Supporting Information for

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

Wiley A. Dunlap-Shohl,[a,b,c] Yuhuan Meng,[a,b,c] Preetham P. Sunkari,[a,b,c] David A. C. Beck,[a,d] Marina Meilă,[d,e] Hugh W. Hillhouse [a,b,c,*]

[a] Department of Chemical Engineering, University of Washington; Seattle, USA.

[b] Clean Energy Institute, University of Washington; Seattle, USA.

[c] Molecular Engineering and Science Institute, University of Washington; Seattle, USA.

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

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

*Corresponding author. E-mail: h2@uw.edu.
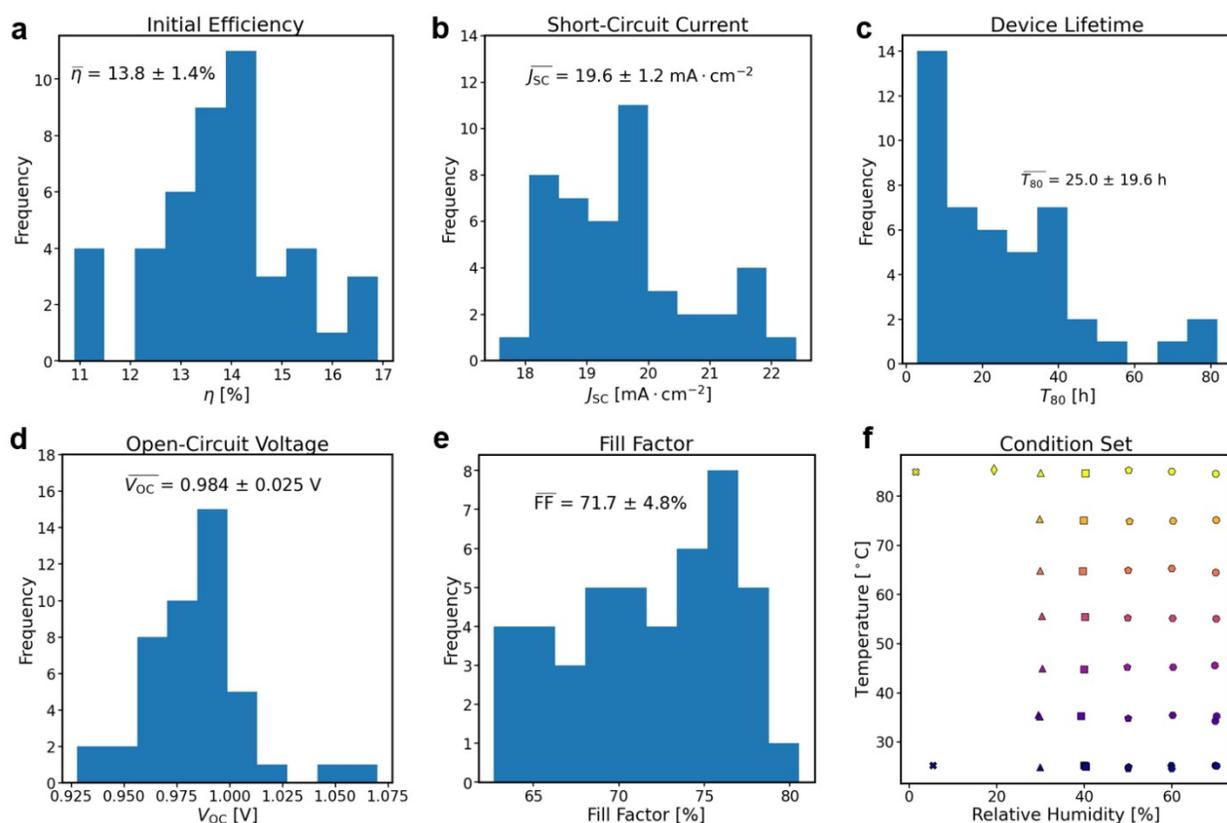
Figure S1. Histograms of device parameters and lifetimes of solar cells used in this study: (a) PCE; (b) J_{sc}; (c) T_{80}; (d) V_{oc}; (e) fill factor. (f) Distribution of environmental conditions for samples in the training set for predictive models developed in this work. Points in (f) are plotted with random jitter to facilitate visualization of conditions for which multiple samples were run (i.e., perturbations in *x*- and *y*-directions for each point are randomly sampled from a normal distribution with mean 0 and standard deviation of 0.25).
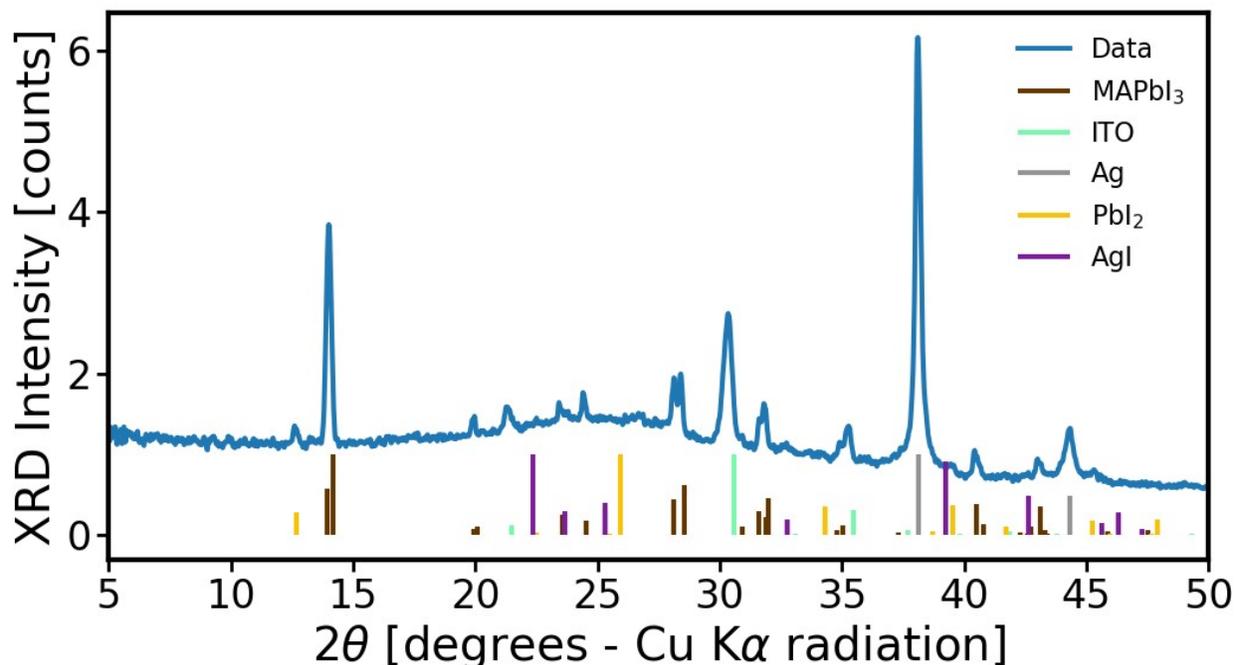
Figure S2. X-ray diffraction pattern of a perovskite solar cell after degradation at 25 °C under 1 sun illumination in air at 60% relative humidity. Comparison against stick patterns shows that the only identifiable degradation product is PbI_2, while the remaining peaks can be accounted for by materials present in pristine devices (i.e., $MAPbI_3$, ITO, and Ag). Note, however, that this measurement does not preclude the presence of other potential degradation products (such as AgI) if they are amorphous or do not build up in domains of sufficient size to scatter X-rays beyond the noise level of the measurement. Stick patterns are calculated from Crystallography Open Database files ($MAPbI_3$: COD pattern #4124388; ITO/In_2O_3: COD pattern #2310009; PbI_2: COD pattern #9009114; AgI: COD #1011025) except for that of Ag, which is taken from Wyckoff.[1]
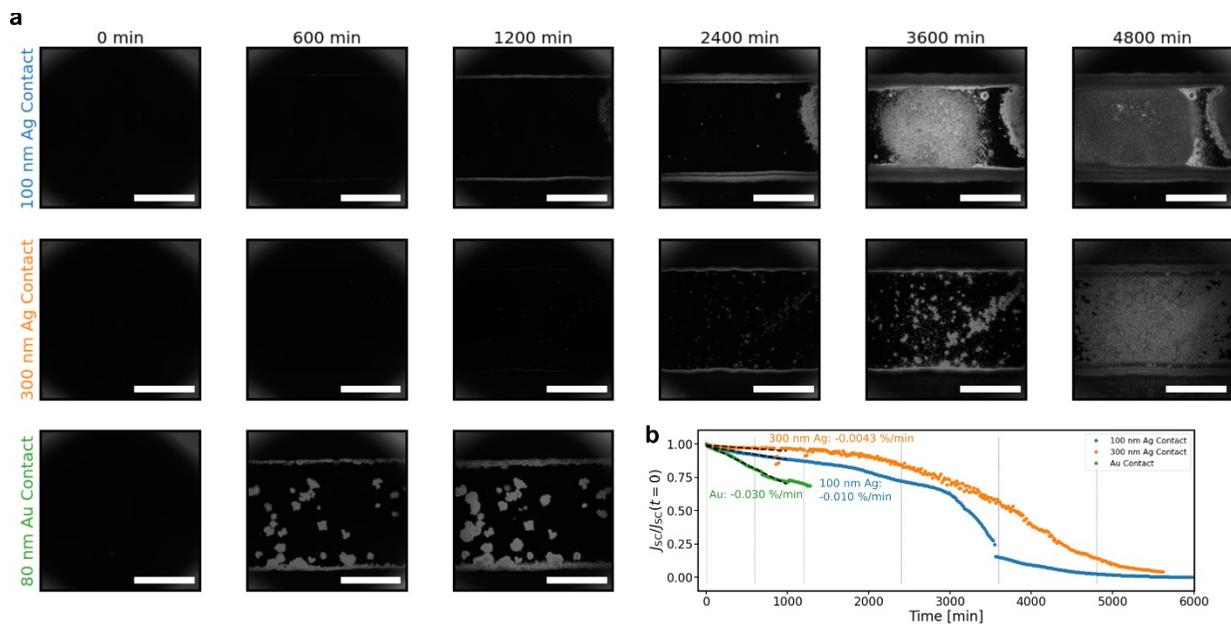
Figure S3. (a) Dark field images of devices with three different types of top contacts tested in air under 1 sun illumination at 25 °C and 50% RH: 100 nm Ag, the standard used in degradation experiments (top row); 300 nm Ag (middle row); and 80 nm Au (bottom row). (b) J_{sc} evolution for the three devices in (a); vertical dashed lines indicate the corresponding columns in (a), while dashed lines at the beginning of each trace represent a linear fit to each curve over the first 1000 min. Decay rates, estimated as the slope of this fit, are given as numerical values in (b). Increasing the thickness of the Ag electrode reduces the speed of both J_{sc} decay and dark field intensity rise overall, although in the middle of decomposition the relative J_{sc} values are comparable, possibly due to a higher density of macroscopic fabrication defects in the device with the thicker Ag film. The device with the Au contact decays significantly faster than those with Ag contacts, but exhibits similar patterns in the dark field images. Pixel brightnesses in all dark field images are scaled to the same absolute intensity (in counts/second, since the image exposure time is auto-set by the acquisition software).
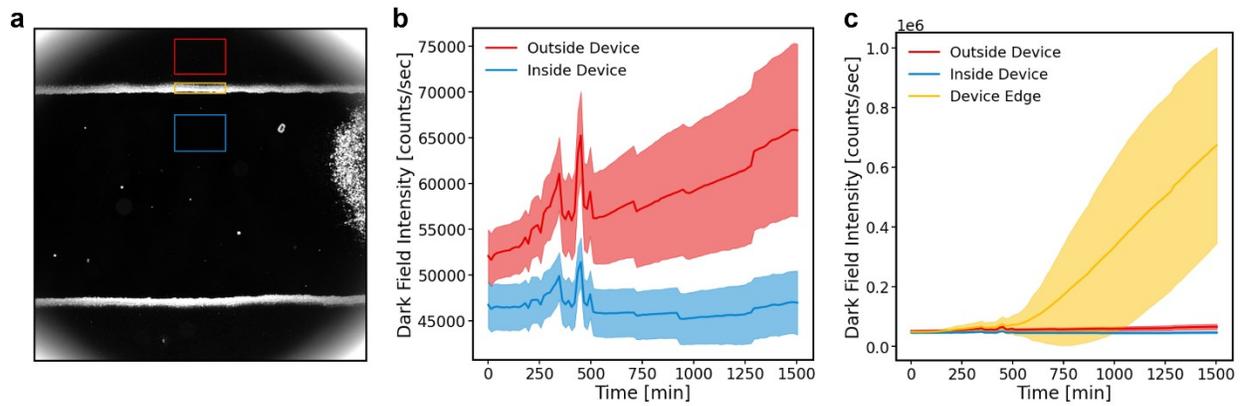
Figure S4. (a) Dark field image of a solar cell exposed to air at 25 °C and 50% RH under 1 sun illumination for 1500 min (same as Figure 2e). (b) Dark field intensity outside and inside the device up to the point at which the image in (a) was taken, calculated from the regions of interest (ROI) bounded by boxes of corresponding color in (a); solid lines represent the mean intensity within the ROI, while the shaded areas correspond to the mean ± the standard deviation of the intensity within the ROI. The ROIs are chosen specifically to avoid confounding effects of any manufacturing defects to best represent the behavior of an ideal device stack. (c) The same data as in (b) plus the dark field intensity corresponding to an ROI on the device edge indicated by the yellow bounding box in (a). Degradation occurs more quickly in the exposed region outside the device (dark field intensity increases by ~25% over the course of 1500 min) than in the device interior (intensity changes are hardly noticeable), but much more slowly than at the edge (increases by ~an order of magnitude over the same time range).
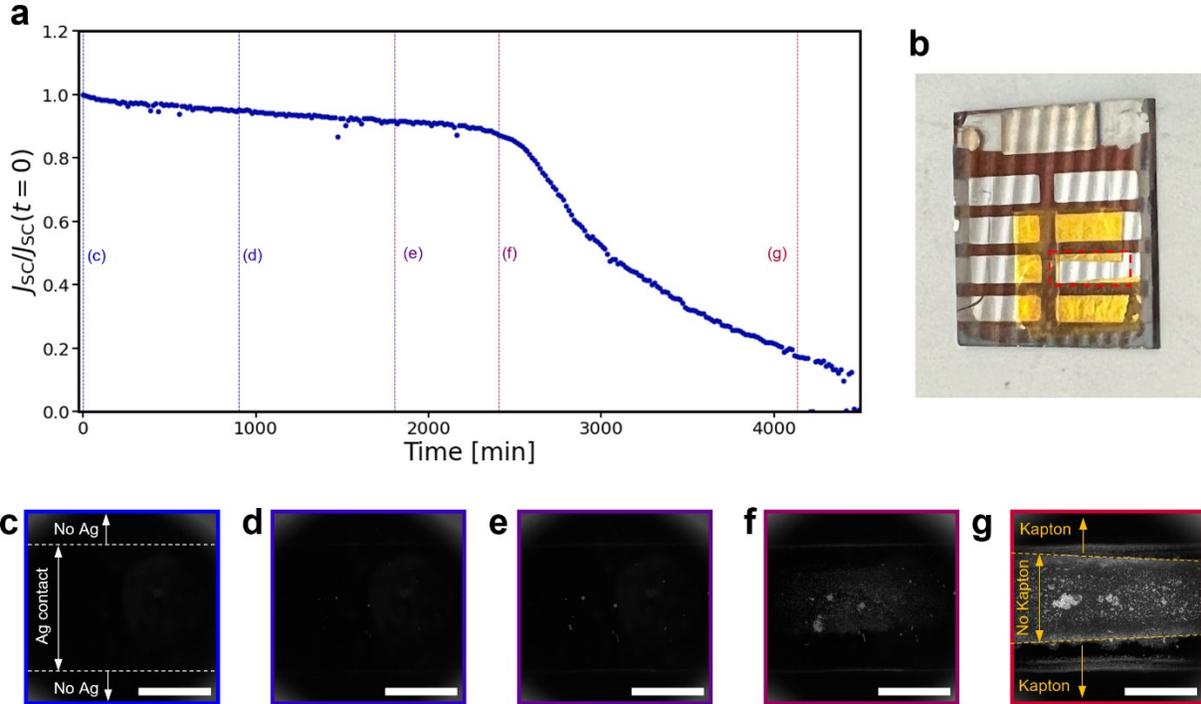
Figure S5. (a) Short-circuit current evolution under 1 sun equivalent illumination at 25 °C in 50% RH air of a $CH_3NH_3PbI_3$ solar cell with the edges covered by Kapton tape to mitigate ingress of atmospheric species through the sides of the device. (b) Photograph of the device measured, indicated by the red dashed box. (c-g) Dark field images of the device at selected times indicated in panel (a). (c) Initially, images of the device are featureless (except for an out-of-focus region on the right side, due to deposits or scratches on the glass side of the device that could not be removed by polishing before measurement). (d,e) During the initial period of slow degradation, isolated defects in the interior "nucleate" and grow over time. There is some degradation at the edges, showing that the Kapton does not entirely prevent this decay mode, but decomposition here is much slower than in the analogous case shown in Figure 2 in which the device is completely unprotected. (f) As in the case of the unprotected device, however, there is a "knee" in J_{sc} that corresponds approximately with relatively homogeneous degradation throughout the region unprotected by the Kapton. (g) Near complete loss of J_{sc} and PCE, the decomposed region has spread almost completely throughout the device; only regions that are covered by Kapton and far from the contact edges remain intact. Extent of the device, as determined by overlap of the Ag and ITO electrodes, is indicated by the dashed lines in (c); regions covered by the Kapton protection are indicated approximately by the dashed lines in (g). Scale bars in (c-g) are 1 mm.
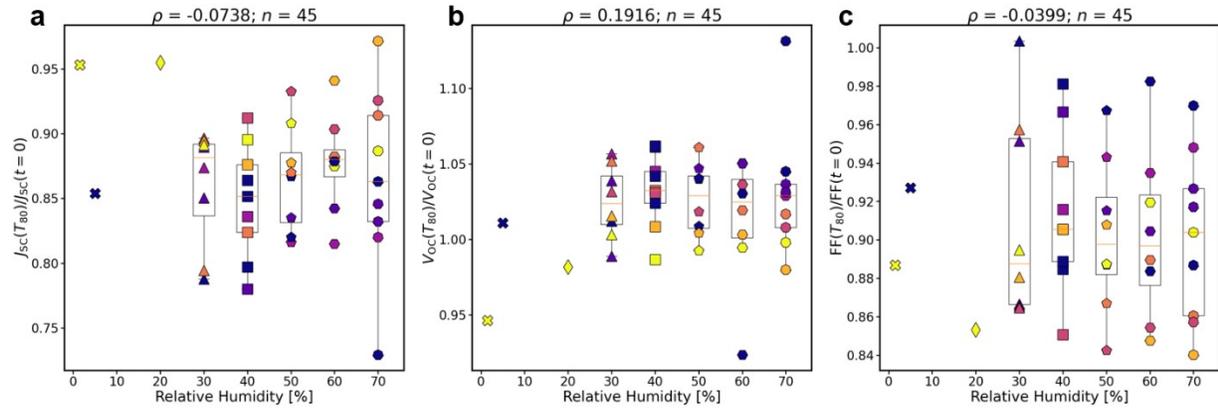
Figure S6. Scatterplots of device parameters at T_{80}, normalized to their initial values, against the relative humidity of the atmosphere under which they were degraded: (a) J_{sc}; (b) V_{oc}; (c) FF. Unlike temperature, the humidity level does not appear to have a major impact on these values.

Supplementary Note 1. Kinetic model for the decomposition rate of $CH_3NH_3PbI_3$ perovskite as a function of ambient environmental conditions.

To predict the behavior of the perovskite absorber under environmental stresses, we use a model that we have recently developed that describes the chemical decomposition of $CH_3NH_3PbI_3$ as an superposition of four distinct pathways:[2] (i) water-accelerated photooxidation (WPO); (ii) dry photooxidation (DPO); (iii) light- and humidity-induced decomposition; and (iv) pure thermal decomposition, of which (i) and (ii) are dominant in air. The decomposition rate is thus expressed as the sum of terms corresponding to each of these pathways, calculated as a function of temperature T (specified in K), respective partial pressures P_{O_2} and P_{H_2O} of O_2 and H_2O (specified in kPa), and incident photon flux above the perovskite band gap I_{in} (specified in photons $\cdot$ m^{-2} $\cdot$ s^{-1}):

$$
r\left(T, P_{O_2}, P_{H_2O}, I_{in}\right)
$$

$$
= k_{0,WPO}\exp\left(-\frac{E_{A,WPO}^{eff}}{k_BT}\right)\frac{P_{O_2}P_{H_2O}I_{in}^{0.7}}{\left(1 + K_{2W}P_{O_2}\left(1 + K_{3W}I_{in}^{0.7}\right)\right)^2} + k_{0,DPO}\exp\left(-\frac{E_{A,DPO}^{eff}}{k_BT}\right)\frac{}{1 + K_{2D}}
$$

$$
+ k_{hum}\exp\left(-\frac{E_{A,hum}^{eff}}{k_BT}\right)P_{H_2O}I_{in}^{0.7} + k_{therm}\exp\left(-\frac{E_{A,therm}^{eff}}{k_BT}\right)
$$

Values of model parameters are given in **Table S1** along with their physical interpretations; k_B is Boltzmann's constant. Further details concerning the development and interpretation of this model may be found in Siegler et al.[2]

Table S1. Values and physical interpretations of parameters of the model used to calculate the $CH_3NH_3PbI_3$ decomposition rate from ambient environmental conditions.

Parameter Symbol	Value	Physical Interpretation
$k_{0,WPO}$	2.9×10^{-25} mol $\cdot$ m^{-2} $\cdot$ s^{-1} $\cdot$ kPa^{-2} $\cdot$ (photons $\cdot$ m^{-2} $\cdot$ s^{-1})$^{-0.7}$	Rate constant for WPO
$E_{A,WPO}^{eff}$	-0.09 eV	Activation energy for WPO
K_{2W}	2.3×10^{-3} kPa^{-1}	Equilibrium constant for O_2 adsorption in WPO
K_{3W}	9.1×10^{-15} (photons $\cdot$ m^{-2} $\cdot$ s^{-1})$^{-0.7}$	Equilibrium constant for superoxide formation in WPO
$k_{0,DPO}$	5.5×10^{-15} mol $\cdot$ m^{-2} $\cdot$ s^{-1} $\cdot$ kPa^{-1} $\cdot$ (photons $\cdot$ m^{-2} $\cdot$ s^{-1})$^{-0.7}$	Rate constant for DPO
$E_{A,DPO}^{eff}$	0.62 eV	Activation energy for DPO
K_{2D}	2.7×10^{-3} kPa^{-1}	Equilibrium constant for O_2 adsorption in DPO
K_{3D}	1.1×10^{-14} (photons $\cdot$ m^{-2} $\cdot$ s^{-1})$^{-0.7}$	Equilibrium constant for superoxide formation in DPO
k_{hum}	9.2×10^{-22} mol $\cdot$ m^{-2} $\cdot$ s^{-1} $\cdot$ kPa^{-1} $\cdot$ (photons $\cdot$ m^{-2} $\cdot$ s^{-1})$^{-0.7}$	Rate constant for humidity-induced decomposition
$E_{A,hum}^{eff}$	0.20 eV	Activation energy for humidity-induced decomposition
k_{therm}	4.1×10^{-4} mol $\cdot$ m^{-2} $\cdot$ s^{-1}	Rate constant for thermal decomposition
$E_{A,therm}^{eff}$	0.45 eV	Activation energy for thermal decomposition
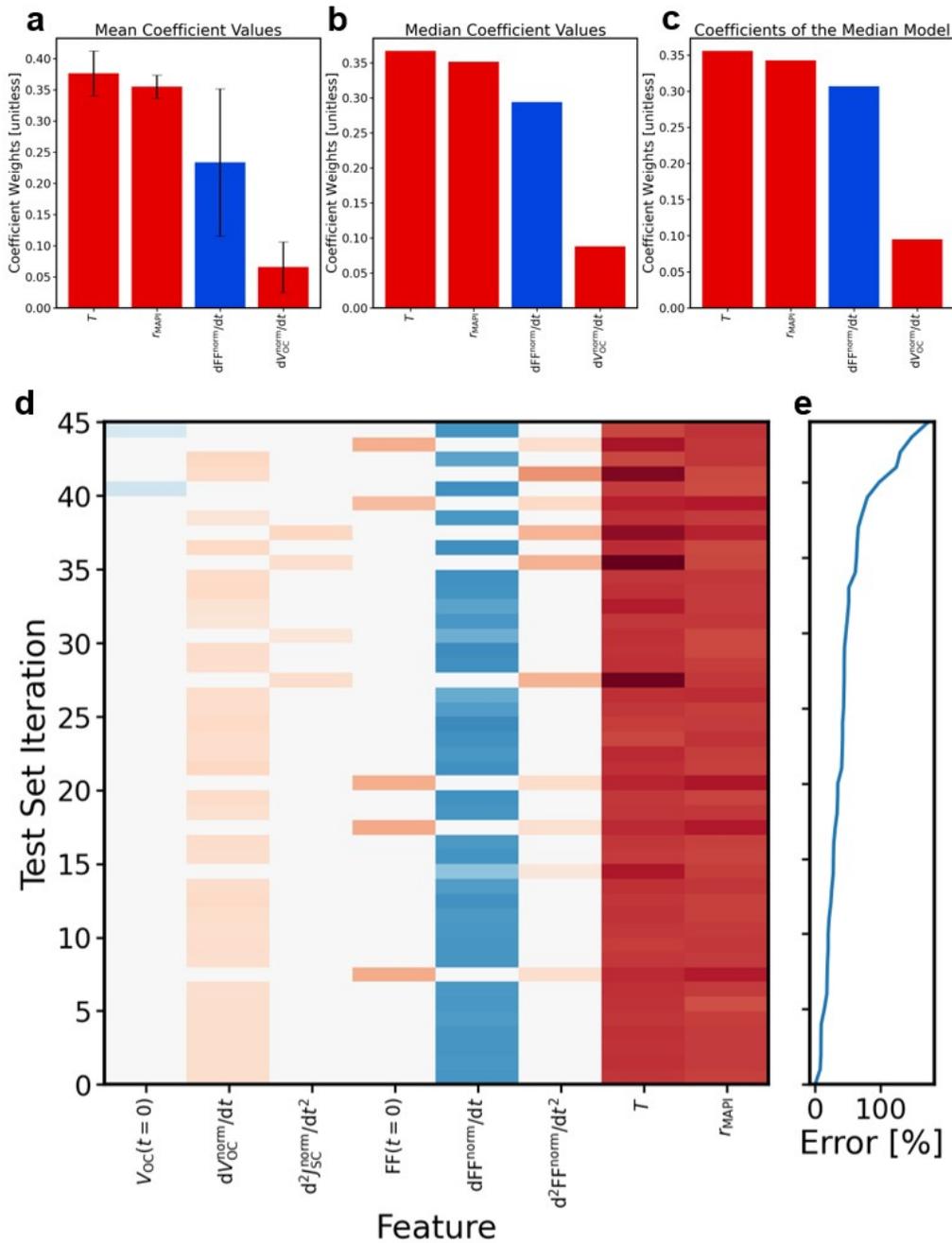
Figure S7. Feature coefficient breakdown for models trained by greedy feature selection by orthogonal matching pursuit. (a) Mean values with standard deviations represented as black bars across all test/train splits. (b) Median values across all test/train splits. (c) Coefficients of the model from the test/train split with median testing error. Color of the bars follows the convention introduced in Figure 4: red if the corresponding coefficient is negative, and blue if it is positive; positive and negative values are indicated by blue and red colors; deeper shades are of greater magnitude. (d) Coefficient values for each test/train split. (e) Test error corresponding to each test/train split. Test/train splits in (d) and (e) are sorted by test sample prediction error to aid in visualization of the structure of models with similar performance.
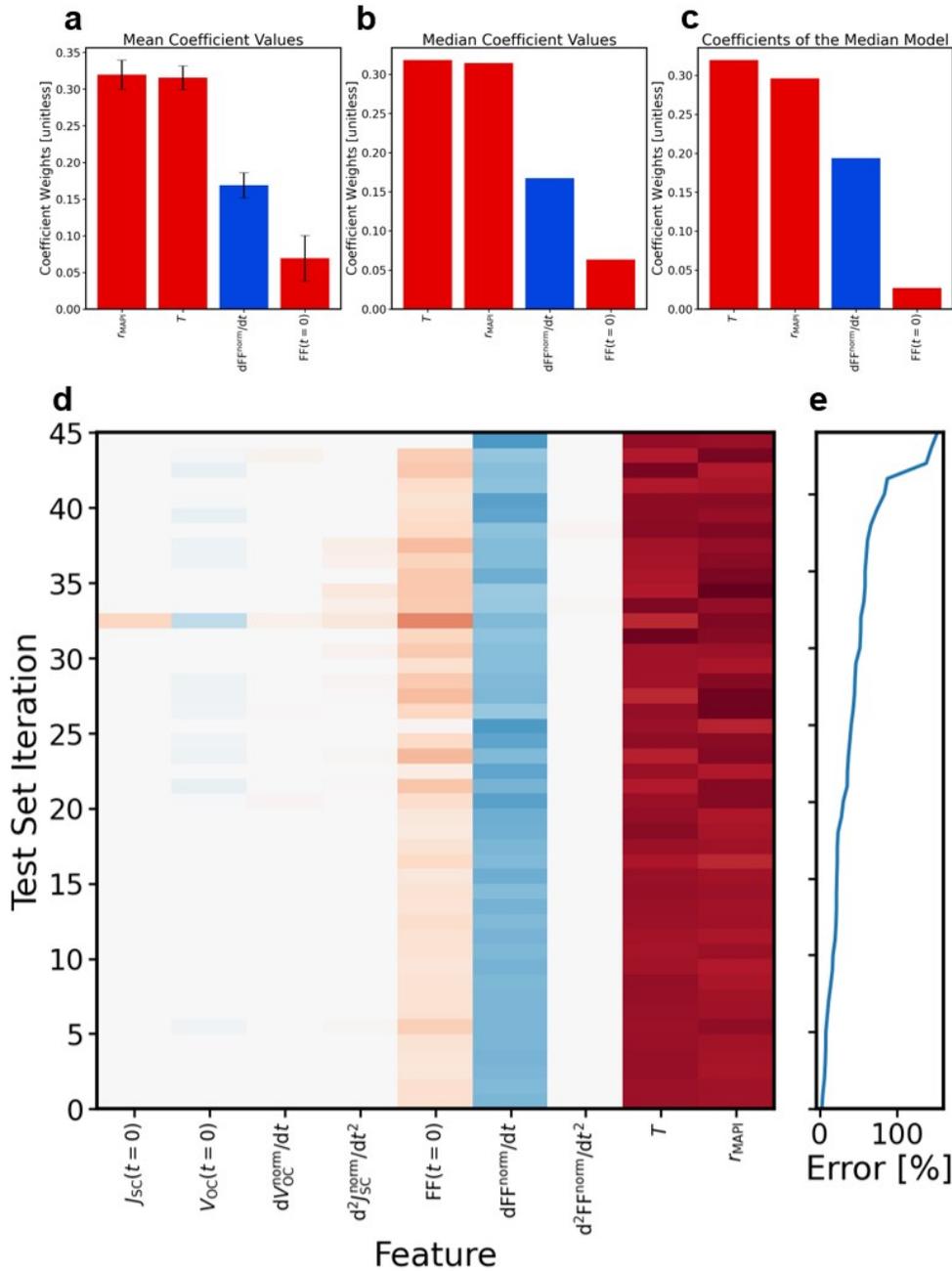
Figure S8. Feature coefficient breakdown for models trained by LASSO. (a) Mean values with standard deviations represented as black bars across all test/train splits. (b) Median values across all test/train splits. (c) Coefficients of the model from the test/train split with median testing error. Color of the bars follows the convention introduced in Figure 4: red if the corresponding coefficient is negative, and blue if it is positive. (d) Coefficient values for each test/train split; positive and negative values are indicated by blue and red colors; deeper shades are of greater magnitude. (e) Test error corresponding to each test/train split. Test/train splits in (d) and (e) are sorted by test sample prediction error to aid in visualization of the structure of models with similar performance.

Figure S9. Feature coefficient breakdown for models trained by ridge regression. (a) Mean values with standard deviations represented as black bars across all test/train splits. (b) Median values across all test/train splits. (c) Coefficients of the model from the test/train split with median testing error. Color of the bars follows the convention introduced in Figure 4: red if the corresponding coefficient is negative, and blue if it is positive. (d) Coefficient values for each test/train split; positive and negative values are indicated by blue and red colors; deeper shades are of greater magnitude. (e) Test error corresponding to each test/train split. Test/train splits in (d) and (e) are sorted by test sample prediction error to aid in visualization of the structure of models with similar performance.

Supplementary Note 2: Sparse modeling using a hybrid best-subset selection method (the $l_0 l_2-$ penalized linear model).

In the simple ordinary least squares (OLS) linear regression method, a sum-squared error loss function is minimized to obtain the optimal coefficients β that best fit the data using a linear model: $y_{pred} = \beta_0 + \beta \cdot x$, where x is the vector of features or independent variables. However, when the data are scarce in comparison to the number of features available, the simple linear regression tends to overfit noise in the data in addition to any real trends that might be present. To reduce overfitting, penalized linear models whose loss functions contain additional terms characterizing the magnitude and complexity of the coefficient vector β have been developed to produce sparse models, in which only a small subset of features from the original menu are selected to build the final linear model. The least absolute selection and shrinkage operator (LASSO)[3] is the most popular penalized linear method that uses an l_1 penalty (the sum of the absolute values of the coefficients in β) to penalize the insignificant features in the loss function. Though it is well known to produce sparse solutions by "shrinking" the coefficients of the insignificant features to zeros and thus selecting a handful of features to enter the model,[3] it is found to render inconsistent solutions (containing many irrelevant features undesirably selected as significant) in cases with very few runs and the presence of collinearity in the feature data.[4] Another popular penalized method is ridge regression[5] which is based on the l_2-penalty (the sum of the squares of the feature coefficients). Compared to LASSO, ridge regression is less aggressive in shrinkage of coefficients and often leads to models using many features. This potentially leads to overfitting especially when the data is scarce. Recently, the best-subset selection (BSS) method (sometimes known as the l_0-penalized method)[5–8] which involves evaluation of all possible subsets of features independently with simple linear regression, has been growing popular. Since a straightforward exploration of all possible subsets is a discrete-optimization problem with NP-hard time-complexity, recent developments in linear programming have offered many faster solutions, with the fastest being the one with a polynomial time-complexity.[8] This method acts as a good alternative to the conventional LASSO method due to its ability to produce consistent solutions and select features fewer than the LASSO method, even when the datasets are small. Recently, a hybrid BSS method based on an $l_0 l_2$ penalty[7] has been developed by augmenting the feature selection ability of the l_0 penalty with that of the l_2 penalty to simultaneously shrink the coefficients. This method is known to not only produce sparse solutions like l_0, but also to be robust in data with large noise like l_2. Table S2 compares the loss functions of the popular penalized models discussed here, where $X = \{X_1, X_2, ..., X_p\}$ is the feature data matrix (i.e., each vector X_j consists of N observations of the associated feature), y is the target data vector (consisting of N corresponding observations of the dependent variable), and λ_0, λ_1, and λ_2 are model hyperparameters used to scale the penalty terms. The pseudo-code below (Algorithm S1) outlines the modeling methodology used in this paper. We use the adaptive best subset selection algorithm developed by Zhu et al.[8] for implementation of the hybrid BSS method.

Table S2. Penalized loss functions that are minimized to obtain the optimal β^*

Method (penalty)	Loss function over $\beta = \{\beta_0, \beta_1, ..., \beta_p\}$
OLS (unpenalized)	$L_{OLS}(\beta \| X, y) = \sum_i^N [y_i - (\beta_0 + \beta_1 X_{1i} + ... + \beta_p X_{pi})]^2$
LASSO (l_1)	$L_{l_1}(\beta \| X, y, \lambda_1) = L_{OLS}(\beta \| X, y) + \lambda_1 \cdot \|\beta_1\| + \|\beta_2\| + ... + \|\beta_p\|$
Ridge (l_2)	$L_{l_2}(\beta \| X, y, \lambda_2) = L_{OLS}(\beta \| X, y) + \lambda_2 \cdot (\beta_1^2 + \beta_2^2 + ... + \beta_p^2)$
Best-subset selection (l_0)	$L_{l_0}(\beta \| X, y, \lambda_0) = L_{OLS}(\beta \| X, y)$ s.t. # of nonzero elements in $\beta \setminus \beta_0$
Hybrid best-subset selection ($l_0 l_2$)	$L_{l_0 l_2}(\beta \| X, y, \lambda_0, \lambda_2) = L_{l_2}(\beta \| X, y, \lambda_2)$ s.t. # of nonzero elements in β

Algorithm S1. Pseudocode for sparse modeling

Start: Feature menu F with p variables and data with N runs.

Loop-1: Leave-one-out testing scheme
for $i = \{1, 2, ..., N\}$:

Test set indices: $S_{test}^i = \{i\}$ $\rightarrow$ 1 run

Train set indices: $S_{train}^i = \{1, 2, ..., N\} \setminus \{i\}$ $\rightarrow$ $N - 1$ runs

Loop-2: Grid-search of combinations of hyperparameters (λ_0, λ_2)
Range of λ_0: $\{1, 2, ..., \lambda_{0,max}\}$ where $\lambda_{0, max}$ is the maximum sparsity
Range of λ_2: $10^{-3} - 10^2$

for $(\lambda_0^l, \lambda_2^l)$ in 2D grid of λ_0 and λ_2:

Loop-3: Leave-one-out cross-validation for hyperparameter tuning
for k in S_{train}:

Tuning validation indices: $S_{val, \lambda}^k = \{k\}$ $\rightarrow$ 1 run

Tuning train indices: $S_{train, \lambda}^k = S_{train}^i \setminus \{k\}$ $\rightarrow$ $N - 2$ runs

- Solve a linear model by minimizing an $l_0 l_2 -$ penalized loss function:

$$L_{l_0 l_2}(\beta \| X, y, \lambda_0^l, \lambda_2^l) = \sum_j^{S_{train, \lambda}^k} [y_j - (\beta_0 + \beta_1 X_{1j} + ... + \beta_p X_{pj})]^2$$
$$+ \lambda_2^l (\beta_1^2 + \beta_2^2 + ... + \beta_p^2)$$

such that the number of non-zero elements in $\{\beta_1, \beta_2, ..., \beta_p\}$ is λ_0^l,

where $X = \{X_1, X_2, ..., X_p\}$ is the feature data matrix and y is the target data array

- Record the validation error over $S_{val,\lambda}^k$ for each $(\lambda_0^l, \lambda_2^l)$

Obtain the mean validation error from loop-3 for each $(\lambda_0^l, \lambda_2^l)$.

Choose the optimal $(\lambda_0^i, \lambda_2^i)$ combination with the smallest mean validation error from loop-2 and obtain the corresponding selected feature subset $F^i \subseteq F$. The size of the selected feature subset is λ_0^i

Finally, perform an ordinary least squares (OLS) linear regression over F^i to obtain the final coefficients for the model corresponding to S_{train}^i

Report the median test error from loop-1 as the test error and the features with large magnitudes of non-zero coefficients across the i-iterations as the overall significant features.
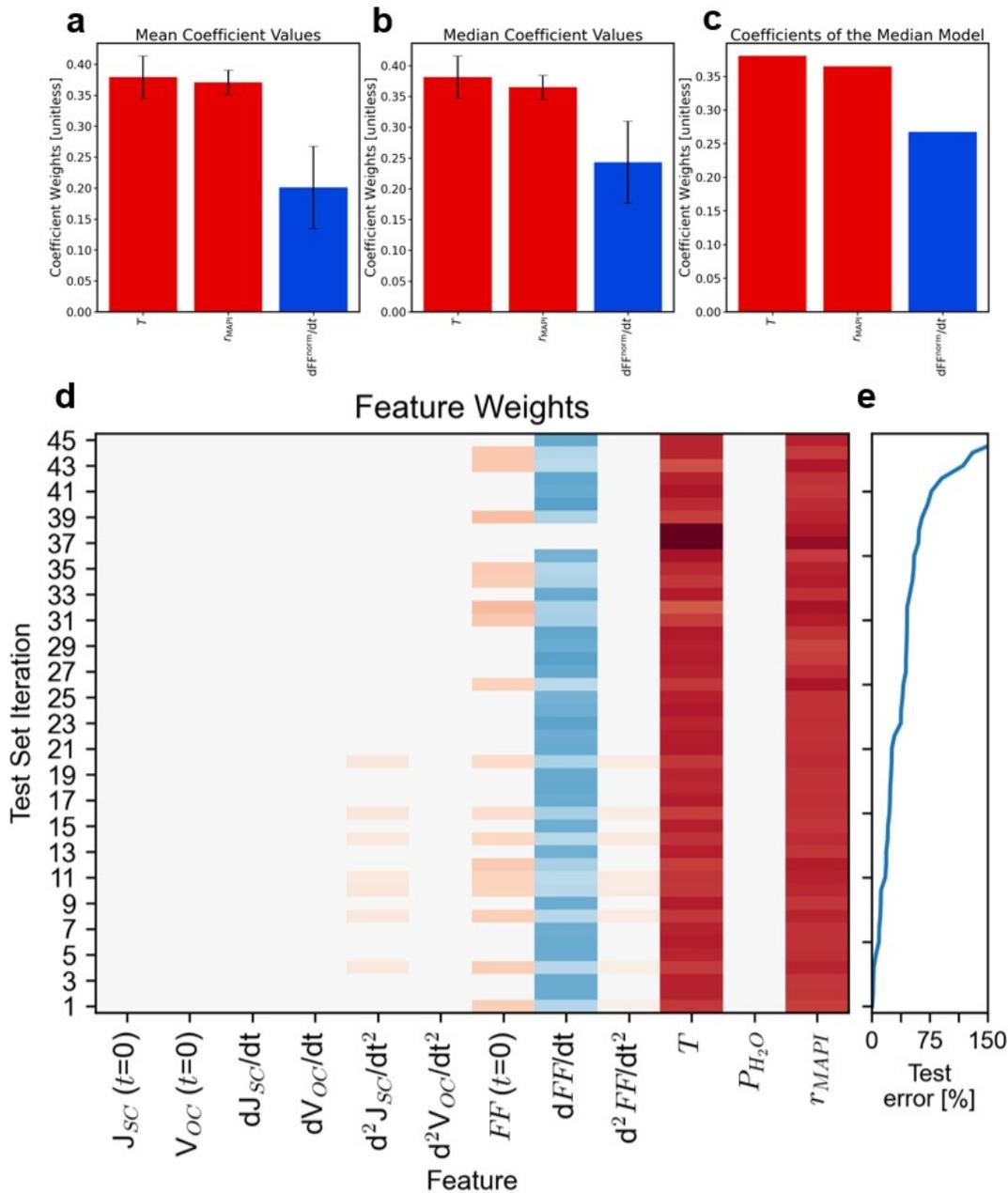
Figure S10. Feature coefficient breakdown for models trained by hybrid best subset selection. (a) Mean values with standard deviations represented as black bars across all test/train splits. (b) Median values across all test/train splits. (c) Coefficients of the model from the test/train split with median testing error. Color of the bars follows the convention introduced in Figure 4: red if the corresponding coefficient is negative, and blue if it is positive. (d) Coefficient values for each test/train split; positive and negative values are indicated by blue and red colors; deeper shades are of greater magnitude. (e) Test error corresponding to each test/train split. Test/train splits in (d) and (e) are sorted by test sample prediction error to aid in visualization of the structure of models with similar performance.

Figure S11. (a) Device characteristics from a degradation experiment performed at 25 °C in air at 20% relative humidity on a solar cell using a 300 nm thick Ag contact. During periods indicated by yellow bands, the device was held under 1 sun-equivalent illumination from a 542 nm LED at maximum power point (when not being actively measured); otherwise, the device was held at short circuit in the dark. (b, c) Scatterplots of fill factor against series resistance measured from forward (b) and reverse (c) J-V scans. Fill factor improvements appear to have both reversible and irreversible components, and have a strong negative correlation with series resistance.

References

1. R. W. G. Wyckoff, *Crystal Structures*, Interscience Publishers, New York, NY, 2nd edn., 1963, vol. 1.
2. T. D. Siegler, W. A. Dunlap-Shohl, Y. Meng, Y. Yang, W. F. Kau, P. P. Sunkari, C. E. Tsai, Z. J. Armstrong, Y.-C. Chen, D. A. C. Beck, M. Meilă and H. W. Hillhouse, Water-Accelerated Photooxidation of $CH_3NH_3PbI_3$ Perovskite, *J. Am. Chem. Soc.*, 2022, **144**, 55s52–5561.
3. R. Tibshirani, Regression Shrinkage and Selection via the Lasso, *J. R. Stat. Soc. Series B, Stat. Methodol.*, 1996, **58**, 267–288.
4. P. Zhao and B. Yu, On model selection consistency of Lasso, *J. Mach. Learn. Res.*, 2006, **7**, 2541–2563.
5. T. Hastie, R. Tibshirani and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2nd edn., 2009.
6. D. Bertsimas, A. King and R. Mazumder, Best subset selection via a modern optimization lens, *Ann. Stat.*, 2016, **44**, 813–852.
7. H. Hazimeh and R. Mazumder, Fast Best Subset Selection: Coordinate Descent and Local Combinatorial Optimization Algorithms. Operations Research, *Oper. Res.*, 2020, **68**, 1517–1537.
8. J. Zhu, C. Wen, J. Zhu, H. Zhang and X. Wang, A polynomial algorithm for best-subset selection problem, *Proc. Natl. Acad. Sci. U.S.A.*, 2020, **117**, 33117–33123.