrics, or details of device architecture and fabrication). ML models have recently been gaining ground in perovskite research, particularly in areas such as stability where important physical processes are insufficiently known to enable fully mechanistic descriptions. For instance, Howard et al.[17] and Srivastava et al.[18] have recently used neural networks to predict the evolution of perovskite photoluminescence intensity in response to humidity fluctuations and $MAPbI_3$ solar cell power conversion efficiency evolution under thermal stress, respectively. However, the evolution was not investigated for times long enough to reach the device T_{80} (the time taken for the power conversion efficiency, PCE, to drop to 80% of its starting value) that serves as the traditional figure of merit for device longevity, making it unclear how well these models would fare in full lifetime prediction. An additional drawback of neural networks is that they are not interpretable. Machine learning has also been used to optimize perovskite stability: Hartono et al.[19] used a variety of supervised ML techniques to optimize 2D perovskite capping layers for protecting $MAPbI_3$ against damp heat environments, finding that random forest regression performed best among all modeling approaches tested; Sun et al.[20] combined DFT calculations and Bayesian optimization to identify the most stable members in the ternary cesium-formamidinium-methylammonium lead iodide perovskite family (all synthesized under identical processing conditions).

Physics-informed machine learning is an especially promising strategy that can address some of the shortcomings of fully empirical ML by incorporating mechanistic knowledge or imposing physical constraints on predictive models. This class of models can, in principle, make maximum use of available mechanistic knowledge while retaining the economy and flexibility of empirical models. With regards to the challenge of predicting PSC service lifetimes, conventional wisdom dictates that chemical decomposition of the absorber is a major factor determining the rate of device performance loss. Mathematical descriptions of perovskite decomposition pathways are therefore expected to be important ingredients of predictive models that may reduce reliance on potentially useful but uninformative empirical relationships between service lifetime and environmental conditions. Recently, we have shown that the initially observed chemical decomposition rate of $CH_3NH_3PbI_3$ thin films is a good predictor of decay of the films' carrier diffusion length over time.[21] Furthermore, we have shown that the initial decomposition rate of $CH_3NH_3PbI_3$ films can be predicted accurately from temperature, above-band gap illumination intensity, and ambient partial pressures of H_2O and O_2.[6] In this work, we develop the first predictive machine learning model of perovskite solar cell operational lifetime (as quantified by the device T_{80}). T_{80} is predicted as a function of the ambient environmental conditions as well as measurements of current-voltage characteristics taken over the first ~90 minutes of operation. The models are trained using 45 experiments carried out under 1-sun equivalent illumination in air at a variety of different temperatures and humidities. This work exemplifies how physics-informed ML models can be used to unite mechanistic physical information with sample-specific observations of performance evolution to both maximize predictive accuracy and model interpretability. The time scale for decay of other metrics such as J_{sc}, V_{oc}, FF, diffusion length, etc. can be formulated as a time to decay to 80% of its initial value ($T_{80,Jsc}$, $T_{80,Voc}$, etc.), but unless otherwise noted, we will use "T_{80}" to refer to T_{80} specifically for the PCE.

Analysis of Device Degradation Data. Device degradation data are collected in a testing station where temperature, humidity, oxygen, and illumination are controlled (**Figure 1a**). Solar cells (approximate area: 0.07 cm^2) are placed underneath the objective lens of a microscope equipped with a light source and low-magnification objective that illuminates the entire device with a 1 sun equivalent above-band gap photon flux of 542±17 nm light (i.e., 1.56×10^{21} photons $\cdot$ m^{-2} $\cdot$ s^{-1}). Electrical characteristics are measured *in situ* using a Keithley 2420 source-measure unit. The device under test is placed on a thermal stage to regulate temperature, which is then enclosed in a controlled-humidity chamber. Device operating characteristics are collected periodically (for most runs, every 15 minutes), including steady state open-circuit voltage, short-circuit current, and maximum power point voltage and current, as well as forward and reverse-scan *J-V* sweeps. In the interim period between measurements, devices are held at the bias corresponding to the most recently determined maximum power point. The device degradation dataset used for T_{80} modeling comprises 45 runs representing 37 distinct environmental conditions collected in air at temperatures ranging from 25 to 85 °C and relative humidities ranging from 0 to 70%. Solar cells are fabricated in a commonly used inverted ITO/NiO_x/$CH_3NH_3PbI_3$/C_{60}/bathocuproine (BCP)/Ag architecture. Devices used in this study have mean initial PCE of 13.8 ± 1.4%, J_{sc} of 19.6 ± 1.2 mA/cm^2, V_{oc} of 0.984 ± 0.025 V, and fill factor of

71.7 ± 4.8% (uncertainty represented by the standard deviation). Note that we do not use a shadow mask to constrain device active area, in order to facilitate microscopy across device edges as well as the interior. Therefore, the active area is assumed to correspond to the overlap area between the ITO and Ag electrodes and may lead to minor inaccuracy in the estimation of J_{sc} due to light-piping effects or fluctuation of individual device active area relative to the nominal value due to variation in substrate-mask alignment during Ag contact deposition. The T_{80} values in this dataset vary over an order of magnitude, ranging from under 200 min in high-thermal stress conditions to almost 5000 min at lower temperature and moderate humidity. Histograms of these quantities and a breakdown of the environmental conditions investigated in this study are plotted in **Figure S1.**

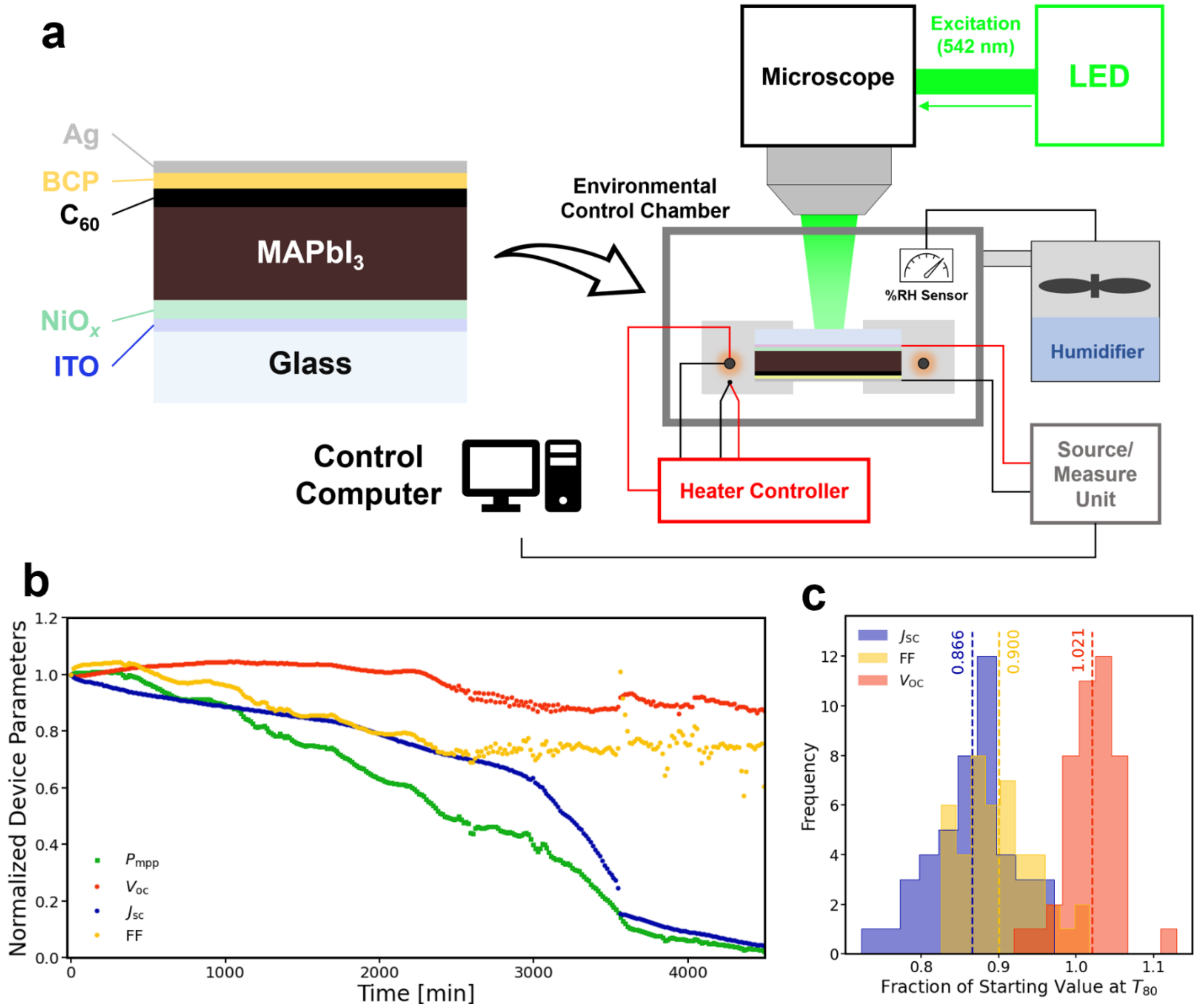

Figure 1. (a) Schematic of device architecture and degradation equipment. (b) Data from a typical run collected under 1 sun illumination at 25 °C in 50% RH air, showing that the power conversion efficiency closely tracks the evolution of short-circuit current, but is also affected by an early decline in fill factor and a later, more modest decline in open-circuit voltage. (c) Histograms of short-circuit current density J_{sc}, fill factor FF, and open-circuit voltage V_{oc} at T_{80} relative to their initial values show that on average, J_{sc} and FF losses account for most of the decline in PCE, while V_{oc} increases modestly. Dashed lines in (c) represent mean values of the histograms with the corresponding colors.

Typical device parameter evolution is shown in **Figure 1b**, using a sample run collected at 25 °C and 50% RH. The overall trajectory of PCE is dominated by the evolution of J_{sc}, which decays monotonically with an initial, relatively slow plateau giving way to a more rapid drop-off as the perovskite absorber decomposes (as indicated by dark field microscopy; see discussion below). PCE is boosted at first by gains in both fill factor and V_{oc}, but while V_{oc} remains relatively high over the course of degradation, fill factor experiences a steady decline after the initial rise. The mid-term decay in fill factor also plays a significant role in PCE decline. Looking across the entire dataset (**Figure 1c**), PCE losses before T_{80} are determined mostly by J_{sc} (on average, at 87% of its starting value at T_{80}), followed closely by fill factor (90% of its starting value), while V_{oc} increases tend to buffer these losses slightly (102% of its starting value). Overall, this behavior is consistent with what we[21,22] and others[23] have observed when examining the degradation of perovskite films' optoelectronic properties: carrier transport (i.e., mobility and diffusion length) is much more sensitive to perovskite decomposition than carrier lifetime is, implying that J_{sc} should decay much more rapidly than V_{oc}. The fill factor (FF) is affected by both mobility and lifetime but also potentially by processes in the device not directly related to absorber decomposition and falls in the middle.

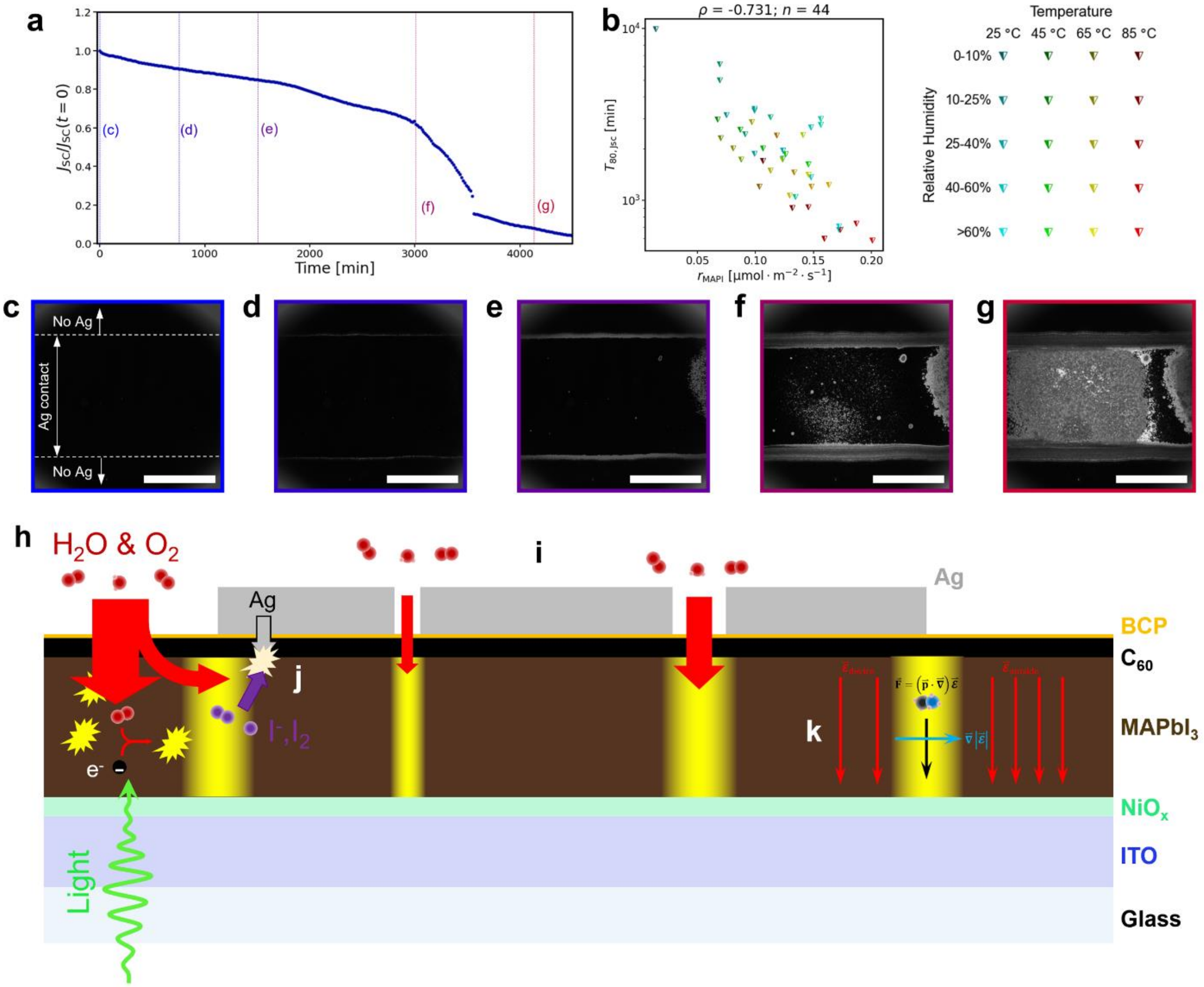

Figure 2. The short-circuit current evolution of $CH_3NH_3PbI_3$ solar cells is closely related to material decomposition. (a) J_{sc} evolution of a representative device at 25 °C and 50% RH. (b) Scatterplot of $T_{80,Jsc}$, the time at which J_{sc} reaches 80% of its maximum value, versus the kinetically modeled decomposition rate of $CH_3NH_3PbI_3$ at the temperature and humidity indicated by the symbol color, showing a strong logarithmic correlation. (c-g) Dark field images of the device taken at representative points in (a) denoted by the labeled vertical lines: (c) Initially, the image is almost

completely featureless, indicating absence of macroscopic defects. (d,e) In early stages of degradation, material decomposition is most pronounced at the edges and weak points in the device interior, which are most vulnerable to ingress of atmospheric species. (f) At the knee at which J_{sc} begins to decrease sharply, decomposition starts to occur homogeneously throughout the interior. (g) When less than 10% of the initial J_{sc} remains, most of the device has experienced severe material degradation, with only a small fraction of the active area remaining intact. Approximate edges of the device, determined by the overlap of the Ag and ITO electrodes, are denoted by dashed lines in (c). Scale bars in (c-g) are 1 mm. (h-k) Schematic of hypothetical degradation modes in $MAPbI_3$ solar cells operating in air. (h) Photooxidation of the absorber is likely to occur at regions unprotected by the top Ag contact or at the edges of the device, where moisture and oxygen may diffuse laterally. (i) Fabrication defects (e.g., thin spots in the contact due to shadowing by dust particles on the surface) may also serve as ingress routes for oxidizing species. (j) Iodine may react with Ag from the contact to produce AgI. (k) Electric field gradients at device edges induce a force on dipolar species such as MA^+ that may break down the perovskite absorber. Defects in the Ag contact as in (i) may also act as device edges, contributing to gradient-induced degradation as well.

The significant average reduction in J_{sc} at T_{80} indicates that perovskite decomposition is a process of major importance in device failure. Moreover, the logarithm of the time at which J_{sc} reaches 80% of its starting value ($T_{80,Jsc}$) obeys a moderately strong correlation (Pearson coefficient ρ = -0.73) with the chemical decomposition rate of $CH_3NH_3PbI_3$ predicted by our kinetic model for a given combination of temperature and humidity (**Figure 2b**), further suggesting that the photooxidation processes that dominate perovskite material degradation when films are exposed to the air also represent a major factor in PCE loss for full devices. Bryant et al.[24] have shown convincing corroborating evidence that photooxidation-like processes driven by injected electrons under dark O_2-containing environments lead to rapid degradation underneath the active electrode of ITO/TiO_2/$CH_3NH_3PbI_3$/Spiro-OMeTAD/Au PSCs while leaving adjacent perovskite regions intact, including those under unbiased contacts. Here the impact of material decomposition in devices is imaged using dark field (DF) microscopy in reflection geometry, which is sensitive to spatial changes in refractive index (which would occur during the conversion of perovskite to secondary phases) or film roughening that increase scattering.[21] The bright regions in dark field images correspond to locations where significant portions of the perovskite absorber have decomposed into PbI_2, which is the only solid degradation product of the dominant water-accelerated photooxidation pathway.[6] PbI_2 is easily recognized by its bright yellow color, which the naked eye can perceive clearly in heavily degraded devices. Dark field images of pristine devices are initially flat and featureless (**Figure 2c**), indicating uniform device layers largely free of macroscopic defects. As degradation progresses (**Figure 2d-g**), material transformation is at first most noticeable at the device edges and at intermediate times, isolated defects in the interior. The "knee" in J_{sc} is approximately concurrent with a strong increase in the scattering of the incident light (the appearance of large numbers of bright spots in the dark field images) throughout the device (**Figure 2f**). These more homogeneous patterns suggest that, at longer timescales, diffusion of O_2 and H_2O directly and more uniformly through the contact (for instance, along grain boundaries as opposed to regions of poor coverage created by fabrication defects) may cause degradation as well. This interpretation is supported by the observation that thicker Ag contacts (300 nm vs. the standard 100 nm used in all other devices in this work) suppress the rates of both J_{sc} decay (0.0043 %/min for the 300 nm contact vs. 0.0102 %/min for the 100 nm contact, both calculated from a linear fit to the time evolution of J_{sc} over the first 1000 min of each experiment) and dark field intensity rise (**Figure S2**), indicating that degradation is mediated to some extent by mass transport directly through the contact. However, performance is already severely compromised by degradation at edges and macroscopic defects by the time this homogeneous degradation mode becomes significant.

The spatial patterns in the dark field images indicate that photooxidation alone may be an incomplete explanation for the progression of solar cell degradation. If it were the only factor, we would expect to see the most degradation where the perovskite is least protected (i.e., the region outside the Ag contact that defines the device), yet the region most vulnerable to decomposition occurs at the device boundary. At early stages of degradation, dark field intensity of the exposed areas degrades much faster than the device interior (probing a region free of macroscopic defects indicates no significant trend of increasing intensity), signifying that the contact still acts as an effective diffusion barrier during this period (**Figure S3a,b**). However, degradation outside the contact is quickly outstripped by the edge region (**Figure S3c**). Overall, these patterns imply that both photooxidation and a separate degradation mode (or modes) introduced by the contact combine to make decomposition most severe at the device boundary. Two broad categories by which the contact might exacerbate degradation are: (i) chemical reactions between halide ions released by the perovskite and the Ag metal and (ii) the influence of the electric field induced by the presence of the contact.

The possibility of chemical reactions is supported by observations made by Besleaga et al.[25] in which FTO/TiO_2/$CH_3NH_3PbI_3$/Spiro-OMeTAD devices under Ag contacts spontaneously degrade even under storage in dark, low-humidity (~10% RH) conditions, while equivalent devices with a protective Mo layer beneath the Ag or with Au substituted for Ag entirely degrade slower or not at all under the same conditions, within the limits of detection. This behavior is attributed to the ability of Ag to act as a chemical sink for I^- ions (producing AgI), which, when combined with the rapid diffusivity of the latter in the Spiro-OMeTAD hole transport layer, creates a chemical gradient that depletes the perovskite of iodine, causing it to break down. We expect that if this process were a major trigger of degradation, it would occur more homogeneously across the device, but we do not exclude the possibility that chemical reactions with the Ag electrode play some role. Contrary to the report of Besleaga et al.,[25] however, when the Ag contact is replaced with Au, J_{sc} decays and dark field intensity rises more rapidly despite the latter ostensibly being more stable (**Figure S2**). Kerner et al.[37] have noted that Au can also react readily with iodine compounds, especially the oxidized forms such as I_2 and I_3^-, the former of which is a likely product of photooxidation while the latter may be formed by subsequent reaction with fresh I^- from the perovskite. It may thus be the case that, when regions near the edges of the contacts are weakened by photooxidation, that decomposition products from the perovskite may facilitate additional reactions with the metal electrodes, accelerating the overall process of degradation.

Alternatively, bias-induced degradation may also be invoked to account for the influence of the contacts. Leijtens et al.[9] observed that sustained application of bias to $CH_3NH_3PbI_3$ films leads to methylammonium accumulation near the cathode and depletion near the anode, eventually leading to irreversible decomposition of the perovskite. They also observed that this process can be accelerated by the presence of water (or other polar solvents like dimethylformamide), suggesting that such species can enhance degradation by enhancing ion mobility. Barbé et al.[26] observed patterns of degradation near device edges similar to those in **Figure 2** when biasing them at 1 V in the dark under atmospheres containing different combinations of O_2 and H_2O. They reported that significant device degradation occurred only when H_2O was present and thus concluded that humidity, not oxygen, was the predominant cause of device degradation due to its ability to facilitate ion drift. However, since the energy bands in a ~1.6 eV bandgap solar cell under 1 V bias should be nearly flat, as this condition is generally close to open-circuit, there shouldn't be significant internal electric fields to drive ion migration. As a result, bias-induced degradation does not seem like the most likely explanation for increased decomposition near device edges. However, even in the reduced field state at maximum power point, the *gradient* of the electric field between the device region and the adjacent exposed regions may be considerable. Under an electric field gradient, dipoles such as MA^+ cations will experience a force that may also lead to decomposition of the material. We emphasize that although the observed effect of this proposed mechanism is the same as that proposed by Leijtens et al.[9] – i.e., local degradation through depletion of MA^+ cations – we propose that the physical origin of the force is an electric field *gradient* acting on *dipoles* ($\mathbf{F} = (\mathbf{p} \cdot \nabla)\boldsymbol{\mathcal{E}}$) rather the electric field *itself* acting on *charges* ($\mathbf{F} = q\boldsymbol{\mathcal{E}}$; here, $\mathbf{F}$ is the electric force acting on a species either of electric charge q or electric dipole moment $\mathbf{p}$ as a result of electric field $\boldsymbol{\mathcal{E}}$). Since the gradient is localized at the device edges, this additional stress may account for the rapid degradation there.

Macroscopic defects in the device interior where Ag coverage is low may effectively act as device edges as well, explaining why degradation fronts from the outer device boundary and interior defects propagate in similar ways. Degradation caused by the field gradient may further disrupt the integrity of the device, making it easier for O_2 and H_2O to penetrate the absorber; thus, these multiple modes of degradation may have a mutual accelerating effect. In a similar experiment to that shown in **Figure 2** conducted at the same conditions (25 °C/50% RH), when the edges of the device are covered with Kapton tape but the center remains exposed, the dark field images show that degradation is partially but not entirely suppressed in the protected region, but occurs at similar rates in the center (**Figure S4**). Comparing this experiment against the one shown in **Figure 2**, the initial decline in J_{sc} is significantly slowed by the partial protection. The fact that degradation can still be observed at the Kapton-protected edges points to the existence of a mode that does not require environmental stimuli (e.g., electric gradient-induced degradation), but the suppression in its rate again indicates that photooxidation also plays a significant role. The overall picture of hypothetical degradation mechanisms is summarized in **Figure 2h-k**: photooxidation is the major cause degradation outside the Ag contact but also contributes to degradation at the device edges and macroscopic defects in the device interior, where H_2O and O_2 can most easily diffuse from areas unprotected by the contact; chemical reactions between decomposition products from the perovskite and Ag from the contact may occur within the device; and electric field gradients at the device edges can exert forces on dipolar species such as MA^+ cations in the perovskite, leading to field-induced decomposition even in the absence of other external stresses.

While the effects of material decomposition on J_{sc} are a major factor in how the PCE evolves over time, fill factor also plays an important role in the critical period leading up to T_{80}. Fill factor is generally interpreted as being impacted by three major processes: series resistance, shunting, and recombination. To determine which are dominant in our experiments (and how these might vary with environmental conditions), we examine the change in series resistance R_s and shunt resistance R_{sh} at T_{80} relative to their starting values (both values are estimated from fits of the reverse-scan J-V sweeps to a diode equation based on the Lambert W-function).[27–29] Starting values of R_s range from 4 to 100 Ω with an average of 36 Ω; the changes in R_s at T_{80} range from -46 to +131 Ω, with an average increase of 17 Ω, and the relative change $\Delta R_s/R_s(t=0)$ ranges from -87% to +565%, with an average relative increase of +65%. In general, although series resistance may increase or decrease by nearly an order of magnitude at T_{80} (**Figure 3**), there does not appear to be a consistent tendency towards either. By contrast, R_{sh} almost always decreases (starting values range from 6.42×10^3 to 1.30×10^6 Ω with an average of 9.56×10^4 Ω; absolute changes from -1.30×10^6 to $+3.30 \times 10^3$ Ω with an average of -8.86×10^4 Ω; relative changes from -99.6% to +40.7% with an average of -64.9%), and in the few cases in which it does not decrease, the increase is marginal (less than 50%). Thus, shunting can be a significant contributor to fill factor loss. We note, however, that only a minority of devices we tested shunted catastrophically (as evidenced by linear, or nearly linear J-V curves), and they are excluded from this analysis on the likelihood that this failure mode indicates significant fabrication defects unrepresentative of well-made devices. In addition to shunting, recombination also likely plays a role in the decline of fill factor, particularly at higher temperatures. We can estimate the influence of recombination by examining the behavior of V_{oc}. There is a clear trend of reduced $V_{oc}(T_{80})/V_{oc}(t = 0)$ with increasing temperature (**Figure 3**), signifying that physical mechanisms that increase carrier recombination over time are activated by heat. A similar trend in $FF(T_{80})/FF(t = 0)$ indicates that thermally-activated recombination processes constraining V_{oc} may constrain FF as well (**Figure 3**). Similar analysis shows that the tendency to shunt is not strongly affected by temperature, while the increase in series resistance at T_{80} is in general much lower at higher temperatures (**Figure 3**). Plots of normalized J_{sc}, V_{oc}, and FF at T_{80} against relative humidity (**Figure S5**) display no obvious correlations, indicating that the same H_2O-mediated degradation mode or modes are active under all conditions probed in this study. In summary, fill factor is predominantly affected by shunting (regardless of temperature) and recombination (at higher temperatures), and the influence of the latter is also evident in the behavior of V_{oc}. The physical origin of temperature-activated recombination is unclear, but it may relate to interdiffusion of device components that lead to defects at interfaces or in the bulk of the perovskite.

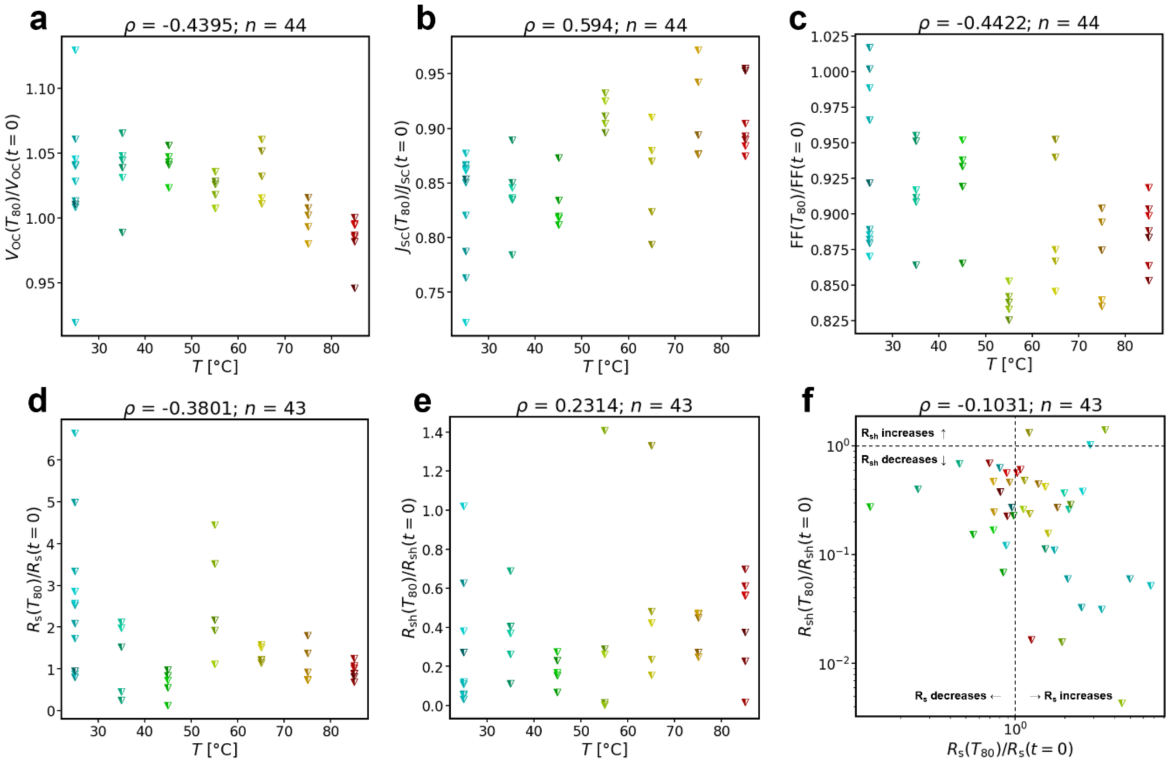

Figure 3. Relative changes in device performance parameters (parameter at $t = T_{80}$ compared to its initial value). (a-e) Scatterplots of relative changes in device parameters as a function of temperature: (a) V_{oc}, (b) J_{sc}, (c) fill factor, (d) series resistance, and (e) shunt resistance. (f) series and shunt resistance, relative to their initial values, plotted against one another. The Pearson correlation ρ between each variable pair and number n of samples plotted are given in the headings. In general, J_{sc} losses have a more pronounced effect on PCE at lower temperatures, while fill factor and to a lesser extent V_{oc} losses become more important as temperature rises. Series resistance effects on fill factor tend to be higher with increased temperature, while shunting effects appear to be insensitive to it. Overall, while series resistance may have increased or decreased at T_{80}, it does not do so in a consistent manner; by contrast, the shunt resistance almost always decreases. Environmental conditions are indicated according to the symbol legend in **Figure 2**.

Development of Machine Learning Models to Predict T_{80}. Having developed a basic understanding of how $CH_3NH_3PbI_3$ solar cell degradation proceeds in different environments, we now turn to the task of developing machine learning models to forecast the evolution of their power conversion efficiency (PCE). Predicting a variable such as T_{80} falls under the category of supervised learning: each degradation experiment is "labeled" by the value of T_{80}, and the objective of the machine learning algorithm is to discover a mathematical relationship between the labels and other data characteristic of each experiment, termed "features." To maximize the models' predictive utility, the features should be calculated from data measured during the early stages of degradation. The data that are available depends on how the experiment is set up and conducted, and might in principle include J-V measurements, environmental conditions, device architecture and processing, photoluminescence or dark field measurements, capacitance spectroscopy, or any other data available from device or film characterization. Which of these data to include in the feature set is a critical decision. On one hand, more extensive feature sets improve the chances of obtaining high predictive accuracy by incorporating as many potentially relevant effects into the model as possible. On the other hand, incorporating too many variables may needlessly increase the experimental burden of data acquisition, and reduce the model's general applicability if it relies on data from techniques that are not readily available to the PV research community. Moreover, models that attempt to incorporate too many features are vulnerable to overfitting—that is, learning noise, rather than the true patterns in the dataset—and generalize poorly when applied to new data beyond the training set. Therefore, we focus on two major classes of features available from instruments that are relatively inexpensive and ubiquitous in photovoltaics research laboratories: *a priori* data that are known from the environmental conditions the solar cells are subjected to, and sample-specific measurements based on *J-V* measurements made during the first few cycles of data acquisition. Explicitly, the *a priori* variables are temperature, partial pressure of H_2O, and the natural logarithm of the kinetically modeled $CH_3NH_3PbI_3$ decomposition rate determined from the ambient environmental conditions, as we have recently reported elsewhere.[6] Features constructed from *J-V* measurements include the initial values of J_{sc}, V_{oc}, and fill factor, as well as the first and second time derivatives of each parameter at the start of each degradation experiment normalized to its starting values. Between the *a priori* and sample-specific variables, this construction yields a total of 12 features, summarized explicitly in **Table 1**.

Table 1. Features provided to machine learning models that predict perovskite solar cell T_{80}, and how they are calculated.

feature symbol	units	physical interpretation/calculation method
r_{MAPI}	$\mathrm{mol \cdot m^{-2} \cdot s^{-1}}$	Decomposition rate of $CH_3NH_3PbI_3$ film under the specified environmental conditions; calculated from kinetic model described in Siegler et al.[6]
T	°C	Solar cell temperature, controlled during the experiment
$P_{\mathrm{H_2O}}$	kPa	Partial pressure of ambient H_2O, calculated from relative humidity measurement
$J_{\mathrm{sc}}(t=0)$	$\mathrm{mA \cdot cm^{-2}}$	Short-circuit current at the start of the experiment, taken from steady state measurements at short circuit
$V_{\mathrm{oc}}(t=0)$	V	Open-circuit current at the start of the experiment, taken from steady state measurements at open circuit

$\mathrm{FF}(t=0)$	%	Fill factor at the start of the experiment, taken from steady state measurements at open circuit, short circuit, and maximum power point
$\mathrm{d}J_{\mathrm{sc}}/\mathrm{d}t$	min^{-1}	1st time derivative of the *normalized* short-circuit current (by its initial value), estimated from the first 90 minutes of each experiment
$\mathrm{d}V_{\mathrm{oc}}/\mathrm{d}t$	min^{-1}	1st time derivative of the *normalized* open-circuit voltage (by its initial value), estimated from the first 90 minutes of each experiment
$\mathrm{dFF}/\mathrm{d}t$	min^{-1}	1st time derivative of the *normalized* fill factor (by its initial value), estimated from the first 90 minutes of each experiment
$\mathrm{d}^2J_{\mathrm{sc}}/\mathrm{d}t^2$	min^{-2}	2nd time derivative of the *normalized* short-circuit current (by its initial value), estimated from the first 90 minutes of each experiment
$\mathrm{d}^2V_{\mathrm{oc}}/\mathrm{d}t^2$	min^{-2}	2nd time derivative of the *normalized* open-circuit voltage (by its initial value), estimated from the first 90 minutes of each experiment
$\mathrm{d}^2\mathrm{FF}/\mathrm{d}t^2$	min^{-2}	2nd time derivative of the *normalized* fill factor (by its initial value), estimated from the first 90 minutes of each experiment

Relative to the size of the dataset (45 runs), the feature set is still large enough to pose a risk of overfitting. To avoid this, we employ modeling techniques that enforce sparsity of the dataset, including linear regression with greedy feature selection (GFS) by orthogonal matching pursuit,[30] LASSO,[31] and ridge regression.[32,33] In all of these models, the natural logarithm of T_{80} is expressed as a linear combination of the features, but the methods for determining the coefficients of each feature are different. In greedy feature selection, features are selected sequentially based on which one most reduces the error of the prior model (starting from a model that includes no features at all); the search is terminated before the number of selected features exceeds 10% of the number of features in the training set (note that it is also possible to terminate the search using error-based criteria, although doing so may not enforce sparsity as stringently as constraints based on the size of the feature set). LASSO and ridge regression assign coefficients by attempting to simultaneously minimize the least squares error of the regression in addition to a penalty term that is proportional to the ℓ_1- (LASSO) or ℓ_2-norm (ridge) of the vector of feature weights. With LASSO, insignificant features are often assigned a weight of precisely zero, strictly enforcing sparsity; with ridge, the weights of insignificant features are suppressed but do not vanish entirely. The models are tested through leave-one-out cross-validation – that is, each experiment in the dataset is sequentially removed as a test sample, and the remaining data are used to train the model. Statistics from the distribution of testing error furnish an unbiased estimate of the model's predictive accuracy on unseen data, representing its ultimate figure of merit. Stability of the models relative to the training set may be assessed by comparing the feature weights across iterations of the test/train split: stable models will repeatably select the same features and assign weights with consistent magnitudes, while feature weights in unstable models may fluctuate considerably.

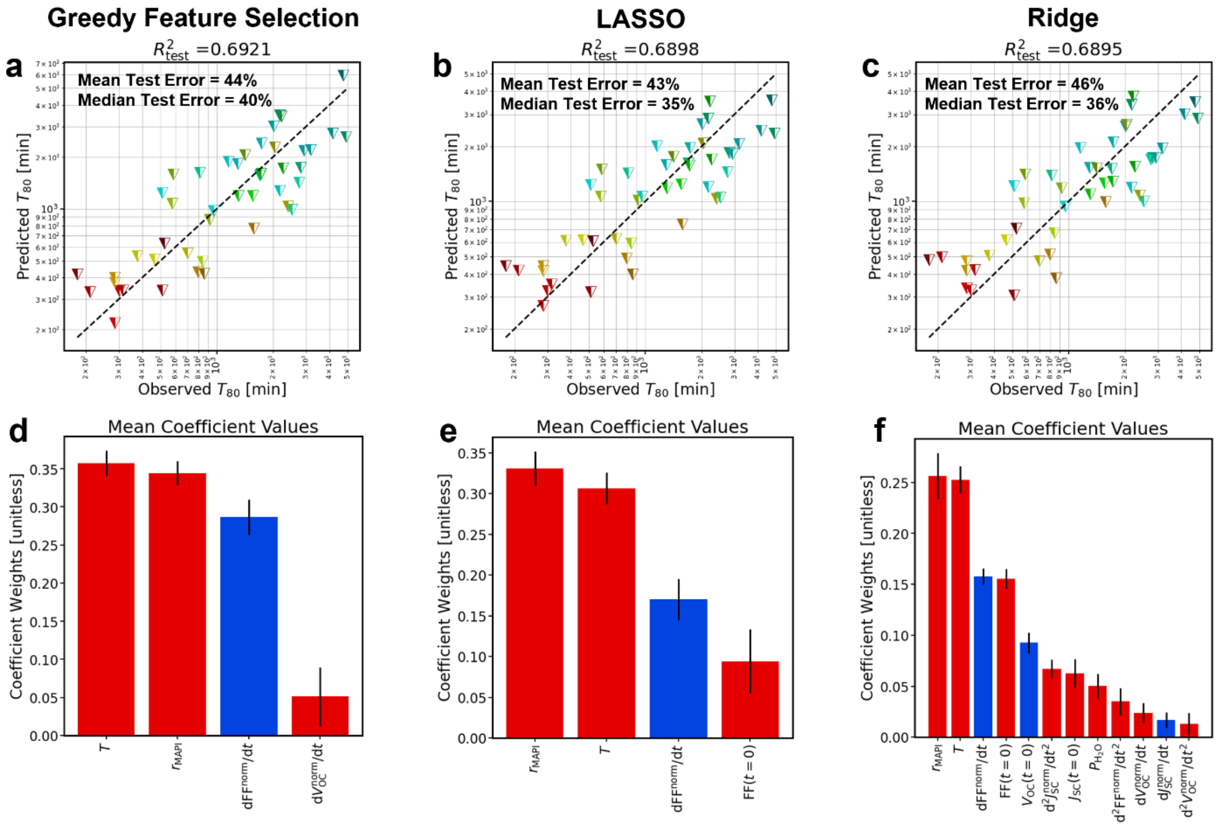

Figure 4. Modeling results for greedy feature selection (a,d), LASSO (b,e), and ridge regression (c,f) with leave-one-out testing. (a-c) Parity plots show that predictions on samples withheld in leave-one-out prediction are in fairly good agreement with the observed values, with average error generally in the range of 40-50%, and that error metrics are relatively close to one another across modeling approaches. (d-f) Bar plots of feature coefficients show that the models also tend to agree on which features are most important. That is, those selected by sparsity-enforcing models (greedy feature selection, LASSO) comprise the top 5 features in the model trained using ridge regression. For clarity, only features for which the mean value is larger (in absolute value) than the standard deviation across all test/train splits are shown; a more complete breakdown of the coefficients is given in **Figures S6-S8.** The standard deviation of coefficient values across all test/train splits is represented by the black bars in (d-f). Environmental conditions are indicated according to the symbol legend in Figure 2.

Models trained using GFS, LASSO, and ridge are relatively consistent with one another. Parity plots of test set predictions (**Figure 4a-c**) show that average prediction accuracy for all models lies in the range of 35-45% while the models themselves are structured similarly. The mean test error is almost the same for each model, while the median is slightly lower for LASSO and ridge than for GFS. The R^2 values for the test set predictions are also similar at 0.69, showing that the models can explain about 2/3 of the variance in unseen samples. Bar plots of the feature weights (**Figure 4d-f**) show that similar features are typically assigned high weights across all three modeling algorithms despite the differences in how they are selected. The models trained by GFS and LASSO are considerably sparser than those trained by ridge regression yet retain comparable predictive accuracy, indicating that most of the features selected by the latter are superfluous and do not meaningfully contribute to prediction. We therefore focus our discussion of model interpretation below on the features selected by GFS. The features selected with mean value greater than their standard deviation across all test-train splits are almost identical for GFS and LASSO (which is a good sign). A full breakdown of the coefficients for each test-train split is given in **Figures S6-S8**.

Interpretation of the Model Predictions. Explicitly, the model trained by GFS (using the values obtained for the test/train split with median error) may be written as:

$$\ln(T_{80}) = c + \beta_1 \frac{r_{\mathrm{MAPI}} - \overline{r_{\mathrm{MAPI}}}}{\sigma_r} + \beta_2 \frac{T - \overline{T}}{\sigma_T} + \beta_3 \frac{\left(\frac{\mathrm{dFF}}{\mathrm{d}t}\right)_{t=0} - \overline{\left(\frac{\mathrm{dFF}}{\mathrm{d}t}\right)}_{t=0}}{\sigma_F} + \beta_4 \frac{\left(\frac{\mathrm{d}V_{\mathrm{OC}}}{\mathrm{d}t}\right)_{t=0} - \overline{\left(\frac{\mathrm{d}V_{\mathrm{OC}}}{\mathrm{d}t}\right)}_{t=0}}{\sigma_V}$$

Here, the terms β_i are the feature coefficients learned by the model; the constant c is the intercept, also learned by the model; and the terms under the bars and σ_i are the means and standard deviations resulting from feature standardization. The standardization parameters are not learned by the model, but are introduced before training to put all features on statistically equivalent (mean of zero and standard deviation of 1) and unitless footing. Coefficient values are provided in **Table 2**.

Table 2. Parameters of the model trained by greedy feature selection corresponding to the test/train split with median testing error.

$$\ln(T_{80}) = c + \sum_{i=1}^{p} \beta_i \frac{x_i - \mu_i}{\sigma_i}$$

Feature x_i	**Units**	**Coefficient β_i**	**Mean μ_i**	**Std. Dev. σ_i**
Intercept c	ln(min)	6.967	--	--
r_{MAPI}	mol·m^{-2}·s^{-1}	-0.3473	1.218×10^{-7}	3.731×10^{-8}
T	°C	-0.3662	52.05	21.69
$\mathrm{dFF}/\mathrm{d}t$	min^{-1}	0.2969	1.848×10^{-4}	7.515×10^{-4}

$\mathrm{d}V_{\mathrm{oc}}/\mathrm{d}t$	min^{-1}	-0.08609	8.245×10^{-5}	1.708×10^{-4}

The most consistently important features in all models described above are the kinetically modeled $CH_3NH_3PbI_3$ decomposition rate, temperature, and the first derivative of fill factor with respect to time. The former two are negatively related to T_{80}, while the latter is positively related. The major role of the kinetically modeled $CH_3NH_3PbI_3$ decomposition rate r_{MAPI} reinforces the observation that photooxidative absorber degradation is an important constraint on solar cell lifetime. As noted above, however, there are likely other processes besides photooxidation that contribute to the observed spatial patterns of degradation that may be specific to the device architecture. The similarly large influence of temperature is in accord with the observed increases in fill factor and V_{oc} losses under heating noted previously, embodying the effects of physical processes that increase recombination (or otherwise compromise performance). There are many possible thermally activated processes that might do so. NiO_x has also been shown to undergo a thermally-induced reaction with MAI, compromising performance of $CH_3NH_3PbI_3$ PSCs processed under short high temperature anneals;[34] it is possible that this reaction may occur at lower temperatures, albeit at a slower rate that nevertheless can compromise device performance over longer periods of operation at more moderate temperatures. Numerous studies have reported that diffusion of other structural components can occur rapidly in perovskite solar cells: In^{3+} from ITO;[34,35] contact metals such as Au;[36,37] and even Na^+, Ca^{2+}, and other components of soda-lime glass substrates.[38,39] Although halide perovskites are generally tolerant of metal ion impurities, sufficient levels may still lead to deterioration of their optoelectronic properties. In view of the many possible avenues by which temperature-activated processes could occur, we take no position on its precise origins here, but note that future generations of predictive models stand to benefit from detailed studies of the relevant physical processes.

The initial rate of fill factor rise is the only sample-specific feature consistently assigned high weight in the models. At lower temperatures, below 55 °C, the fill factor initially increases before settling into its longer-term decline (as seen in **Figure 1b**). At higher temperatures (≥65 °C), an initial rise in FF is rarely observed, and there is an overall negative linear correlation between dFF/dt and temperature (**Figure 5a**). However, the initial value of the fill factor also increases with temperature (**Figure 5b**), indicating that the initial rise observed at lower temperature may also occur at higher ones, but much faster, within the first data acquisition cycle. This initial rise correlates very weakly with initial changes in shunt resistance and modestly with series resistance (**Figure 5c,d**). These results suggest that multiple thermally activated processes influence fill factor: a beneficial process acting on short-moderate timescales that increases it, and a longer-timescale process that deteriorates it along with V_{oc}, as discussed above. In the model, the fill factor time derivative captures the shorter-timescale beneficial processes (or more accurately, the net effect of the beneficial processes and the detrimental ones), while temperature captures the longer-term effects of the adverse ones. We consider two hypotheses for the origin of processes that initially benefit the fill factor: either they are related to ion redistribution as a result of the device transitioning from equilibrium in the dark (its storage state) to maximum power point under illumination (its main operating condition); or they may reflect irreversible improvements associated with elimination of trap states in the bulk or at interfaces. In the former case, when the device is in the dark, mobile ions (e.g., iodine or methylammonium vacancies) will drift under the built-in electric field and accumulate at the contacts until the electrochemical potential associated with their buildup is strong enough to resist further drift. Domanski et al.[40] investigated transient changes in perovskite solar cell power output and concluded that different ionic species drift on different timescales, with halide vacancies moving relatively quickly, on the order of 0.1-100 s, and cation vacancies moving much more slowly, on the order of >1000 s. This difference in ionic mobility implies that halide vacancies are responsible for J-V hysteresis, while cation vacancies are more likely to account for longer-term evolution in device performance. When the device is measured, it spends most of its time operating near the maximum power point—i.e., at relatively high forward bias. When placed under these conditions, the applied bias acts in opposition to the built-in electric field, and thus the ions will no longer be pushed as strongly toward the contacts as they were in the dark. As they reestablish a new equilibrium, the perovskite energy bands will also shift in response to the changing charge distribution. Reconfiguration of the bands in this manner may also change the relative positions of the Fermi level and midgap trap states, with the possibility of these states shifting from mostly unfilled (active) to mostly filled (inactive) and thus resulting in reduced recombination. Band bending due to ion accumulation at the interfaces driven by the built-in field may also create charge extraction barriers there. When the device shifts from its dark equilibrium state to maximum power point, these barriers may be alleviated as mobile ions diffuse back into the bulk. To test whether ion migration plays a role in fill factor improvement, we assess whether these effects are reversible when the device is transitioned between periods of operation under illumination at maximum power point (when the electric field is low) and in the dark at short circuit (when it is high). The results of this experiment indicate that the initial FF improvement is partially reversible and strongly correlated with series resistance. (**Figure S9**). We note also that the rise in fill factor over the first light cycle is anti-correlated with the change in V_{oc}, suggesting that its initial enhancement occurs in spite of increased carrier recombination rather than due to a reduction in defect activity and implying that interface barrier alleviation may be the best explanation. However, during successive cycles FF and V_{oc} display a similar upward trend, indicating that increases following the initial rise may have a different physical origin such as passivation of trap states by O_2 or H_2O. After several light/dark cycles, the FF improvement is maintained across the dark periods (i.e., appears to become less reversible). This may indicate that drift of ions in the dark is slower than their diffusion in the light, allowing performance gains to build up over time with a 50% light/dark duty cycle. Hysteresis in the device begins low and remains so over the course of the experiment, indicating that there are no major changes in the mobile ion population. In view of the partially reversible nature of the initial fill factor improvement and its clear connection with series resistance, we therefore believe it can be plausibly explained by interfacial band reconfiguration due to ion migration.

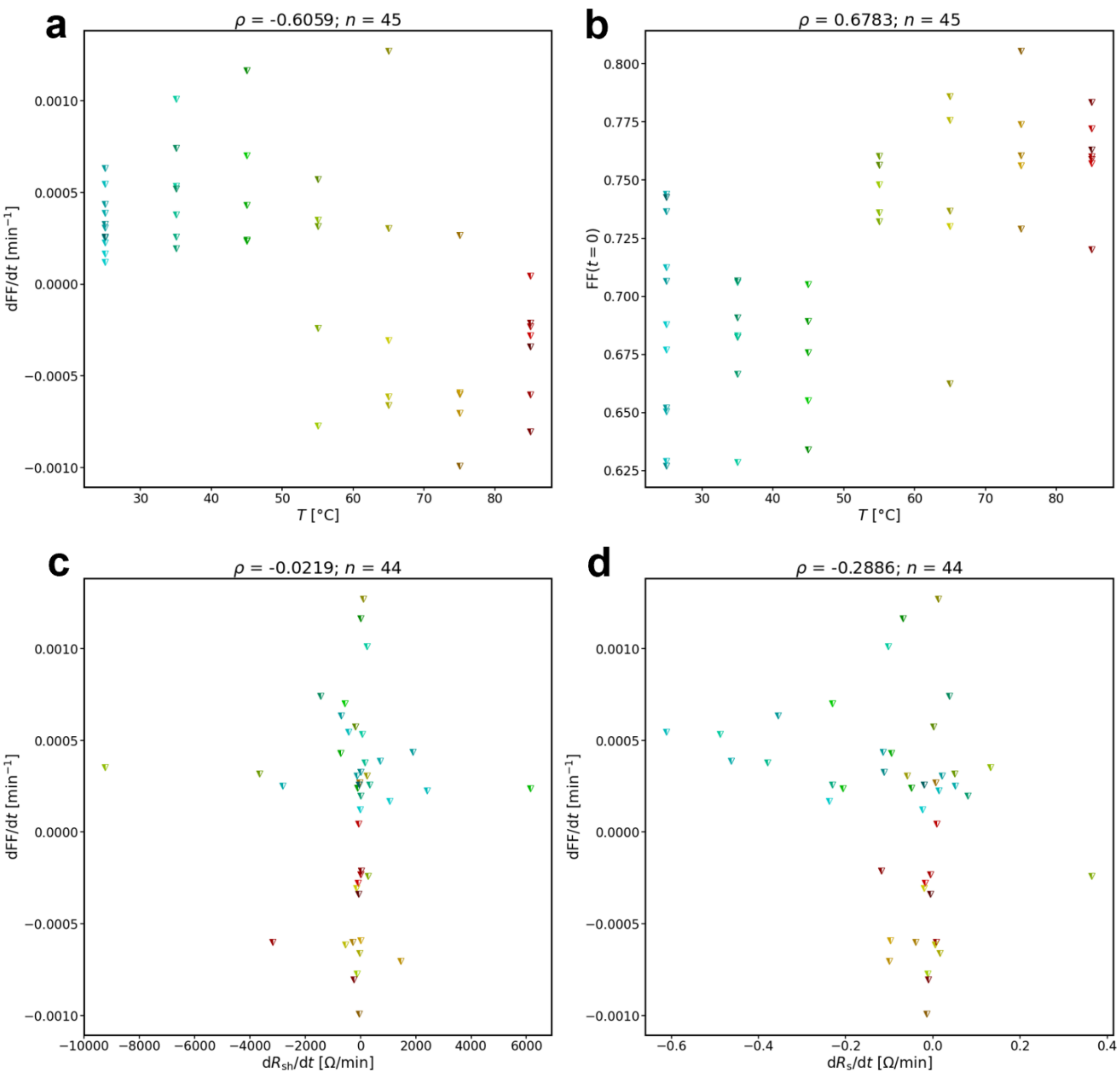

Figure 5. (a) Scatterplot of initial 1st time derivative of fill factor against temperature. (b) Scatterplot of initial fill factor against temperature. (c) Scatterplot of initial 1st time derivative of fill factor against initial 1st time derivative of shunt resistance. (d) Scatterplot of initial 1st time derivative of fill factor against initial 1st time derivative of series resistance. The Pearson correlation ρ between each variable pair and number n of samples plotted are given in the headings.

Conclusions. In this work, we have experimentally examined the degradation of 45 ITO/NiO$_x$/CH$_3$NH$_3$PbI$_3$/C$_{60}$/BCP/Ag solar cells under a wide range of environmental conditions. We find that power conversion efficiency tends to decline mostly because of losses in short-circuit current and fill factor, while open-circuit voltage typically remains high over the useful lifetime of the devices. Short-circuit current losses are strongly associated with water-accelerated photooxidation and electric field-related decomposition of the perovskite absorber (via electric field gradients). Fill factor losses are almost universally attributable at least in part to reduction in shunt resistance, but increases in recombination play a larger role at higher temperatures, at which open-circuit voltage also starts to become affected. Machine learning models trained to predict T_{80} have accuracy of 35-45% on average, and can attain this level of performance using sparse feature sets relying on the kinetically modeled $CH_3NH_3PbI_3$ decomposition rate, the temperature, and the initial time derivative of the device's fill factor. Both our analysis of the data and the model's choice of features suggests that decomposition of the perovskite plays a large role in the gradual loss of power conversion efficiency by way of short-circuit current reduction, but additional thermally-activated processes likely contribute to increased recombination over time, contributing fill factor and V_{oc} losses (particularly under higher temperatures). This work demonstrates how incorporating physical knowledge of the processes constraining device performance can contribute to ML model accuracy when the size of the dataset is restricted by allowing the development of highly informative features. The success of our models should motivate future efforts to develop quantitative understanding of decomposition processes of other perovskites of interest (particularly compositions rich in formamidinium), as well as interlayer device interactions that may lead to increased carrier recombination rates.

References.

(1) Dualeh, A.; Gao, P.; Seok, S. I.; Nazeeruddin, M. K.; Grätzel, M. Thermal Behavior of Methylammonium Lead-Trihalide Perovskite Photovoltaic Light Harvesters. *Chem. Mater.* **2014**, *26* (21), 6160–6164. https://doi.org/10.1021/cm502468k.

(2) Aristidou, N.; Sanchez-Molina, I.; Chotchuangchutchaval, T.; Brown, M.; Martinez, L.; Rath, T.; Haque, S. A. The Role of Oxygen in the Degradation of Methylammonium Lead Trihalide Perovskite Photoactive Layers. *Angew. Chem.* **2015**, *127* (28), 8326–8330. https://doi.org/10.1002/ange.201503153.

(3) Akbulatov, A. F.; Ustinova, M. I.; Shilov, G. V.; Dremova, N. N.; Zhidkov, I. S.; Kurmaev, E. Z.; Frolova, L. A.; Shestakov, A. F.; Aldoshin, S. M.; Troshin, P. A. Temperature Dynamics of $MAPbI_3$ and PbI_2 Photolysis: Revealing the Interplay between Light and Heat, Two Enemies of Perovskite Photovoltaics. *J. Phys. Chem. Lett.* **2021**, *12* (18), 4362–4367. https://doi.org/10.1021/acs.jpclett.1c00883.

(4) Leguy, A. M. A.; Hu, Y.; Campoy-Quiles, M.; Alonso, M. I.; Weber, O. J.; Azarhoosh, P.; van Schilfgaarde, M.; Weller, M. T.; Bein, T.; Nelson, J.; Docampo, P.; Barnes, P. R. F. Reversible Hydration of $CH_3NH_3PbI_3$ in Films, Single Crystals, and Solar Cells. *Chem. Mater.* **2015**, *27* (9), 3397–3407. https://doi.org/10.1021/acs.chemmater.5b00660.

(5) Christians, J. A.; Miranda Herrera, P. A.; Kamat, P. V. Transformation of the Excited State and Photovoltaic Efficiency of $CH_3NH_3PbI_3$ Perovskite upon Controlled Exposure to Humidified Air. *J. Am. Chem. Soc.* **2015**, *137* (4), 1530–1538. https://doi.org/10.1021/ja511132a.

(6) Siegler, T. D.; Dunlap-Shohl, W. A.; Meng, Y.; Yang, Y.; Kau, W. F.; Sunkari, P. P.; Tsai, C. E.; Armstrong, Z. J.; Chen, Y.-C.; Beck, D. A. C.; Meilă, M.; Hillhouse, H. W. Water-Accelerated Photooxidation of $CH_3NH_3PbI_3$ Perovskite. *J. Am. Chem. Soc.* **2022**, jacs.2c00391. https://doi.org/10.1021/jacs.2c00391.

(7) Bowring, A. R.; Bertoluzzi, L.; O'Regan, B. C.; McGehee, M. D. Reverse Bias Behavior of Halide Perovskite Solar Cells. *Adv. Energy Mater.* **2018**, *8* (8), 1702365. https://doi.org/10.1002/aenm.201702365.

(8) Rajagopal, A.; Williams, S. T.; Chueh, C.-C.; Jen, A. K.-Y. Abnormal Current–Voltage Hysteresis Induced by Reverse Bias in Organic–Inorganic Hybrid Perovskite Photovoltaics. *J. Phys. Chem. Lett.* **2016**, *7* (6), 995–1003. https://doi.org/10.1021/acs.jpclett.6b00058.

(9) Leijtens, T.; Hoke, E. T.; Grancini, G.; Slotcavage, D. J.; Eperon, G. E.; Ball, J. M.; De Bastiani, M.; Bowring, A. R.; Martino, N.; Wojciechowski, K.; McGehee, M. D.; Snaith, H. J.; Petrozza, A. Mapping Electric Field-Induced Switchable Poling and Structural Degradation in Hybrid Lead Halide Perovskite Thin Films. *Adv. Energy Mater.* **2015**, *5* (20), 1500962. https://doi.org/10.1002/aenm.201500962.

(10) Khenkin, M. V.; K. M., A.; Katz, E. A.; Visoly-Fisher, I. Bias-Dependent Degradation of Various Solar Cells: Lessons for Stability of Perovskite Photovoltaics. *Energy Environ. Sci.* **2019**, *12* (2), 550–558. https://doi.org/10.1039/C8EE03475C.

(11) Bush, K. A.; Palmstrom, A. F.; Yu, Z. J.; Boccard, M.; Cheacharoen, R.; Mailoa, J. P.; McMeekin, D. P.; Hoye, R. L. Z.; Bailie, C. D.; Leijtens, T.; Peters, I. M.; Minichetti, M. C.; Rolston, N.; Prasanna, R.; Sofia, S.; Harwood, D.; Ma, W.; Moghadam, F.; Snaith, H. J.; Buonassisi, T.; Holman, Z. C.; Bent, S. F.; McGehee, M. D. 23.6%-Efficient Monolithic Perovskite/Silicon Tandem Solar Cells with Improved Stability. *Nat. Energy* **2017**, *2* (4), 17009. https://doi.org/10.1038/nenergy.2017.9.

(12) Shi, L.; Bucknall, M. P.; Young, T. L.; Zhang, M.; Hu, L.; Bing, J.; Lee, D. S.; Kim, J.; Wu, T.; Takamure, N.; McKenzie, D. R.; Huang, S.; Green, M. A.; Ho-Baillie, A. W. Y. Gas Chromatography–Mass Spectrometry Analyses of Encapsulated Stable Perovskite Solar Cells. *Science* **2020**, *368* (6497), eaba2412. https://doi.org/10.1126/science.aba2412.

(13) Jang, Y.-W.; Lee, S.; Yeom, K. M.; Jeong, K.; Choi, K.; Choi, M.; Noh, J. H. Intact 2D/3D Halide Junction Perovskite Solar Cells via Solid-Phase in-Plane Growth. *Nat. Energy* **2021**, *6* (1), 63–71. https://doi.org/10.1038/s41560-020-00749-7.

(14) Shi, L.; Young, T. L.; Kim, J.; Sheng, Y.; Wang, L.; Chen, Y.; Feng, Z.; Keevers, M. J.; Hao, X.; Verlinden, P. J.; Green, M. A.; Ho-Baillie, A. W. Y. Accelerated Lifetime Testing of Organic–Inorganic Perovskite Solar Cells Encapsulated by Polyisobutylene. *ACS Appl. Mater. Interfaces* **2017**, *9* (30), 25073–25081. https://doi.org/10.1021/acsami.7b07625.

(15) Khenkin, M. V.; Katz, E. A.; Abate, A.; Bardizza, G.; Berry, J. J.; Brabec, C.; Brunetti, F.; Bulović, V.; Burlingame, Q.; Di Carlo, A.; Cheacharoen, R.; Cheng, Y.-B.; Colsmann, A.; Cros, S.; Domanski, K.; Dusza, M.; Fell, C. J.; Forrest, S. R.; Galagan, Y.; Di Girolamo, D.; Grätzel, M.; Hagfeldt, A.; von Hauff, E.; Hoppe, H.; Kettle, J.; Köbler, H.; Leite, M. S.; Liu, S.; Loo, Y.-L.; Luther, J. M.; Ma, C.-Q.; Madsen, M.; Manceau, M.; Matheron, M.; McGehee, M.; Meitzner, R.; Nazeeruddin, M. K.; Nogueira, A. F.; Odabaşı, Ç.; Osherov, A.; Park, N.-G.; Reese, M. O.; De Rossi, F.; Saliba, M.; Schubert, U. S.; Snaith, H. J.; Stranks, S. D.; Tress, W.; Troshin, P. A.; Turkovic, V.; Veenstra, S.; Visoly-Fisher, I.; Walsh, A.; Watson, T.; Xie, H.; Yıldırım, R.; Zakeeruddin, S. M.; Zhu, K.; Lira-Cantu, M. Consensus Statement for Stability Assessment and Reporting for Perovskite Photovoltaics Based on ISOS Procedures. *Nat. Energy* **2020**, *5* (1), 35–49. https://doi.org/10.1038/s41560-019-0529-5.

(16) Jacobsson, T. J.; Hultqvist, A.; García-Fernández, A.; Anand, A.; Al-Ashouri, A.; Hagfeldt, A.; Crovetto, A.; Abate, A.; Ricciardulli, A. G.; Vijayan, A.; Kulkarni, A.; Anderson, A. Y.; Darwich, B.

P.; Yang, B.; Coles, B. L.; Perini, C. A. R.; Rehermann, C.; Ramirez, D.; Fairen-Jimenez, D.; Di Girolamo, D.; Jia, D.; Avila, E.; Juarez-Perez, E. J.; Baumann, F.; Mathies, F.; González, G. S. A.; Boschloo, G.; Nasti, G.; Paramasivam, G.; Martínez-Denegri, G.; Näsström, H.; Michaels, H.; Köbler, H.; Wu, H.; Benesperi, I.; Dar, M. I.; Bayrak Pehlivan, I.; Gould, I. E.; Vagott, J. N.; Dagar, J.; Kettle, J.; Yang, J.; Li, J.; Smith, J. A.; Pascual, J.; Jerónimo-Rendón, J. J.; Montoya, J. F.; Correa-Baena, J.-P.; Qiu, J.; Wang, J.; Sveinbjörnsson, K.; Hirselandt, K.; Dey, K.; Frohna, K.; Mathies, L.; Castriotta, L. A.; Aldamasy, Mahmoud. H.; Vasquez-Montoya, M.; Ruiz-Preciado, M. A.; Flatken, M. A.; Khenkin, M. V.; Grischek, M.; Kedia, M.; Saliba, M.; Anaya, M.; Veldhoen, M.; Arora, N.; Shargaieva, O.; Maus, O.; Game, O. S.; Yudilevich, O.; Fassl, P.; Zhou, Q.; Betancur, R.; Munir, R.; Patidar, R.; Stranks, S. D.; Alam, S.; Kar, S.; Unold, T.; Abzieher, T.; Edvinsson, T.; David, T. W.; Paetzold, U. W.; Zia, W.; Fu, W.; Zuo, W.; Schröder, V. R. F.; Tress, W.; Zhang, X.; Chiang, Y.-H.; Iqbal, Z.; Xie, Z.; Unger, E. An Open-Access Database and Analysis Tool for Perovskite Solar Cells Based on the FAIR Data Principles. *Nat. Energy* **2022**, *7* (1), 107–115. https://doi.org/10.1038/s41560-021-00941-3.

(17) Howard, J. M.; Wang, Q.; Srivastava, M.; Gong, T.; Lee, E.; Abate, A.; Leite, M. S. Quantitative Predictions of Moisture-Driven Photoemission Dynamics in Metal Halide Perovskites via Machine Learning. *J. Phys. Chem. Lett.* **2022**, *13* (9), 2254–2263. https://doi.org/10.1021/acs.jpclett.2c00131.

(18) Srivastava, M.; Howard, J. M.; Gong, T.; Rebello Sousa Dias, M.; Leite, M. S. Machine Learning Roadmap for Perovskite Photovoltaics. *J. Phys. Chem. Lett.* **2021**, *12* (32), 7866–7877. https://doi.org/10.1021/acs.jpclett.1c01961.

(19) Hartono, N. T. P.; Thapa, J.; Tiihonen, A.; Oviedo, F.; Batali, C.; Yoo, J. J.; Liu, Z.; Li, R.; Marrón, D. F.; Bawendi, M. G.; Buonassisi, T.; Sun, S. How Machine Learning Can Help Select Capping Layers to Suppress Perovskite Degradation. *Nat. Commun.* **2020**, *11* (1), 4172. https://doi.org/10.1038/s41467-020-17945-4.

(20) Sun, S.; Tiihonen, A.; Oviedo, F.; Liu, Z.; Thapa, J.; Zhao, Y.; Hartono, N. T. P.; Goyal, A.; Heumueller, T.; Batali, C.; Encinas, A.; Yoo, J. J.; Li, R.; Ren, Z.; Peters, I. M.; Brabec, C. J.; Bawendi, M. G.; Stevanovic, V.; Fisher, J. I.; Buonassisi, T. A Data Fusion Approach to Optimize Compositional Stability of Halide Perovskites. *Matter* **2021**, *4* (4), 1305–1322. https://doi.org/10.1016/j.matt.2021.01.008.

(21) Stoddard, R. J.; Dunlap-Shohl, W. A.; Qiao, H.; Meng, Y.; Kau, W. F.; Hillhouse, H. W. Forecasting the Decay of Hybrid Perovskite Performance Using Optical Transmittance or Reflected Dark-Field Imaging. *ACS Energy Lett.* **2020**, *5* (3), 946–954. https://doi.org/10.1021/acsenergylett.0c00164.

(22) Stoddard, R. J.; Eickemeyer, F. T.; Katahara, J. K.; Hillhouse, H. W. Correlation between Photoluminescence and Carrier Transport and a Simple In Situ Passivation Method for High-Bandgap Hybrid Perovskites. *J. Phys. Chem. Lett.* **2017**, *8* (14), 3289–3298. https://doi.org/10.1021/acs.jpclett.7b01185.

(23) Lin, H.-J.; Cacovich, S.; Rebai, A.; Rousset, J.; Longeaud, C. Influence of Environment and Light-Stress on the Optoelectronic Properties of Triple-Cation Perovskite Thin Films. *ACS Appl. Mater. Interfaces* **2020**, *12* (17), 19495–19503. https://doi.org/10.1021/acsami.0c01732.

(24) Bryant, D.; Aristidou, N.; Pont, S.; Sanchez-Molina, I.; Chotchunangatchaval, T.; Wheeler, S.; Durrant, J. R.; Haque, S. A. Light and Oxygen Induced Degradation Limits the Operational Stability of Methylammonium Lead Triiodide Perovskite Solar Cells. *Energy Environ. Sci.* **2016**, *9* (5), 1655–1660. https://doi.org/10.1039/C6EE00409A.

(25) Besleaga, C.; Abramiuc, L. E.; Stancu, V.; Tomulescu, A. G.; Sima, M.; Trinca, L.; Plugaru, N.; Pintilie, L.; Nemnes, G. A.; Iliescu, M.; Svavarsson, H. G.; Manolescu, A.; Pintilie, I. Iodine Migration and Degradation of Perovskite Solar Cells Enhanced by Metallic Electrodes. *J. Phys. Chem. Lett.* **2016**, *7* (24), 5168–5175. https://doi.org/10.1021/acs.jpclett.6b02375.

(26) Barbé, J.; Kumar, V.; Newman, M. J.; Lee, H. K. H.; Jain, S. M.; Chen, H.; Charbonneau, C.; Rodenburg, C.; Tsoi, W. C. Dark Electrical Bias Effects on Moisture-Induced Degradation in Inverted Lead Halide Perovskite Solar Cells Measured by Using Advanced Chemical Probes. *Sustain. Energy Fuels* **2018**, *2* (4), 905–914. https://doi.org/10.1039/C7SE00545H.

(27) Jain, A.; Kapoor, A. A New Approach to Study Organic Solar Cell Using Lambert W-Function. *Sol. Energy Mater. Sol. Cells* **2005**, *86* (2), 197–205. https://doi.org/10.1016/j.solmat.2004.07.004.

(28) Zhang, C.; Zhang, J.; Hao, Y.; Lin, Z.; Zhu, C. A Simple and Efficient Solar Cell Parameter Extraction Method from a Single Current-Voltage Curve. *J. Appl. Phys.* **2011**, *110* (6), 064504. https://doi.org/10.1063/1.3632971.

(29) Díaz, S. R. A Generalized Theoretical Approach for Solar Cells Fill Factors by Using Shockley Diode Model and Lambert W-Function: A Review Comparing Theory and Experimental Data. *Phys. B Condens. Matter* **2022**, *624*, 413427. https://doi.org/10.1016/j.physb.2021.413427.

(30) Tropp, J. A.; Gilbert, A. C. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* **2007**, *53* (12), 4655–4666. https://doi.org/10.1109/TIT.2007.909108.

(31) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser. B Stat Methodol* **1996**, *58* (1), 267–288.

(32) Hoerl, A. E.; Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12* (1), 55–67.

(33) Hoerl, A. E.; Kennard, R. W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **1970**, *12* (1), 69–82.

(34) Dunlap-Shohl, W. A.; Li, T.; Mitzi, D. B. Interfacial Effects during Rapid Lamination within $MAPbI_3$ Thin Films and Solar Cells. *ACS Appl. Energy Mater.* **2019**, *2* (7), 5083–5093. https://doi.org/10.1021/acsaem.9b00747.

(35) Kerner, R. A.; Rand, B. P. Electrochemical and Thermal Etching of Indium Tin Oxide by Solid-State Hybrid Organic–Inorganic Perovskites. *ACS Appl. Energy Mater.* **2019**, *2* (8), 6097–6101. https://doi.org/10.1021/acsaem.9b01356.

(36) Domanski, K.; Correa-Baena, J.-P.; Mine, N.; Nazeeruddin, M. K.; Abate, A.; Saliba, M.; Tress, W.; Hagfeldt, A.; Grätzel, M. Not All That Glitters Is Gold: Metal-Migration-Induced Degradation in Perovskite Solar Cells. *ACS Nano* **2016**, *10* (6), 6306–6314. https://doi.org/10.1021/acsnano.6b02613.

(37) Kerner, R. A.; Zhao, L.; Harvey, S. P.; Berry, J. J.; Schwartz, J.; Rand, B. P. Low Threshold Voltages Electrochemically Drive Gold Migration in Halide Perovskite Devices. *ACS Energy Lett.* **2020**, *5* (11), 3352–3356. https://doi.org/10.1021/acsenergylett.0c01805.

(38) Bi, C.; Zheng, X.; Chen, B.; Wei, H.; Huang, J. Spontaneous Passivation of Hybrid Perovskite by Sodium Ions from Glass Substrates: Mysterious Enhancement of Device Efficiency Revealed. *ACS Energy Lett.* **2017**, *2* (6), 1400–1406. https://doi.org/10.1021/acsenergylett.7b00356.

(39) Dunlap-Shohl, W. A.; Barraza, E. T.; Barrette, A.; Gundogdu, K.; Stiff-Roberts, A. D.; Mitzi, D. B. $MAPbI_3$ Solar Cells with Absorber Deposited by Resonant Infrared Matrix-Assisted Pulsed Laser Evaporation. *ACS Energy Lett.* **2018**, *3* (2), 270–275. https://doi.org/10.1021/acsenergylett.7b01144.

(40) Domanski, K.; Roose, B.; Matsui, T.; Saliba, M.; Turren-Cruz, S.-H.; Correa-Baena, J.-P.; Carmona, C. R.; Richardson, G.; Foster, J. M.; De Angelis, F.; Ball, J. M.; Petrozza, A.; Mine, N.; Nazeeruddin, M. K.; Tress, W.; Grätzel, M.; Steiner, U.; Hagfeldt, A.; Abate, A. Migration of Cations Induces Reversible Performance Losses over Day/Night Cycling in Perovskite Solar Cells. *Energy Environ. Sci.* **2017**, *10* (2), 604–613. https://doi.org/10.1039/C6EE03352K.