Joule, Volume *6*

Supplemental information

Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing

Zhe Liu, Nicholas Rolston, Austin C. Flick, Thomas W. Colburn, Zekun Ren, Reinhold H. Dauskardt, and Tonio Buonassisi

Supplemental Items

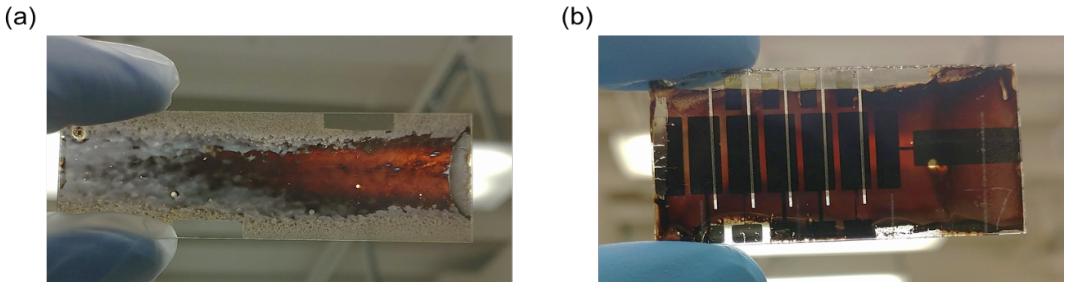

Figure S1: The pictures of visual inspections. (a) A poor-quality film that failed the visual inspection. (b) A good-quality film that were fabricated into devices.

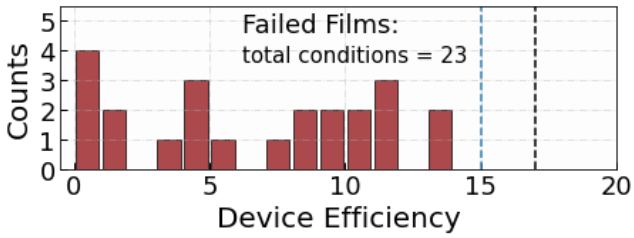

Figure S2: The efficiency distribution for those devices that were fabricated using low-quality films, as a confirmation of the visual inspection. The blue dashed line marks 15% efficiency, and the black dashed line marks 17% efficiency.

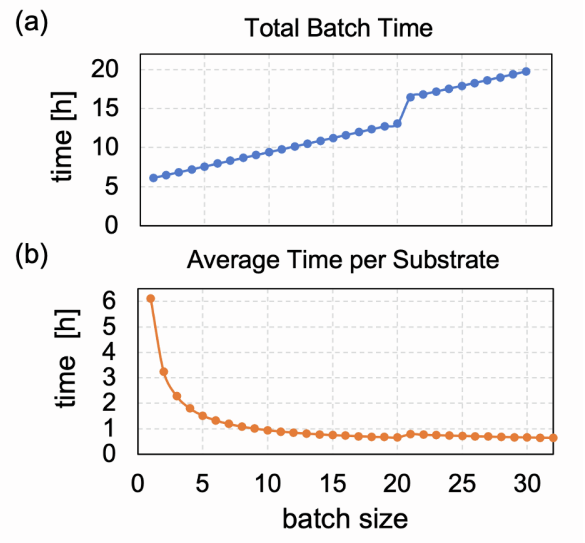

Figure S3: Estimated consumption of experimental time for fabricating perovskite devices. (a) The total time per batch for different batch size. (b) The average time per substrate for different batch size.

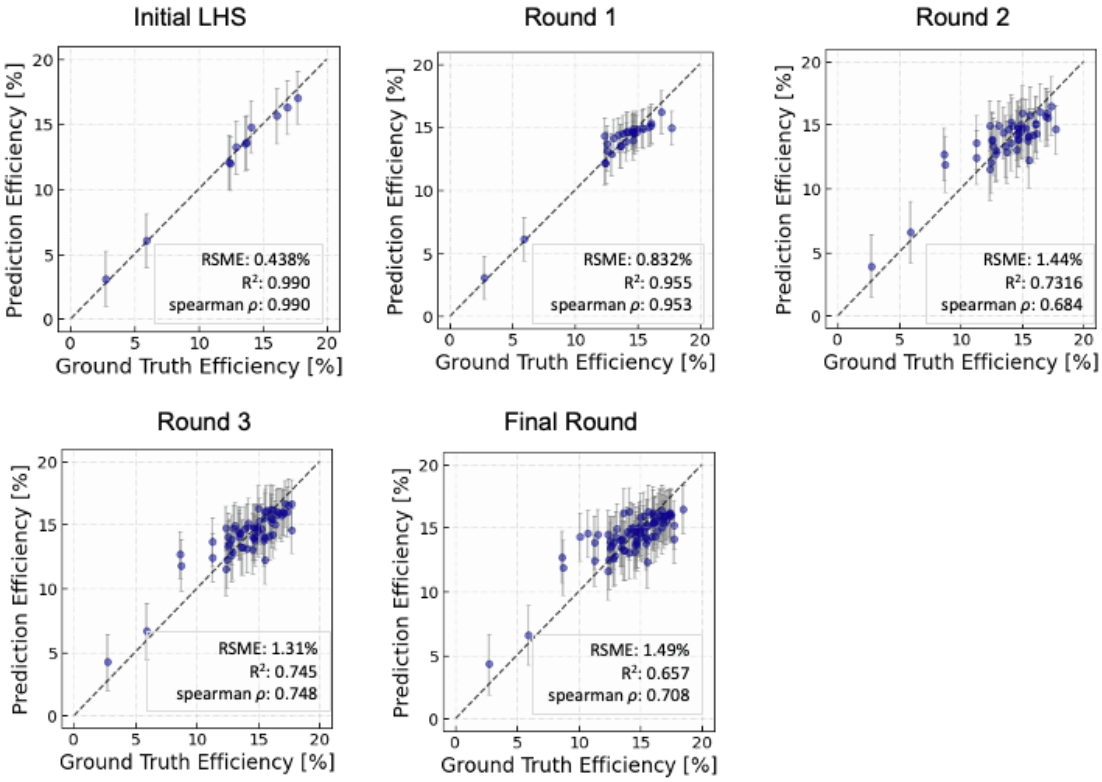

Figure S4: The predictions by surrogate model versus the ground truth of the measured PCEs after every experimental iteration. The error bars are the estimated standard deviation for the prediction uncertainty (*i.e.*, 68.2% confidence interval). The numerical metric for the model errors (i.e., root mean squared error RMSE, goodness of fit R^2, spearman coefficient) are shown in the inserted box.

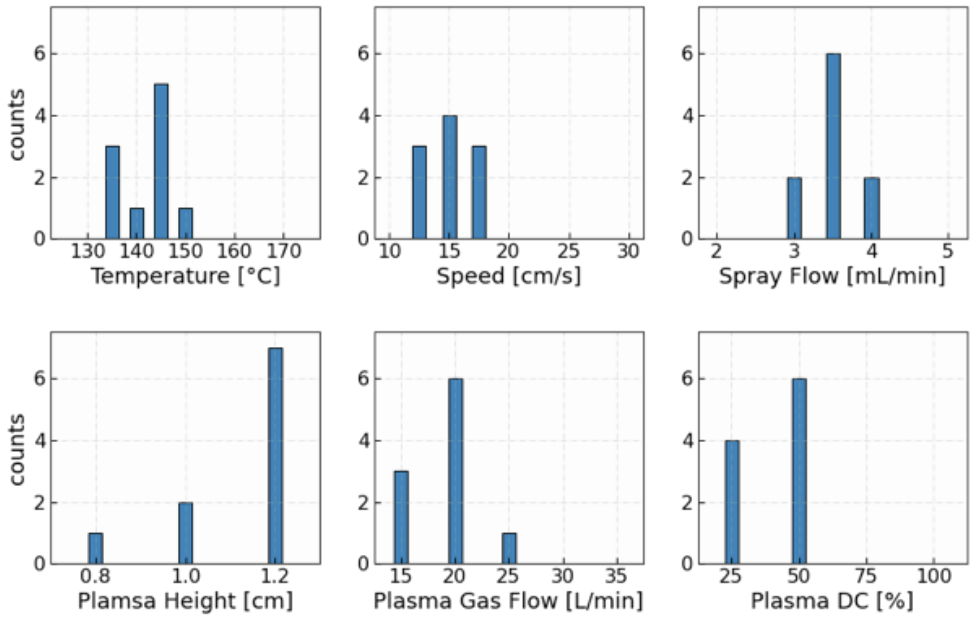

Figure S5: The histogram of the process conditions for the 10 top-performer (PCE ≥17% devices).

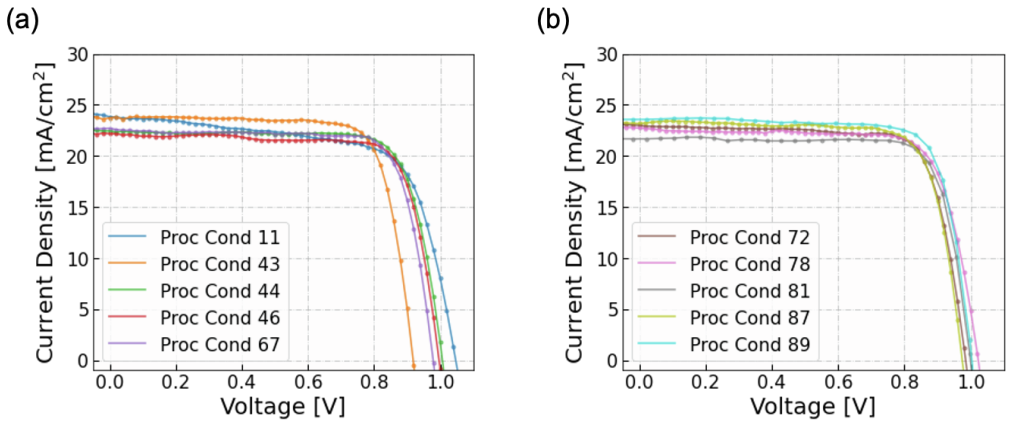

Figure S6: The in-house measured *I-V* curves of the top efficiency devices (≥17%) that were produced in this work. Process condition no. 89 achieves the champion efficiency of 18.5% efficiency.

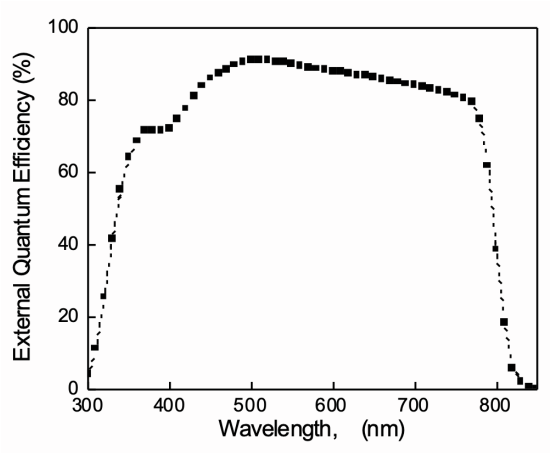

Figure S7: The external quantum efficiency (EQE) of the champion device fabricated under process condition 89. The integrated current density 23.02 mA/cm^2 (weighted under AM1.5G)

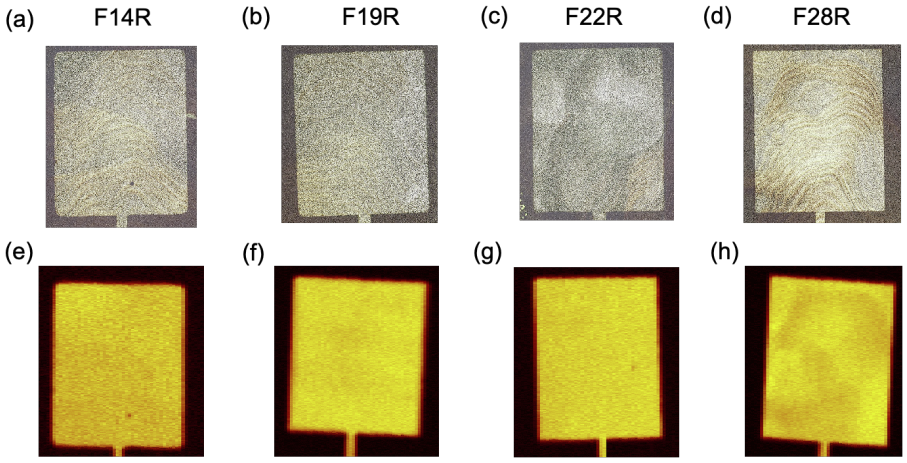

Figure S8: The optical images (in a, b, c, d) and the light-beam-induced-current maps (in e, f, g, h) are shown for four devices fabricated under different process conditions.

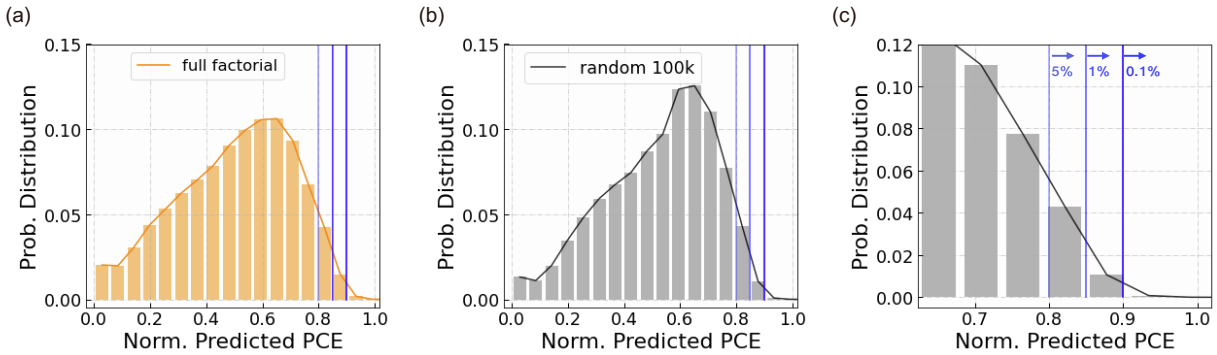

Figure S9: The histogram distribution of the predicted efficiencies by the "teacher" regression model. (a) The full factorial sampling of all the grid levels defined in Table I. (b) The randomly sampled 100k process conditions. (c) The zoom-in view of (b) to show the blue lines of the top 5%, 1% and 0.1% in the random 100k samples, which are approximately corresponding to 0.8, 0.85 and 0.9 respectively in the normalized PCE.

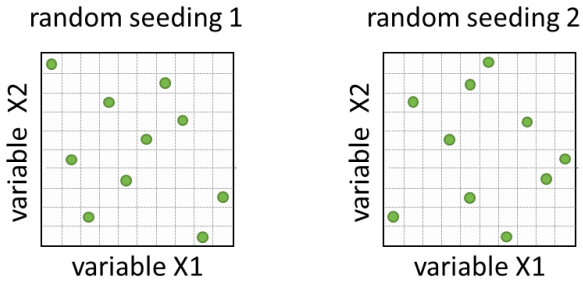

Figure S10: Schematic illustration of Latin hypercube sampling (LHS) with different runs of random seeding in an exemplary 2D space. The green solid circles are the sampled conditions.

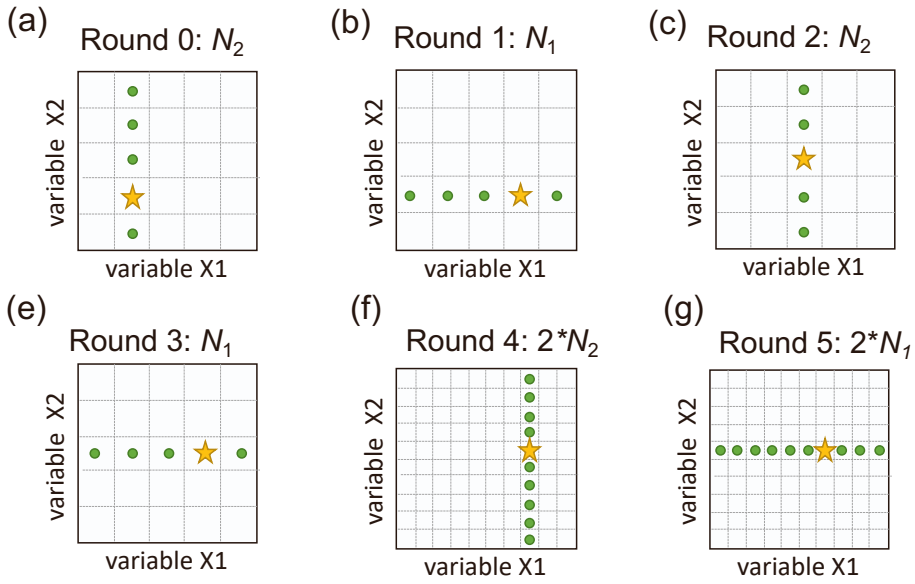

Figure S11: Schematic illustration of one variable at a time sampling (OVATS). The green solid circles are sampled conditions in that round. The yellow star is the best condition of all the samples until that specific round.

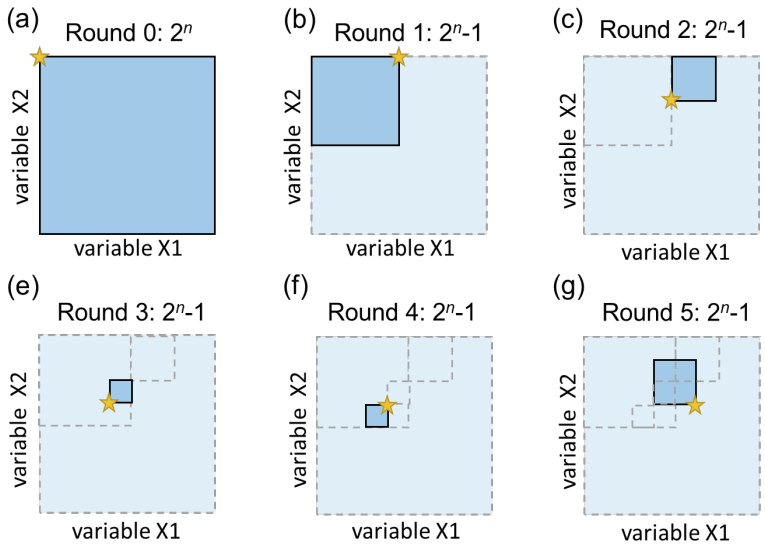

Figure S12: Schematic illustration of an exemplary implementation of the factorial sampling with progressive grid subdivision. The darker blue box (with black lines) is the parameter space for two-level factorial sampling in the specific round. The dashed gray boxes are the sampled space in previous rounds, where corner points or intersections are the conditions that have already been sampled. The yellow star is the best device of all the samples. The smallest box is 1/8 of the entire parameter space, so the total level of each variable is nine. In this illustration, the total number of sampling points is 379 after all six rounds.

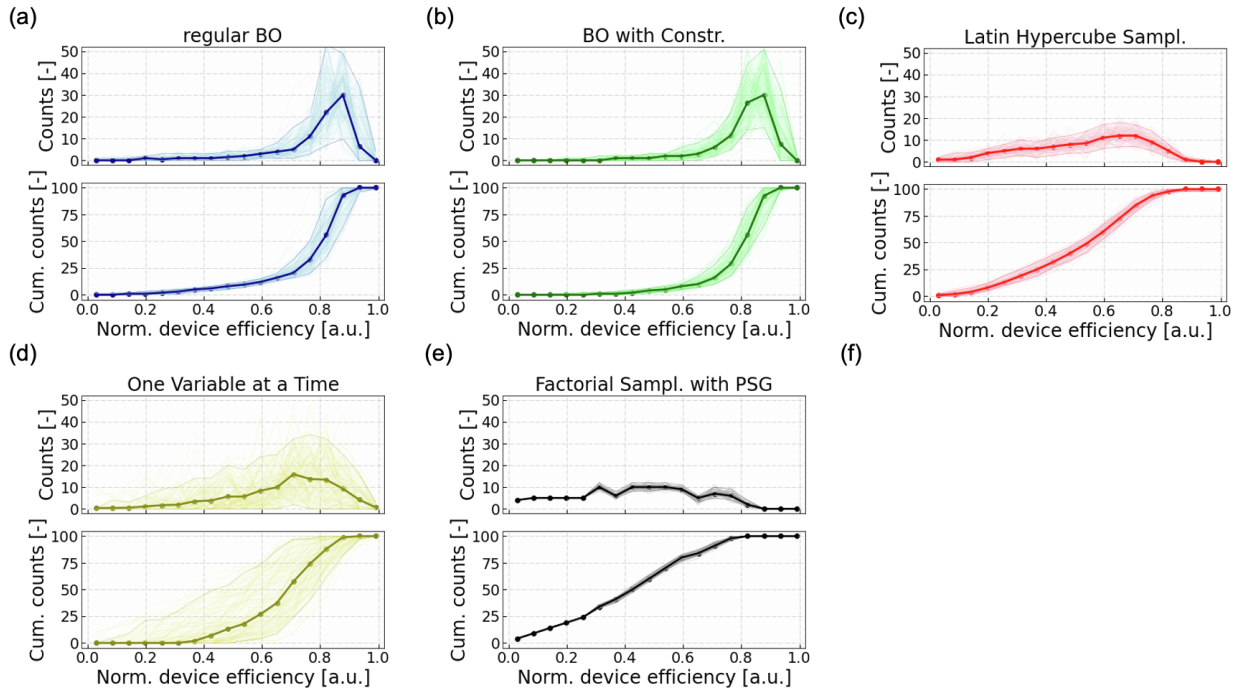

Figure S13: The histogram distribution within the first 100 experimental budget by the simulated process optimization. The lighter lines in each figure represents the 100 separate runs with random seeding. The darker solid line represents the median of the 100 runs, and the dashed lines indicates the confidence interval of 5% – 95%.

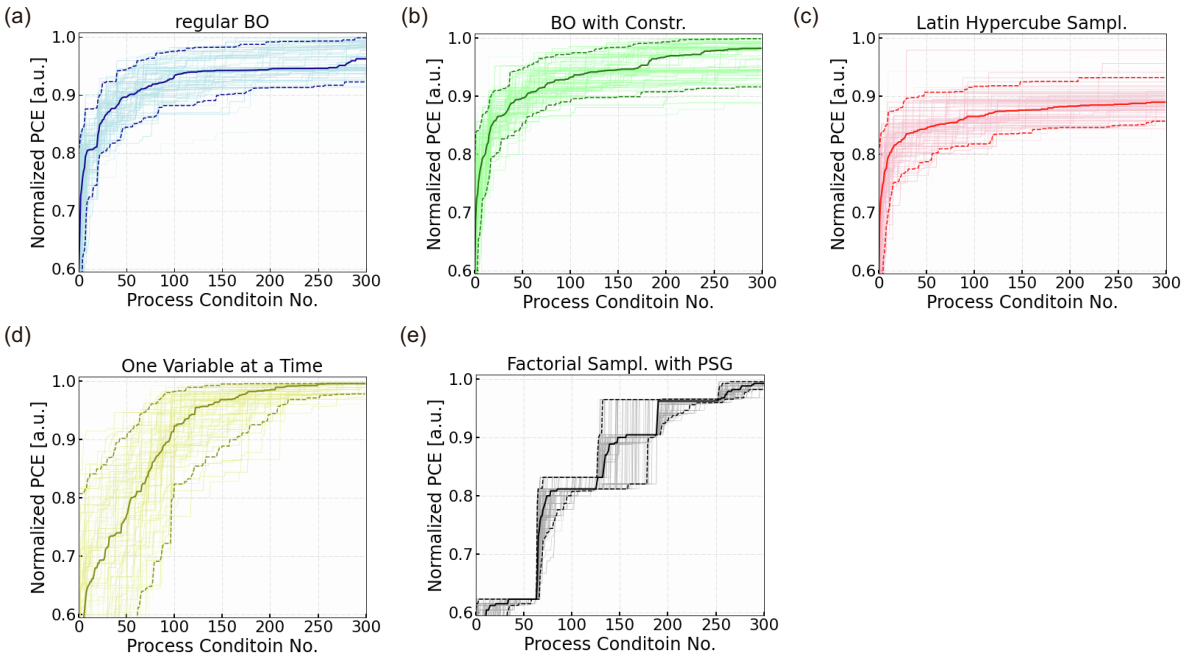

Figure S14: The best efficiency at a given number of process conditions by the simulated process optimization. The lighter lines in each figure represents the 100 separate runs with random seeding. The darker solid line represents the median of the 100 runs, and the dashed lines indicates the confidence interval of 5% – 95%.

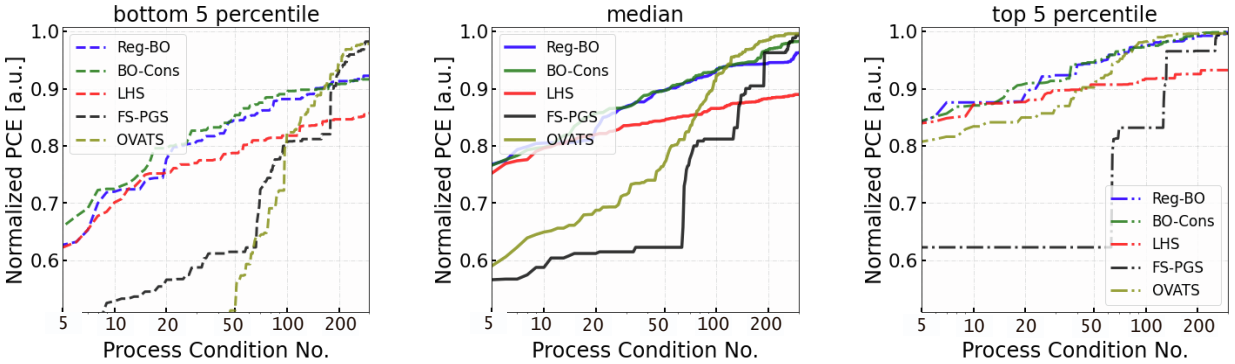

Figure S15: The comparison of the best PCE vs the number of process conditions explored among the different sampling methods. Different percentiles are plotted in the (a), (b) and (c) respectively.

Table S1: Relative humidity in the RSSP deposition chamber for each batch of devices.

	Initial	Round 1	Round 2	Round 3	Final
Process Condition	0 – 19	20 – 39	40 – 59	60 – 79	80 – 99
Acquisition Method	LHS	BO with UCB	BO with UCB	BO with UCB	Local region search
Relative Humidity [%]	43%	27%	25%	50%	35%
The ambient temperature is 23°C±2°C.					

Table S2: The parameter space for the final round of the local region search.

Process Variable	Total Range (Interval)	Process Variable	Total Range (Interval)
Temperature	140 – 150 °C (1°C)	Plasma height	1.0 – 1.2 cm (0.05 cm)
Speed	12.5 – 17.5 cm/s (0.1 cm/s)	Plasma gas flowrate	16 – 20 L/min (1 L/min)
Spray flowrate	3 – 3.5 mL/s (0.01 mL/min)	Plasma duty cycle	25 – 50 % (1%)

Table S3: Estimated time consumption for some key steps in device fabrication.

Steps	Setup Time	Processing Time
Substrate Cleaning	10 mins to fill cleaning agents	50 mins for up to 20 substrates
RSPP Deposition	1hr to reconfigure & test the tool	10 – 15 mins per condition
Evaporation (batch-20)	30 mins to pump the pressure down	3 hrs. for up to 20 substrates
I-V characterization	15 mins to warm up the lamp + circuit setup	10 mins per substrate

Table S4. All the sampled experimental process conditions and the fabricated device efficiency.

ML Condition	Tempe-rature [°C]	Speed [mm/s]	Spray Flow [uL/min]	Plamsa Height [cm]	Plasma Gas Flow [L/min]	Plasma DC [%]	Efficiency [%]	Film Good or Not?
0	155	200	5000	1.2	35	50	5.91	Yes
1	160	225	3000	0.8	30	75	0.06	No
2	135	200	2500	1	25	75	12.89	Yes
3	150	175	2000	1	20	100	4.77	No
4	170	275	4500	1.2	16	50	11.82	No
5	175	250	3500	1.2	25	75	9.06	No
6	140	150	4000	1	20	75	14.05	Yes
7	155	225	4000	0.8	30	25	9.88	No
8	130	175	3500	0.8	25	75	16.90	Yes
9	135	125	2500	1.2	20	25	16.05	Yes
10	130	250	4500	1.2	30	50	12.34	Yes
11	145	125	3500	1	25	50	17.70	Yes
12	145	150	4500	1	16	100	12.43	Yes
14	165	225	4000	1	25	25	13.66	Yes
15	125	275	3000	0.8	20	50	13.61	Yes
16	170	175	5000	1	16	75	13.40	No
17	150	100	2500	1	35	50	2.70	Yes
18	140	300	2000	1	30	50	-	No
19	165	125	3000	1	30	75	11.42	No

20	130	200	4500	1.2	25	50	15.87	Yes
21	135	125	3500	1	25	25	12.54	Yes
22	135	150	2500	1.2	20	25	14.16	Yes
23	135	150	3500	1.2	25	25	13.84	Yes
24	135	175	4000	1.2	25	25	16.12	Yes
25	135	175	4000	1.2	25	50	13.04	Yes
26	140	100	3500	1	25	50	14.72	Yes
27	140	100	3500	1.2	25	25	15.38	Yes
28	140	100	4000	1	25	50	13.41	Yes
29	140	125	2500	1	25	50	14.09	Yes
30	140	125	3500	1.2	25	25	14.89	Yes
31	140	150	3500	1	25	50	12.34	Yes
32	140	150	4000	1	25	50	14.85	Yes
33	145	100	3500	1.2	25	50	14.61	Yes
34	145	100	4500	1	25	50	14.67	Yes
35	145	125	3000	1	25	50	14.54	Yes
36	145	150	3500	1	25	50	14.61	Yes
37	145	150	4000	1	25	50	14.61	Yes
38	150	125	3500	1.2	25	50	12.54	Yes
39	150	150	4500	1.2	25	50	15.41	Yes
40	130	150	5000	1.2	25	25	12.82	Yes
41	130	225	4500	1.2	20	25	8.72	Yes
42	135	100	4500	1.2	20	25	15.50	Yes
43	135	150	3500	1.2	20	25	16.97	Yes
44	135	150	4000	0.8	20	50	17.37	Yes
45	135	175	3000	0.8	20	75	16.42	Yes
46	135	175	4000	1.2	20	50	17.04	Yes
47	140	100	3500	1	15	25	14.97	Yes
48	140	175	4500	1	20	25	15.22	Yes
49	140	175	5000	1.2	20	25	14.55	Yes
50	140	200	4000	0.8	20	75	15.65	Yes
51	140	200	5000	1.2	20	50	12.61	Yes
52	145	100	4500	1.2	20	25	11.23	Yes
53	145	100	4500	1.2	25	25	11.25	Yes
54	145	175	3500	1	20	50	14.93	Yes
55	145	175	4000	1.2	20	25	16.08	Yes
56	145	200	5000	1.2	25	25	11.18	No
57	145	200	5000	1.2	25	25	10.68	No
58	155	150	4500	1.2	25	25	8.58	Yes
59	160	150	4500	1.2	25	25	15.49	Yes
60	130	125	4000	1	20	25	12.84	Yes

61	130	175	3000	0.8	16	50	13.55	Yes
62	135	100	3500	1	16	50	13.38	Yes
63	135	125	2500	1	20	50	13.45	Yes
64	135	125	3500	1	20	50	15.08	Yes
65	135	175	3500	1.2	20	50	16.15	Yes
66	135	175	4000	1	16	75	16.32	Yes
67	140	150	3500	1.2	20	50	17.13	Yes
68	140	175	4000	1	20	50	15.56	Yes
69	140	175	4500	1	20	50	15.93	Yes
70	140	200	2500	1	20	25	8.15	No
71	140	200	4000	0.8	20	50	15.95	Yes
72	145	125	3000	1.2	16	50	17.42	Yes
73	145	150	3500	0.8	16	50	15.48	Yes
74	145	150	4000	1	20	75	12.99	Yes
75	145	175	3000	1	16	25	16.73	Yes
76	145	175	4000	0.8	20	75	16.49	Yes
77	150	150	3000	1	20	50	7.75	No
78	150	175	3500	1	16	50	17.66	Yes
79	155	150	3000	1.2	20	50	10.98	No
80	144	166	3270	1.15	19	44	14.98	Yes
81	145	130	3090	1.1	17	29	17.41	Yes
82	150	141	3300	1.1	20	50	14.86	Yes
83	149	170	3160	1	20	35	14.43	Yes
84	142	126	3230	1.05	16	40	16.90	Yes
85	140	155	3200	1.2	17	33	13.57	Yes
86	149	148	3120	1	18	37	14.34	Yes
87	143	163	3340	1.15	19	31	17.32	Yes
88	148	137	3410	1.15	18	46	13.02	No
89	145	145	3490	1.2	19	27	18.45	Yes
90	141	134	3050	1.2	20	48	15.28	Yes
91	146	159	3450	1.05	16	38	16.76	Yes
92	141	174	3010	1	16	42	11.53	Yes
93	147	152	3380	1.2	17	25	16.71	Yes
94	147	151	3280	1.15	17	39	14.03	Yes
95	140	150	3500	1.2	20	50	15.96	Yes
96	140	150	3500	1	20	50	10.74	Yes
97	145	125	3000	1.2	16	50	16.62	Yes
98	150	175	3500	1	16	50	14.99	Yes
99	145	125	3500	1	20	50	10.07	Yes

Table S5: The process conditions and the *I-V* characteristics for the 10 top-performer (PCE ≥17% devices)

Process condition	Temperature [ºC]	Speed [cm/s]	Spray flowrate [mL/min]	Plasma height [cm]	Plasma gas flowrate [L/min]	Plasma duty cycle [%]	j_{SC} [mA/cm^2]	V_{OC} [V]	*FF* [-]	**PCE [%]**
11	145	125	3.5	1	25	50	23.9	1.047	0.68	**17.0**
43	135	150	3.5	1.2	20	25	23.7	0.919	0.78	**17.0**
44	135	150	4.0	0.8	20	50	22.4	1.007	0.77	**17.4**
46	135	175	4.0	1.2	20	50	22.2	0.997	0.77	**17.0**
67	140	150	3.5	1.2	20	50	22.7	0.980	0.77	**17.1**
72	145	125	3.0	1.2	16	50	23.0	0.985	0.77	**17.4**
78	150	175	3.5	1	16	50	22.7	1.023	0.76	**17.7**
81	145	130	3.1	1.1	17	29	23.2	0.975	0.77	**17.4**
87	143	163	3.3	1.15	19	31	21.7	1.000	0.80	**17.3**
89	145	145	3.5	1.2	19	27	23.6	1.004	0.78	**18.5**

Table S6: The selected comparison of the process conditions and device *I-V* characteristic parameters in the last batch of experimental iteration

ML Condition	Sample	Temperature [°C]	Speed [mm/s]	Spray flowrate [mL/min]	Plasma Height [cm]	Plasma Gas Flow [L/min]	Plasma DC [%]	j_{SC} [mA/cm^2]	V_{OC} [V]	*FF* [-]	PCE [%]
84	F14R	142	126	3.23	1.05	16	40	24.1	1.062	0.66	16.9
81	F19R	145	130	3.09	1.1	17	29	23.2	0.975	0.77	17.4
89	F22R	145	145	3.49	1.2	19	27	23.6	1.004	0.78	18.5
91	F28R	146	159	3.45	1.05	16	38	22.3	0.991	0.76	16.8

Supplemental Experimental Procedures

1. Sequential Learning Framework to Guide Iterative Learning Cycles

The sequential learning contains five steps in each iterative cycle as discussed in the main text. In addition, some key features used in this work has been summarized as follows.

- Initial sampling strategy: For a general framework, we start with experimental planning of process conditions with a Latin-hypercube sampling (LHS) method for the initial round. LHS offers good coverage of the entire parameter space, with as few shots on goal as possible (because experiments are expensive), yielding high information at the start of the learning.
- Manufacturing & testing: Perovskite solar cells are fabricated by the RSPP method at Stanford, and PCE is measured with a solar simulator under standard testing conditions.
- Model training*:* With the experimental data of the process parameters and the device PCEs, we train the regression model to learn the process-efficiency relation. The surrogate model selected was Gaussian Process (GP) regression with anisotropic Matern 5/2 kernel.
- Predicting & planning: The prediction results were evaluated by an acquisition function together with the constraint information, and therefore a new round of experiments was planned using Bayesian optimization (BO). We explain some details of the constraint information by film quality and the acquisition function choice in this section.

1.1 Perovskite Film Quality as Probabilistic Constraint

In the proposed BO framework in this work, visual assessment has been used a quick method to create a probabilistic constraint in the acquisition decisions. Fig. S1 shows the examples for low-quality and good-quality films with visual inspection. The low-quality film has a pale gray color on the edges (which is an indication of poor crystallization kinetics and a porous film). This low-quality film is also not very uniform and has many large holes. In contrast, the good film has a uniform brown color across the substrate.

To validate the film quality assessment, we made the low-quality films into devices and measured the power conversion efficiencies. The histogram of these device efficiencies is shown in Fig. S2, and they are all in the low-efficiency regime. None is above 15% efficiency.

1.2 Choice of Acquisition and Constraint Functions

a. Upper Confidence Bound

For the acquisition function of the objective function (*i.e.*, process-efficiency relation), we used the commonly used utility function named upper confidence bound (UCB) [1], which is also known as Negative Lower Confidence Bound (LCB) in the *Emukit* package and other literature. UCB achieves a better balance between exploration and exploitation in the acquisition process, in comparison to the other popular utility function named Expected Improvement [2]. The UCB function is defined as,

$$acq_{\mathrm{UCB}} = y_{\mathrm{pred}} + \beta \cdot \sigma_{\mathrm{pred}}$$

where β is the parameter that adjusts the relative weight to prediction uncertainty σ_{pred} over the prediction mean y_{pred}. In the term of balancing exploration and exploitation, higher β value leads to more exploration, while lower β value leads to more exploitation. In this work, we set β equals to 1, because we found β = 1 achieved a good balance in the previous works [3].

b. Probability of Satisfying the Constraint Criteria

For converting the film quality model and the regression model of the previous device data, we calculate the probability of exceeding a threshold value δ:

$$pr_{\text{cons}} = \text{CDF}\left(\frac{y_{\text{cons}} - \delta}{\sigma_{\text{cons}}}\right)$$

where CDF is the cumulative distribution function of normal distribution.

2. Apply the Sequential Learning to RSPP Experiments

2.1 Five Cycles of Learning

We have listed the different sampling methods in Table S1 for the different batches (*i.e.*, 5 experimental batches in total of 20 process conditions). The initial round and final round were guided by global LHS and local region search respectively (as explained in the main text). We recorded the relative humidity (RH) changes in the process chamber in the table to track a potential source for the variability, although we demonstrated the PCEs of RSPP devices were fairly independent of ambient RH [4].

2.2 Choice of Batch Size in Each Cycle

During sequential learning, the batch size in every iteration is an importance parameter to choose for Bayesian optimization. Although smaller batches can provide quicker feedback to revise the regression model, it could be much more costly (in term of time consumption for the experiments) since many processes are batch process. We estimated the time consumption for the device fabrication and listed them in Table S3. We only listed the key steps that consumes the most time; other steps, such as precursor preparation and data analysis, can be added as well.

Based on Table S3, we calculated the time needed for each batch and average time per substrate while varying the batch size. Fig. S3 gives a quantitative perspective of time consumption per batch and average time per substrate for different batch size. Especially observed in Fig. S3b, the average time per substrate is diminishing very quickly with increasing batch size. That is why the batch mode of Bayesian optimization (with local penalization algorithm) is more cost-effective than single-sample iteration when experiments are more time-consuming than computation. A batch size is chosen to achieve a good balance between low average time cost per substrate and the batch time to feedback the regression model. Because the most time-consuming step – thermal evaporation – can accommodate up to 20 substrates per batch, we used the batch size of 20 in this work. Note the sudden jump in both Fig. S3a and S3b at the batch size of 21.

2.3 Evaluation of Surrogate Regression Model

After every experimental batch being made, the new data were added to the training dataset to revise the regression model. Fig. S4 shows the model prediction vs the ground truth (*i.e.*, the experimentally measured) at each iteration.

2.4 Characterization Results of Perovskite PV Devices

a. All sampled experimental process conditions and solar cell device efficiencies

Table S4 shows the 100 experimental conditions, the PV device efficiency, and whether it was considered as good quality film.

b. The process conditions *vs I-V* characteristics of top-performing devices

We documented the process conditions and the *I-V* characteristics in Table S5 for the top performers (PCE ≥17%). The histogram distribution of these 10 process conditions (binned into specific levels) are shown in Fig. S5. The corresponding *I-V* curves are shown in Fig. S6.

Fig. S7 shows the spectral-dependent external quantum efficiency (EQE) of the device under process condition 89. We integrated the EQE curve with the AM1.5G solar spectrum and obtained a current density value of 23.0 mA/cm^2. The main reason for the discrepancy between the EQE-integrated current density and the short-circuit current density may possibly come from device degradation in the storage, insufficient light soaking prior to EQE measurement, slight spectral mismatch between AM1.5G and solar simulator and so on.

We selected several samples in the same batch as the best performing solar cells, and obtained the optical microscope images (under reflection mode) and the maps of light beam induced current. We found that non-uniformity is one of the key factors affecting both the LBIC and the optical images (see Figs. 8a & 8e and 8d & 8h). Please note that both thickness variations (which affect optical absorption) and carrier collection is reflected in the LBIC maps.

3. Compare Sequential Learning vs. Design of Experiment for Device Optimization

3.1. Deep Dive to the Entire Optimization Space with Random Sampling and Full Factorial Sampling

To understand the overall efficiency distribution of the entire parameter space, we ran both the random sampling of 100,000 process conditions and the full factorial sampling of 41,580 process conditions (see Table I in the main text). The predicted efficiency distributions are shown in Fig. S9. The full factorial sampling and the 100k random sampling have very similar distributions (with only minor difference for the range of 10 – 13%). Based on the 100k random samples, we estimated the top 5 percentile is around 0.8 at the normalized PCE (14.2%/17.7%) in the predicted PCE (marked by blue line in Fig. S9). This top one percentile mark is also used in Fig. 6 from the main text to compare different sampling methods. Again, note that the “teacher” model slightly underestimates the top performers. Therefore, we normalized the predicted efficiency by 17.7% in Figs. 6b and 6c (from the main text), which is the highest predicted efficiency among all the experimental conditions in this work.

3.2. Design of Experiment Method 1: Latin Hypercube Sampling (LHS)

Latin hypercube sampling is known as a low-discrepancy pseudo-random sampling method. The parameter space is stratified into a hypercube with *m* levels in every dimension, and samples are picked sequentially so that each parameter level is only sampled once. At every iteration except for the last sample, the parameters are chosen randomly in the unoccupied hypercubes. As shown in Fig. S10, the pseudo-randomness can result in different samples. When being run many times, the statistical distribution will converge to the same as the random sampling.

3.3. Design of Experiment Method 2: One Variable at a Time Sampling (OVATS)

The one-variable-at-a-time method is one of the most popular methods for the experimental planning in academic research. The idea is simply illustrated in Fig. S11. The initial round will sample all levels in one variable while the rest variables are set to randomly picked values, and the best condition can be found. Then, the first variable is fixed to that best condition while all levels in another variable are sampled. The sequence of the sampled variables is determined at random. This procedure iterates through all the variables, and it repeats until no improvement can be achieved. In that case, we subdivide the grid, and it restarts the sampling procedure (see Round 4 in Fig. S11).

3.4. Design of Experiment Method 3: Factorial Sampling in Progressive Grid Division (FS-PGS)

The full factorial sampling suffers from the curse of dimensionality, which means the number of samples becomes infeasible to achieve experimentally as the dimensionality increases. To sample m levels in n-variable optimization, we need to sample m^n conditions in a full grid. For example, 5 levels in 6-variable optimization result in 15625 conditions. In practice, we can use two-level grid sampling sequentially to achieve a more efficient design of the experiments. After each round, we rely on the location of the best device in the parameter space to define the new optimization grid in the next round.

An initial sampling (or Round 0) is started with a full two-level grid in each variable, *i.e.*, high (H) and low (L) conditions. The total number of samples in Round 0 is 2^n, where n is the total number of variables in the optimization (*e.g.*, six variables in this work). After Round 0 experiments, we see the best condition is found at the (L, H) condition. Therefore, the new reduced space of Round 1 is moved toward the best condition in Round 0. The parameter space keeps reducing after each round until the minimum sampling space is reached (see Round3 in Fig. S12e). After that, the sampling is repeated iteratively with the minimum sampling space, until the location of the best device is not changing (see Round 4 in Fig. S12f). In this case, the sampling space in the next round is increased to explore a broader region (see Round 5 in Fig. S12g). This is just one way to implement the factorial sampling with progressive grid subdivision, but the key concept is to exploit the regions where the best device is found.

3.5 Benchmarking: the BO Methods vs Model-Free Design of Experiment Methods

In addition to the Fig. 6 in the main text, we plotted the several virtual benchmarking results in full details with separate figures (Figs. S13-S15).

Supplemental References

1. Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger (2012). Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting. IEEE Transactions on Information Theory *58*, 3250–3265.

2. J. Mockus, V. Tiesis, and A. Zilinskas (1978). The application of Bayesian methods for seeking the extremum. In Towards Global Optimisation, L.C.W.Dixon and G.P. Szego, eds. (North-Holand).

3. Rohr, B., Stein, H.S., Guevarra, D., Wang, Y., Haber, J.A., Aykol, M., Suram, S.K., and Gregoire, J.M. (2020). Benchmarking the acceleration of materials discovery by sequential learning. Chemical Science *11*, 2696–2706.

4. Rolston, N., Scheideler, W.J., Flick, A.C., Chen, J.P., Elmaraghi, H., Sleugh, A., Zhao, O., Woodhouse, M., and Dauskardt, R.H. (2020). Rapid Open-Air Fabrication of Perovskite Solar Modules. Joule *4*, 2675–2692.