

Accelerating device characterization in perovskite solar cells via neural network approach

Xinhai Zhao ^{a,b,1}, Chaopeng Huang ^{a,e,1}, Erik Bergersson ^{b,c}, Nikita Suprun ^d, Hu Quee Tan ^{b,c}, Yurou Zhang ^a, Yuxia Jiang ^a, Chunhui Shou ^a, Jingsong Sun ^{a,*}, Jun Peng ^{a,e,*}, Hansong Xue ^{f,**}

^a Zhejiang Baima Lake Laboratory Co., Ltd, Hangzhou, Zhejiang 310000, China

^b Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore

^c Solar Energy Research Institute of Singapore (SERIS), National University of Singapore, 7 Engineering Drive 1, 117574, Singapore

^d School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

^e Institute of Functional Nano & Soft Materials (FUNSOM), Jiangsu Key Laboratory of Advanced Negative Carbon Technologies, Joint International Research Laboratory of Carbon-Based Functional Materials and Devices, Soochow University, Suzhou, Jiangsu 215123, China

^f Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, 487372, Singapore

HIGHLIGHTS

- A three-layered ANN can treat the highly non-linear drift-diffusion system in PSCs.
- Efficient NNs with MSEs $< 4 \times 10^{-4}$ are trained for predicting key indices in PSCs.
- Fast model calibration in 24 s accelerates device characterization with modeling.
- Experimental 2 % increase in PCE is achieved with tuning of SAMs and passivators.
- Tracking of recombination losses is enabled with NNs during in-house experiments.
- Optimization with NN finds a theoretical PCE of 28.9 % for 1.56 eV in-house PSCs.

ARTICLE INFO

Keywords:

Machine learning
Artificial neural network
Perovskite solar cell
Device characterization
Loss analysis
Optimization

ABSTRACT

Perovskite solar cells are promising candidates for next-generation high-efficiency photovoltaic devices, especially as top cells in tandem applications. Using a physical-based optoelectronic model, we collect big data of one hundred thousand sample size to train neural network models for efficient prediction of device performance and recombination losses. Latin hypercube sampling, Bayesian regularization, and Bayesian optimization are adopted for data preparation, model training, and optimization of the neural networks, respectively. The best neural network models achieved mean squared errors below 4×10^{-4} on a reserved testing dataset. The computational speed of the neural network is more than one thousand times faster than traditional optoelectronic models. As a result, fast device calibration can be conducted in twenty-four seconds. The reduced computational cost allows for efficient device characterization, parametric studies, sensitivity analysis, loss analysis, and optimization. After optimizing interface recombination in our in-house fabricated devices, we observed an experimental improvement of approximately 2 % in power conversion efficiency. Additionally, we predict theoretical power conversion efficiencies of 28.9 % and 25.5 % for perovskite solar cells with band gaps of 1.56 eV and 1.63 eV, respectively.

1. Introduction

With the rapid advancement of artificial intelligence, machine

learning and neural networks have gained prominence for solving complex problems, making accurate predictions, and aiding decision-making. The applications of machine learning span multiple

* Corresponding authors at: Zhejiang Baima Lake Laboratory Co., Ltd., Hangzhou, Zhejiang 310000, China.

** Corresponding author.

E-mail addresses: sunjingsong@zjenergy.com.cn (J. Sun), jpeng@suda.edu.cn (J. Peng), Hansong.xue@sutd.edu.sg (H. Xue).

¹ Xinhai Zhao (X.Z.) and Chaopeng Huang (C.H.) contributed equally and are co-first authors in this work.

disciplines, and in photovoltaics, they offer promising extensions to traditional models with greater computational efficiency.

Researchers have applied machine learning for modeling of perovskite solar cells. Several studies have focused on predicting key properties of perovskite materials, such as band gap [1–3], lattice parameters [4,5], formability [3,6,7], and stability [6], using various machine learning approaches. Beyond material properties, ongoing research also aims to predict the efficiencies and other metrics of perovskite solar cells. For instance, Ichiwani et al. [8] have explored the correlation between spray parameters and the efficiency of spray-coated perovskite solar cells with multiple linear/polynomial regressions and convolutional neural networks, predicting an optimal power conversion efficiency (PCE) of 29 %. Hu et al. [9] have used a dataset of around one hundred samples, incorporating five material and geometrical properties to predict five key metrics—open-circuit voltage (V_{oc}), short-circuit current density (i_{sc}), fill factor (FF), PCE, and stability—via regression models such as support vector machine (SVM) and Gaussian process regression (GPR). Their optimized perovskite solar cell has achieved a PCE of 23.4 % and demonstrated long-term stability.

Artificial neural network is an efficient machine learning algorithm that has been adopted to predict the performance of perovskite solar cells. For example, Oboh et al. [10] have trained artificial neural networks to predict the performance of planar perovskite solar cells using layer thicknesses and dopant concentrations as input parameters. They showed good agreement between the predicted efficiencies from neural networks and numerical simulations. Sun et al. [11] have applied neural networks to assist in the optical design of perovskite solar cells, and they have predicted an over 30 % enhancement in absorptance. Wu et al. [12] have developed a neural network to predict charge-carrier recombination rates in perovskite solar cells, providing insights into interface recombination, hysteresis, and ion migration. Tan et al. [13] have conducted a Pareto front optimization on a four-terminal perovskite-on-copper indium selenide tandem solar cell with a neural network trained with Bayesian regularization-backpropagation algorithm, and they have optimized the layer thicknesses for the best cost-efficiency.

The aforementioned works have advanced the use of artificial neural networks in modeling of perovskite solar cells. Nonetheless, the following aspects of applying neural networks to planar perovskite solar cells can be further improved. First, incorporating more input parameters, such as charge-carrier mobility, density of states, and energy band levels, could make a neural network better learn the electronics of a perovskite solar cell. The second point is about data sampling. A well-distributed dataset covering a wide range of parameters could help generalize the neural network, ensuring it is not tailored to a specific device. Lastly, optimizing the neural network is crucial to minimize the training error and prevent overfitting.

In this work, we address the above-mentioned aspects by training a physical-based neural network model for planar p-i-n perovskite solar cells, which will replace the full electronic model. The parameter space includes thirty-one inputs related to the device's electronics and fourteen outputs about cell performance and various recombination losses. To begin with, we prepare a set of evenly distributed training data of one hundred thousand samples from an electronic model, along with three additional datasets for testing and validation. Next, data processing, such as logarithmic scaling and normalization, are applied. Finally, we adopt Bayesian optimization to optimize the hyperparameters in the neural network, followed by training and testing.

We then employ an optics model that uses a vectorized matrix method and combine it with the obtained neural networks to form a surrogate model. The obtained model is calibrated and validated with four reported solar cells and five champion devices from in-house experiments, showing good agreements between simulated and experimental results. The time taken are approximately 24 s for model calibration and around 28 s for device optimization. Our objective is to provide a fast, reliable tool for characterizing and optimizing of planar perovskite solar cells, and hence efficiently guide decision making.

2. Mathematical formulation

We consider the following planar perovskite solar cell shown in Fig. 1(A), with schematics of its one-dimensional optical electrical field (Transverse Electro-Magnetic mode, N -layer thin-film stack) and energy band levels shown in Fig. 1(B) and 1(C), respectively. Here, the superscripts {G,F,H,P,E,B} stand for the glass, front-contact, hole-transporting, perovskite, electron-transporting, and back-contact layers, respectively, the roman numerals denote the interfaces between layers, l is the layer thickness, E_0 , E_R , E_T , $E_{0,b}$ are the incident, reflected, transmitted, and rear-incident optical electric field magnitudes, respectively, A_j and B_j describe the forward and backward optical electric field magnitude in the j -th thin-film layer, respectively, L_j is the total thickness from the origin to the end of j -th layer, E_{vac} , E_F , E_c , and E_v are the vacuum, Fermi, conduction band and valence band energy levels, respectively, W is the work function, and χ_e and χ_h are the electron affinity and hole ionization energy, respectively.

To simulate the $i - V$ characteristics of perovskite solar cells, we used to adopt the optoelectronic model from our previous works [14,15]. A flowchart of this full mathematical model is shown in Fig. 1(D); here, n is the refractive index, κ is the extinction coefficient, I_0 refers to the spectral irradiance of the front-incident light (ASTM G173-03 Global tilt standard spectrum [16], with the solar irradiance, P_m , being 1000 W/m²), R_a is the spectral albedo for the rear-incident light, N_d and N_a are the donor and acceptor impurity concentrations, respectively, ϵ is the permittivity, N_c and N_v are the density of states in the conduction and valence bands, respectively, μ is the charge carrier mobility, the subscripts {e, h} denote the properties associated with electrons and holes, respectively, v_{IV} and v_V are the net surface recombination velocities at interfaces IV and V, respectively, G is the charge carrier generation rate, B_{rad} is the radiative recombination coefficient, τ is the charge carrier lifetime, E_t is the energy level of the trap state, A_i is the auger recombination coefficient for species i , T is the operating temperature, V_a is the applied voltage, V_{oc} is the open-circuit voltage, FF stands for the fill factor and is denoted by $\tilde{\delta}$ later as well, η is the PCE, i_{rad} , i_{srh} , i_{aug} , i_s^{IV} and i_s^V are the current density losses from radiative recombination, Shockley-Read-Hall (SRH) recombination, auger recombination, and interfacial recombinations at interfaces IV and V, respectively.

A complete mathematical formulation of the optoelectronic model is provided in Table 1 with ion transport considered [17]. The optics part corresponds to Fig. 1(B); here, k_N and k_0 are the wave numbers in layer N and air, respectively, E_z is the electric field magnitude in the z-direction, Q is the time-averaged energy dissipation rate, λ is the wavelength of light, λ_u and λ_l are the upper and lower bounds for the wavelength, G_{avg} is the averaged charge-carrier generation rate, e is the elemental charge, c is the speed of light in vacuum, and ϵ_0 is the permittivity in vacuum. A typical perovskite solar cell includes a thick glass layer, and we consider spatially incoherent light propagation inside, where the phase information is destroyed, and no light interference effects occur. The phase elimination method [18] is adopted to deal with this incoherency in multilayer calculations.

For the electronics, j is the charge-carrier flux, the superscripts ' and " denote first and second order differentiations, respectively, the subscript I denotes mobile ions, the superscripts n and p indicate that the ions are negative or positive charged, respectively, ψ is the electric potential, c_e and c_h are the electron and hole concentrations, respectively, $c_{I,n}$ and $c_{I,p}$ are the negative and positive charged ion vacancies, respectively, i_{tot} is the total current density, V_b is the built-in voltage, c_h^0 and c_e^0 are the reference hole and electron concentrations at interfaces III and VI, respectively, the subscripts {+, -} denote properties below and above an interface, respectively, Π is the jump in charge carrier concentration, j_{IV} and j_V are the interfacial recombination fluxes, D is the diffusion coefficient, k_B is the Boltzmann constant, R_{aug} , R_{rad} and R_{srh} are the auger, radiative, and SRH recombination rates, respectively, c_i is the intrinsic charge-carrier concentration, $c_{(e,h),trap}$ is the electron or hole

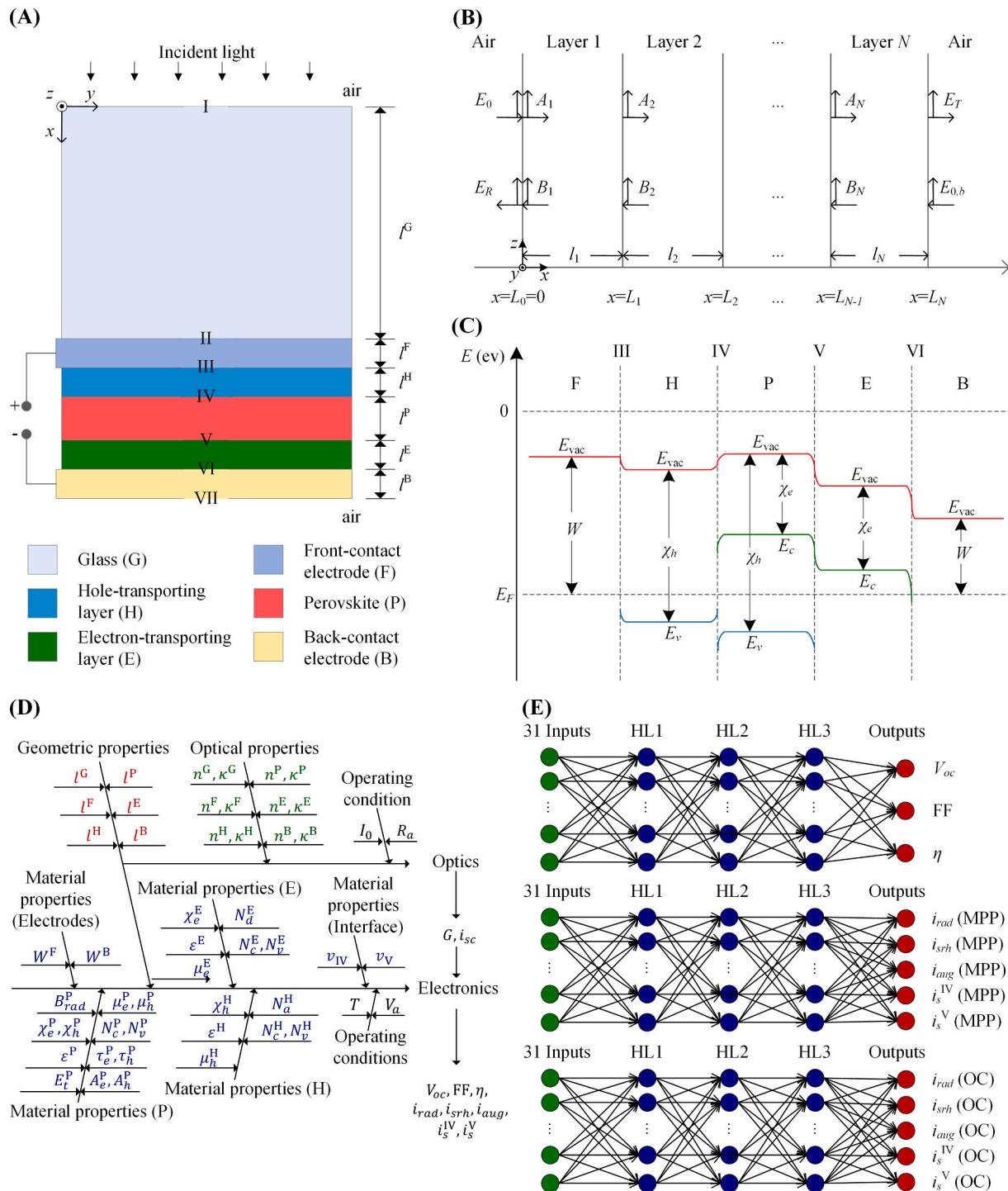


Fig. 1. Schematics of (A) a p-i-n type planar perovskite solar cell; (B) optical electric field in a N -layer thin-film stack; here, the vertical arrows represent the electric field magnitudes, and the horizontal arrows represent the direction of light propagation; (C) energy band diagram of the p-i-n type planar perovskite solar cell; (D) input and output parameters of the full optoelectronic model for planar perovskite solar cells; and (E) Neural network architectures for a planar perovskite solar cell with three hidden layers—HL1, HL2, and HL3. The outputs are the device performances, recombination losses at MPP, and recombination losses at OC. Here, not all the nodes in the inputs and hidden layers are shown for the sake of simplicity.

concentration at the trap state, and P_{in} and P_{out} are the input and output power densities, respectively.

The electronic model discussed above is employed to generate big data for training artificial neural networks, which are then integrated with the optics to create a coupled model. These networks predict device performances, recombination losses at maximum power point (MPP), and recombination losses at open circuit (OC), as shown in Fig. 1(E).

During collecting of big data later the model calibration and validation, we have considered perovskite solar cells without hysteresis effect [19] shown in their $i - V$ curves, and assumed that ion transport is not leading order. We will later detail the neural network training process, discuss the model performances, and demonstrate its application for rapid device characterization.

Table 1

Vectorized matrix formulation for the optics, together with the governing equations, boundary conditions, and constitutive relations for the electronics in p-i-n type planar perovskite solar cells. Here, the value of N is 6 for a five-layered thin-film stack with a glass layer in front.

Optics	Electronics
$\mathbf{X} = \mathbf{A} \setminus \mathbf{B}$, $\mathbf{A} = \begin{bmatrix} k_0 + k_1 & k_0 - k_1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \mathbf{S}_1 & \mathbf{S} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \mathbf{S}_2 & \mathbf{S} & \ddots & \vdots & 0 & 0 \\ 0 & 0 & \vdots & \ddots & \ddots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \mathbf{S}_{N-1} & \mathbf{S} \\ 0 & 0 & 0 & 0 & \mathbf{S}_{N-1} & \mathbf{S} \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & s_1 & s_2 \end{bmatrix},$ $\mathbf{X} = \begin{bmatrix} B_1 \\ \vdots \\ A_N \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix},$ $B_N = \begin{bmatrix} 2k_0 E_0 \\ 2k_0 E_{0,b} \end{bmatrix}$, $S_n = \begin{bmatrix} \exp(-ik_nl_n) & \exp(ik_nl_n) \\ (k_n/k_{n+1})\exp(-ik_nl_n) & -(k_n/k_{n+1})\exp(ik_nl_n) \end{bmatrix},$ $s_1 = (k_0 - k_N)\exp(-ik_Nl_N), s_2 = (k_0 + k_N)\exp(ik_Nl_N),$ $k_0 = 2\pi/\lambda, k_n = k_0(n_n - ik_n),$ $E_z = A_n \exp(-ik_n(x - L_{n-1})) + B_n \exp[ik_n(x - L_{n-1})],$ $E_0 = \sqrt{2I_0/ce_0}, E_{0,b} = \sqrt{2R_a I_0/ce_0},$ $E_R = A_1 + B_1 - E_0,$ $E_T = A_N \exp(-ik_Nl_N) + B_N \exp(ik_Nl_N),$ $Q = \frac{1}{2} ce_0 \alpha n E_z ^2, \alpha = 4\pi\kappa/\lambda,$ $G = \frac{1}{\hbar c} \int_{\lambda_l}^{\lambda_u} (\lambda Q) d\lambda,$ $i_{sc} = e \int_p G dx,$ $G_{avg} = i_{sc} / (eI^p).$	<p><i>Governing equations</i></p> $\dot{J}_{(e,h)} = S_j, \dot{J}_{I,(n,p)} = 0, e\psi'' = S_\psi, S_j = \begin{cases} 0, (H, E) \\ G_{avg} - R, (P) \end{cases}, S_\psi = e \begin{cases} N_a - c_h, (H) \\ c_e - c_h \pm c_{I,(n,p)}, (P) \\ c_e - N_d, (E) \end{cases}, i_{tot} = -e(j_e - j_h \pm j_{I,(n,p)}).$ <p><i>Boundary conditions</i></p> $\psi^{III} = V_a - V_b, \psi^{VI} = 0, c_h^{III} = c_h^0, c_{h,+}^{IV} = \Pi_h c_{h,-}^{IV}, c_{e,-}^V = \Pi_e c_{e,+}^V, c_e^{VI} = c_e^0, \dot{J}_{h,-}^V = \dot{J}_{h,+}^V - j_{IV}, \dot{J}_{e,+}^V = j_{IV}, \dot{J}_{I,(n,p)}^V = 0, \dot{J}_{e,-}^V = \dot{J}_{e,+}^V - j_{IV}, \dot{J}_{h,-}^V = j_{IV}, \dot{J}_{I,(n,p)}^V = 0.$ <p><i>Constitutive relations</i></p> $j_{(e,h)} = \pm \mu_{(e,h)} c_{(e,h)} \psi' - D_{(e,h)} c'_{(e,h)}, j_{I,(n,p)} = \pm \mu_{I,(n,p)} c_{I,(n,p)} \psi' - D_{I,(n,p)} c'_{I,(n,p)}, D_{(e,h)} = k_B T \mu_{(e,h)} / e, V_b = (W^F - W^B) / e, c_e^0 = N_v^H \exp[(\chi_e^H - W^F) / (k_B T)], c_e^0 = N_c^E \exp[(\chi_e^E - W^B) / (k_B T)], \Pi_h = (N_v^P / N_v^H) \exp[(\chi_h^H - \chi_h^P) / (k_B T)], \Pi_e = (N_c^P / N_c^E) \exp[(\chi_e^P - \chi_e^E) / (k_B T)], R = R_{aug} + R_{rad} + R_{srh}, R_{aug} = (A_e c_e + A_h c_h) (c_e c_h - c_i^2), R_{rad} = B_{rad} (c_e c_h - c_i^2), R_{srh} = \frac{c_e c_h - c_i^2}{\tau_e (c_h + c_{h,trap}) + \tau_h (c_e + c_{e,trap})}, j_{IV} = \nu_{IV} [c_{e,+}(c_{h,-} + c_{h,+}) - c_i^P (c_i^H + c_i^P)], j_V = \nu_V [c_{h,+}(c_{e,+} + c_{e,-}) - c_i^P (c_e^E + c_i^P)], i_{avg} = e \int_p R_{avg} dx, i_{rad} = e \int_p R_{rad} dx, i_{srh} = e \int_p R_{srh} dx, i_{IV} = e j_{IV}, i_V = e j_V, c_i = \sqrt{N_v N_c \exp[-E_g / (k_B T)]}, E_g = \chi_h - \chi_e, c_{(e,h),trap} = N_{(c,v)} \exp[-\Delta E_{(e,h)} / (k_B T)], \Delta E_e = E_g - E_s, \Delta E_h = E_t, \eta = V_{oc} i_{sc} \tilde{\delta}, \tilde{\delta} = \max(P_{out}/P_{in}, P_{out}) / V_a i_{tot}.$

3. Methodology

3.1. Data preparation

For modeling of typical p-i-n structured perovskite solar cells, we consider in total thirty-one input parameters, as summarized in Table S1. The parameter values are based on our previous works [13–15,20–24], and sufficient margins at the lower and upper bounds are added for training a more holistic model. Thirteen outputs are then generated from the optoelectronic model, including V_{oc} , FF, PCE, and various recombination losses.

For training of the artificial neural networks, we prepare one hundred thousand samples as training data based on Latin Hypercube sampling method. This large sample size ensures good generalization of a three-layered neural network shown in Fig. 1(E) [25]. Here, for parameters with more than one order of magnitude difference between their lower and upper bounds, we use a logarithmic scale instead of a linear scale for even sampling across a large range. Three additional datasets, each with a sample size of ten thousand but different seeds, are prepared as the testing data (testing data A), the validation data, and another reserved testing data (testing data B). After data generation, we remove samples with misaligned energy band levels that would impede smooth charge-carrier transport; i.e.

$$\chi_h^p < \chi_h^H < W^F \quad (1)$$

$$W^B < \chi_e^E < \chi_e^P \quad (2)$$

Here, we further remove samples with numerical issues, such as convergence problems where the input parameters fail to return a valid $i - V$ characteristic. As a result, the final sample sizes are 54,374, 5407, 5416, and 5479 for the training data, validation data, testing data A, and testing data B, respectively. We then normalize all inputs and outputs to a range of -1 to 1 due to large order of magnitude differences among the parameters.

3.2. Training of conventional regression models

Before training the neural networks, we evaluate whether conventional regression models in machine learning could well predict the outputs. We train four types of models in MATLAB—fine tree, interaction linear regression, cubic SVM, and isotropic Matern 5/2 GPR—with the prepared training data, each targeting three different outputs— V_{oc} , FF, and PCE. A five-fold cross-validation is applied to each training process. We apply Bayesian optimization to determine the hyperparameters—box constraint and epsilon for training the SVM model. For the GPR model, we optimize its kernel scale and sigma value, and set its basic function to be constant. The mean squared error (MSE) between the normalized predicted and true values is recorded with the reserved testing data B to assess the model performance.

3.3. Training of neural network

As shown in Fig. 1(E), we train three neural networks—A, B, and C—each with a specific output group: device performance metrics (V_{oc} , FF, and PCE), recombination losses at MPP, and recombination losses at OC, respectively. With a known $i - V$ curve, model calibration and parameter extraction can be performed using neural network A alone, while neural networks B and C enable further device characterization, such as loss analysis.

Training of the three-layered neural networks is conducted in MATLAB with Bayesian regularization [26], which automatically updates and optimizes the weights and bias. Instead of randomly splitting the data, we fix the training data, validating data, and testing data (testing data A) for even sampling and consistent training results. The reserved unseen dataset, testing data B, is used to assess model performance and check for overfitting. The number of epochs is set to be one thousand for thorough training with Bayesian regularization. The transfer function at the output layer is set to be linear (purelin). An early stopping technique is considered here with a maximum number of failures—number of successive iterations when the MSE for the validating data increases while that for the training data decreases—being ten to prevent overfitting. The rest of the hyperparameters, including the number of neurons, transfer functions, initial learning rate, momentum, and minimum gradient are determined through Bayesian optimization.

3.4. Optimization of hyperparameters

We employ Bayesian optimization in MATLAB to find the hyperparameters. For the SVM model, the lower and upper bounds for box constraint are 10^{-3} and 10^3 , respectively, and epsilon ranges from 3×10^{-4} to 30. For the GPR model, kernel scale ranges from 2×10^{-3} to 2, and sigma is searched among real values in the range 10^{-4} to ten times the standard deviation of the training data. The objective function is the training MSE. Termination conditions are set at a maximum of 100 objective iterations or two days of runtime for SVM and four days of runtime for GPR.

Regarding hyperparameters in the neural networks, for the hidden layer sizes, the minimum and maximum number of neurons are set to be 2 and 100, respectively. The initial learning rate, minimum gradient, and momentum have their lower bounds at 0.01, 10^{-8} , and 0.5, respectively, and upper bounds at 1, 10^{-6} , and 0.98, respectively. For the transfer functions in the three hidden layers, we select from common options in regression problems, and they are hyperbolic tangent sigmoid (tansig), radial basis (radbas), log-sigmoid (logsig), and linear (purelin). The objective function is the training MSE. Termination conditions are set at a maximum of 10^4 objective iterations or seven days of runtime, with 500 initial seed points. The Bayesian optimization is coupled with neural network training to identify the best-performing design.

4. Numerics

We adopt the finite element method to solve the differential equations for the electronics part in COMSOL MULTIPHYSICS 5.2a. The direct solver MUMPS is selected as the nonlinear solver with a 10^{-3} relative convergence tolerance. The computer used to run the simulations has a CPU processor Intel® Core™ i5-7600 @ 3.50 GHz and a total installed RAM of 16.0 GB. The computational domains for charge carrier transport—hole-transporting, perovskite, and electron-transporting layers—are solved with a mesh size of 1 nm in the domains and 0.1 nm at the boundaries. The time taken for one thousand rounds of $i - V$ simulations, each with 45 voltage points, is around 4800 s (wall-clock time). The $i - V$ data are exported from COMSOL to MATLAB for postprocessing.

5. Results and discussions

5.1. Data distribution

Data quality are crucial for neural network design and optimization. We have generated well-distributed input samples based on Latin Hypercube sampling method. The distributions of all training data are shown in Fig. 2. A full list of inputs and outputs is provided in Table S1, and their distributions after normalization are shown in Figure S1.

As shown in Fig. 2(A-J), most input data remain evenly distributed after postprocessing (as mentioned in section 3.1). An exception is the average generation rate, G_{avg} , which is calculated from evenly distributed short-circuit current densities and perovskite layer thicknesses (see Table 1). This results in a narrowed interquartile range, resembling a normally distributed dataset in a logarithmic scale, as illustrated in Fig. 2(F).

For the output data shown in Fig. 2(K-N), the V_{oc} approximate a normal distribution, the FFs are close to an even distribution, the PCEs are positively skewed, and the recombination losses are negatively skewed. The maximum PCE reaches 42.1 %, which has far exceeded the Shockley-Queisser limit. Having said that, we have set the ranges of input data to be wide enough for training a more general neural network. As such, the maximum short-circuit current density is set to be 360 A/m^2 , which leads to broad ranges of the output data.

5.2. Regression model performances

Conventional regression models in machine learning, such as fine tree, linear regression, SVM, and GPR are commonly employed to address regression problems. From Bayesian optimizations with the prepared data, we have determined the hyperparameters for SVM and GPR models. For SVM models, the box constraints are found to be 1.89, 6.37, and 2.05 and the epsilons are 9.03×10^{-3} , 1.13×10^{-2} , and 7.23×10^{-3} for predicting the V_{oc} , FF, and PCE, respectively. For GPR models, the kernel scales are optimized to be 1.48, 0.446, and 1.00 and the sigma are 5.28×10^{-3} , 2.04×10^{-2} , and 2.22 for predicting the V_{oc} , FF, and PCE, respectively. For the four regression models, their capabilities to predict the device performance are demonstrated in Fig. 3.

We find that the Matern 5/2 GPR models, in overall, exhibit the lowest testing MSEs between true and predicted values among the four models. However, we observe serious overfitting problem in the GPR model that predicts the PCE. As shown in Fig. 3(L), the unseen testing data points (red) deviate from the diagonal line while the training data points (blue) stay close to the diagonal line, even with a five-fold cross-validation applied during model training. Furthermore, both cubic SVM and Matern 5/2 GPR models require significantly longer training times—approximately two hours for the SVM model and over ten hours for the GPR model—compared to less than two minutes for the fine tree and linear regression models.

Among the three outputs, the predicted values for V_{oc} show the lowest discrepancies from the true values, as indicated by the overall good clustering of data around the diagonal lines. This could be due to the fact that at OC, an additional relation,

$$\int_P (G - R) dx = 0 \quad (3)$$

holds, which pulls the current density to zero and drastically reduces the complexity of the electronics model. In contrast, the models struggle to accurately predict the FF and, consequently, the PCE, with absolute differences between predicted and true values exceeding one. We attribute this to increased non-linearity at MPP. Given these substantial errors, it is necessary to pursue a more advanced method, such as an artificial neural network, for this regression problem.

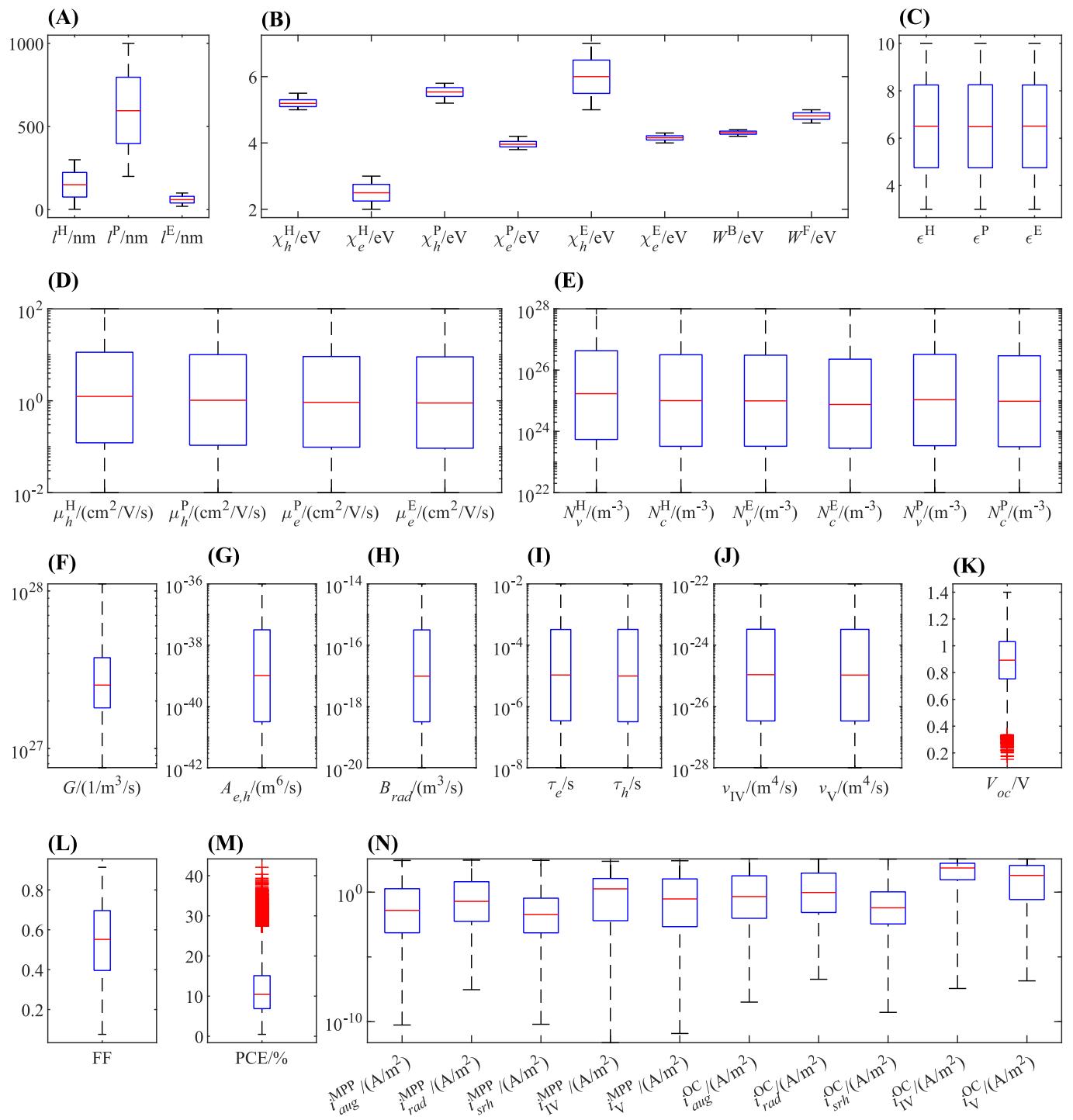


Fig. 2. Distributions of the (A–J) input and (K–N) output parameters in the training data.

5.3. Neural network performance

We verify the optimized neural networks by comparing their predicted outputs with the actual outputs from optoelectronic model. For neural network A that predicts V_{oc} , FF and PCE, the verification results are shown in Fig. 4(A–C), with the MSEs of the reserved testing data B indicated. The corresponding optimized hyperparameters from Bayesian optimization, along with the neural network performances, are summarized in Table 2.

We find that neural network A predicts the outputs significantly better than the conventional regression models. The coefficients of

correlation are close to 1, with MSEs at an order of magnitude of 10^{-4} for both training and testing data. Comparing to the MSEs from 1.8×10^{-3} to 8.3×10^{-2} for the regression models, the errors are about ten times smaller for the neural networks. The predicted and actual data are close to each other, with only a few outliers observed for the FF and no obvious overfitting problem. When looking at the absolute errors (AE) after reversing the normalization, 95 % of the training data AND 94 % of the testing data have $AE \leq 0.01$ V for V_{oc} ; 88 % of the training data and 79 % of the testing data have $AE \leq 0.01$ for FF; and 98 % of the training data AND 95 % of the testing data have $AE \leq 0.5\%$ for PCE.

Additionally, we have trained three separate neural networks for

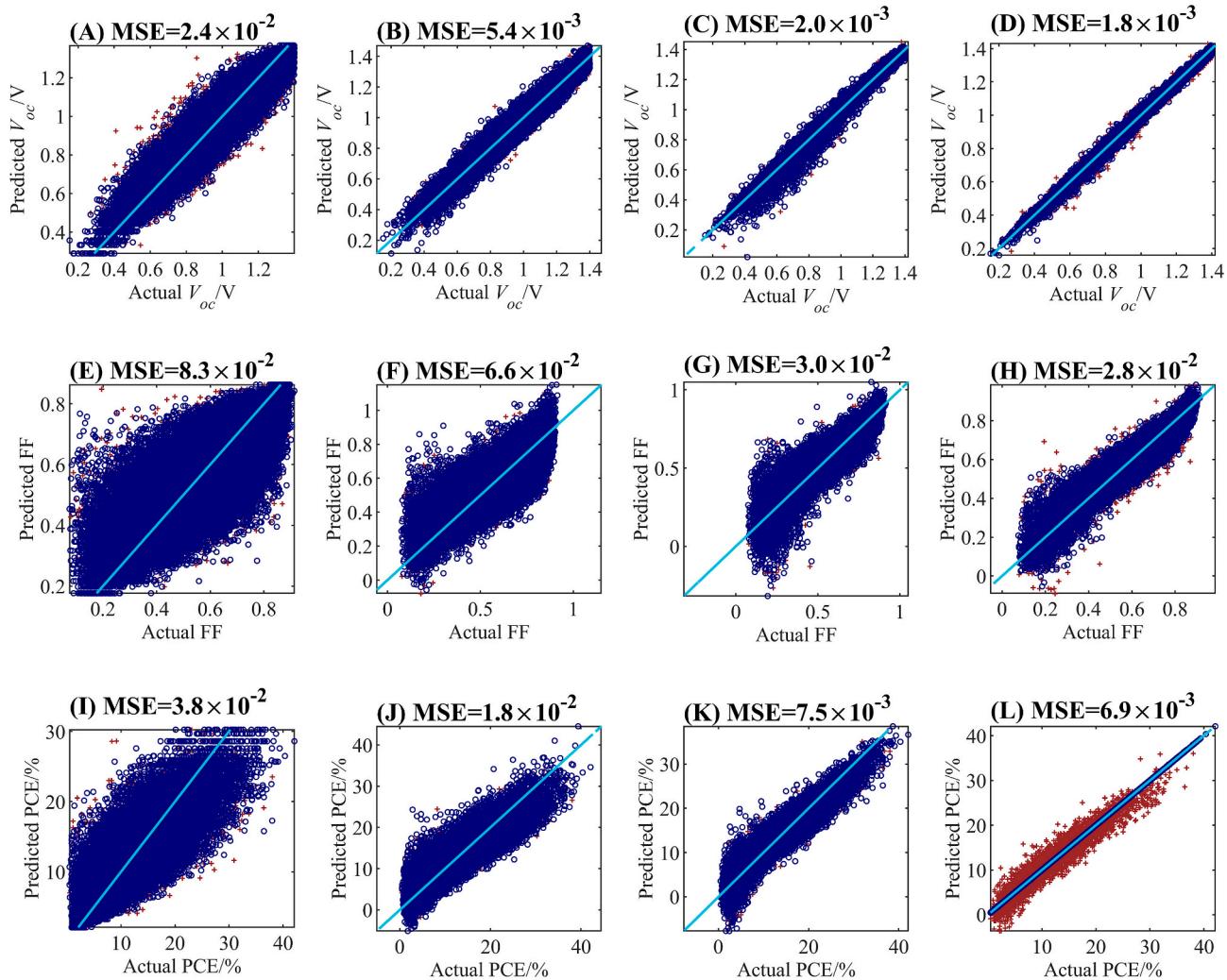


Fig. 3. True versus predicted values for (A–D) V_{oc} , (E–H) FF, and (I–L) PCE from (A, E, and I) fine tree, (B, F, and J) interaction linear regression, (C, G, and K) third order SVM, and (D, H, and L) Matern 5/2 GPR models with the testing MSEs indicated. Here, the blue circles represent the training data, the red plus signs represent the testing data (reserved testing data B), and the cyan diagonal line indicates perfect matching of predicted and true values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each output— V_{oc} , FF and PCE—whose performances are shown in Figure S2. Comparing to neural network A, smaller MSEs of 7.8×10^{-6} are achieved for predicting the V_{oc} . The MSEs for predicting the FF and PCE, however, are increased slightly to 8.4×10^{-4} and 1.5×10^{-4} , respectively. We attribute this larger MSE to the absence of covariance among the outputs. Instead of these separate models, neural network A is more efficient for model calibration and device characterization, as it accounts for parameter correlations.

The above results indicate the overall good performance of neural network A, which we attribute to several factors. First, Bayesian regularization with backpropagation effectively handles highly nonlinear regression by automatically updating weights and biases. Next, instead of k-fold cross-validation, we use a separate validation dataset of ten thousand samples, evenly distributed with input ranges matching the training set, to check for overfitting every epoch. This validation mimics unseen data due to a different seed used during data sampling. Additionally, we apply early stopping with a failure threshold of ten, effectively reducing overfitting (see Figure S3 for a comparison without early stopping). These methodologies are sufficient to prevent overfitting problem in this work. While dropout layers may further mitigate overfitting, they are omitted to avoid extended training time [27].

In terms of computational speed, time taken for our neural network to conduct one thousand rounds of simulations is about 4.5 s (wall-clock

time). When with vectorization, only about 0.3 s (wall-clock time) is required for the neural network to finish the computation. Comparing to the full optoelectronic model that requires around 4800 s, this new model is estimated to be more than one thousand times faster. Therefore, we expect significant reductions in the time taken for model calibration and large-scale optimizations.

Moreover, a neural network capable of estimating the recombination losses can provide more insights into device characterization. Since artificial neural network models are much more computationally efficient than traditional optoelectronic models, they enable quick device calibration and loss analysis. For neural network B that predicts the recombination losses at MPP, the optimized hyperparameters and its performance are shown in Table 3 and Fig. 4(D–H). The neural network achieves a coefficient of correlation close to 1, with MSEs of less than 4×10^{-4} .

In addition, we have trained one more neural network C that predicts the recombination losses at OC. The optimized hyperparameters and performance of this third neural network are summarized in Table 4. Among the trained models, neural network C has the best coefficients of correlation. Excellent agreement between the true and predicted values has been achieved, as shown in Fig. 4(I–M), supported by low MSEs of smaller than 4×10^{-5} .

Beyond low training and testing errors, a key advantage of these

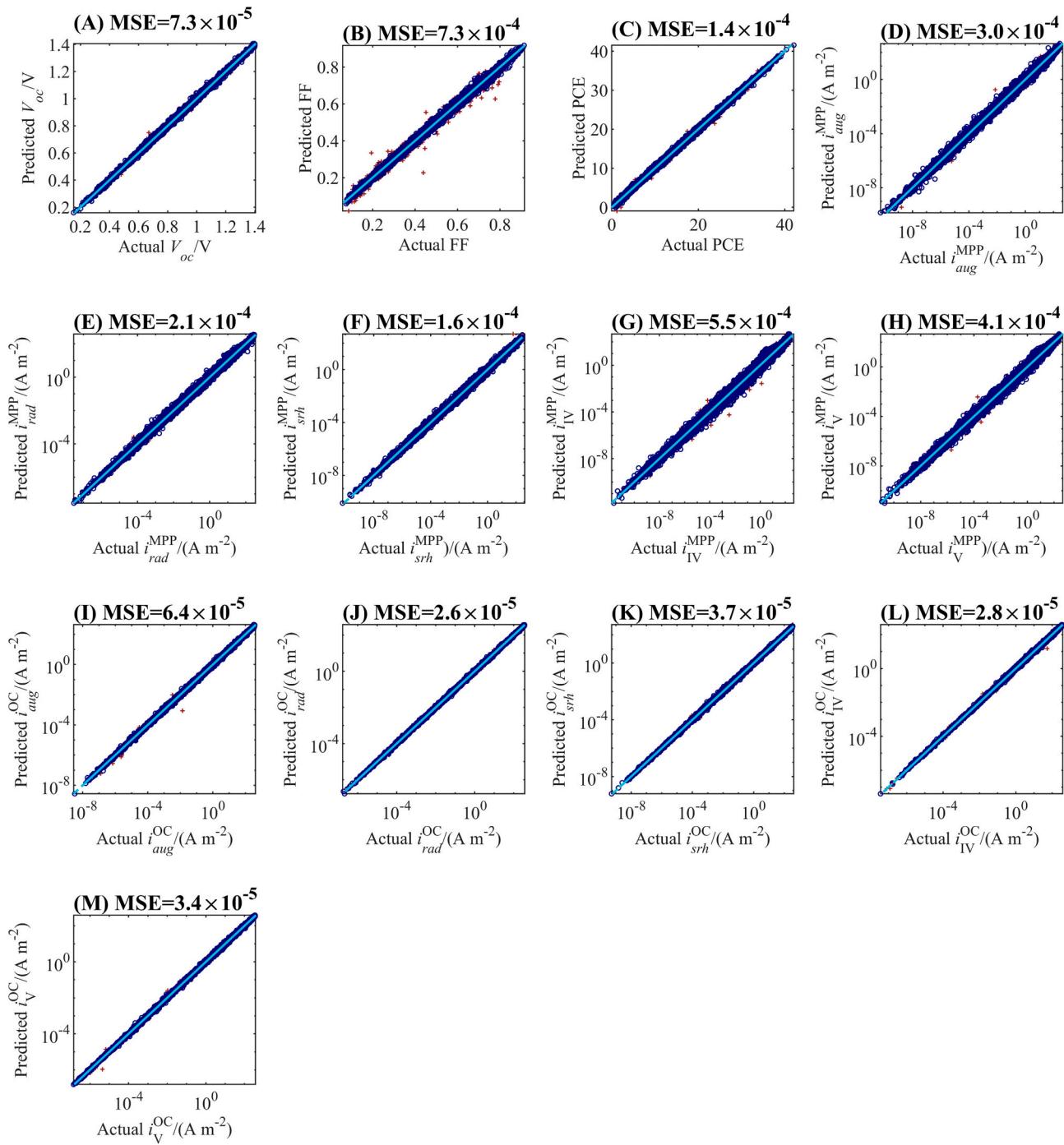


Fig. 4. Actual values from full optoelectronic model and results from neural network A for (A) V_{oc} , (B) FF, (C) PCE, together with Auger, radiative, and SRH recombination losses, and surface recombination losses at interfaces IV and V at (D–H) MPP and (I–M) OC. The blue circles and the red plus signs represent the training data AND the testing data (testing data B), respectively, and the cyan diagonal lines contain all points when the predicted and actual values are equal to each other. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

neural networks is their scalability. With wide parameter bounds (Table S1), including material properties, recombination coefficients, and layer thicknesses, the model can adapt to new materials and fabrication techniques. While the models are trained specifically on p-i-n structured perovskite solar cells, they could also apply to n-i-p structures by incorporating additional experimental or simulated data. These data should focus on energy band levels, charge carrier mobility, and density of states for typical electron- and hole-transporting materials in n-i-p devices. After data preparation, by incorporating classification techniques or integer-based classifiers and retraining, the model could

generalize to both structures, enhancing its versatility.

Moreover, with a wide band gap range of 1.0 to 2.0 eV, the neural networks can confidently be applied to four-terminal all-perovskite tandem solar cells. Since the subcells in four-terminal tandems are electrically decoupled, they can be characterized individually using our model. In contrast, two-terminal all-perovskite device, which requires an interconnecting layer and current matching between subcells, would need a different physical model [20] and new data. Nonetheless, our current model remains applicable to standalone perovskite subcells in two-terminal tandem devices.

Table 2

Optimized hyperparameters and performances for neural network A. Here, N_1 to N_3 are the number of neurons in hidden layers 1 to 3.

Hyperparameters	Results	Neural network performances	Results
Number of epochs	162	Coefficient of correlation (training data)	0.9993
Initial learning rate	0.0118	Coefficient of correlation (validation data)	0.9983
Minimum gradient	7.83×10^{-8}	Coefficient of correlation (testing data A)	0.9986
Momentum	0.540	Mean squared error (training data)	1.4×10^{-4}
HL1 transfer function	tansig	Mean squared error (validation data)	3.0×10^{-4}
HL2 transfer function	logsig	Mean squared error (testing data A)	2.7×10^{-4}
HL3 transfer function	radbas	Mean squared error (testing data B)	3.1×10^{-4}
N_1	65		
N_2	55		
N_3	75		

Table 3

Optimized hyperparameters and performances for neural network B.

Hyperparameters	Results	Neural network performances	Results
Number of epochs	451	Coefficient of correlation (training data)	0.9981
Initial learning rate	0.0135	Coefficient of correlation (validation data)	0.9973
Minimum gradient	7.08×10^{-8}	Coefficient of correlation (testing data A)	0.9975
Momentum	0.727	Mean squared error (training data)	2.6×10^{-4}
HL1 transfer function	logsig	Mean squared error (validation data)	3.6×10^{-4}
HL2 transfer function	radbas	Mean squared error (testing data A)	3.3×10^{-4}
HL3 transfer function	tansig	Mean squared error (testing data B)	3.3×10^{-4}
N_1	61		
N_2	39		
N_3	56		

Table 4

Optimized hyperparameters and performances for neural network C.

Hyperparameters	Results	Neural network performances	Results
Number of epochs	630	Coefficient of correlation (training data)	0.9998
Initial learning rate	0.0194	Coefficient of correlation (validation data)	0.9997
Minimum gradient	6.08×10^{-8}	Coefficient of correlation (testing data A)	0.9997
Momentum	0.873	Mean squared error (training data)	2.4×10^{-5}
HL1 transfer function	logsig	Mean squared error (validation data)	3.7×10^{-5}
HL2 transfer function	radbas	Mean squared error (testing data A)	3.6×10^{-5}
HL3 transfer function	tansig	Mean squared error (testing data B)	3.8×10^{-5}
N_1	40		
N_2	46		
N_3	96		

With its low computational cost, the model can calibrate key physical parameters every 30 s during repetitive $i - V$ measurements, facilitating real-time tracking of material properties such as charge carrier mobility, lifetime, and energy band levels for stability analysis. As an extension, our neural network model can be adopted to provide more insights to

device stability by incorporating density function theory [9], and this could involve retraining the neural networks with molecular dynamics parameters as inputs and stability as an additional output. Furthermore, when our model is integrated with device fabrication conditions during controlled experiments, it forms a two-step Bayesian inference framework [28]. The first step links process conditions to material properties, while the second connects these properties to device performance. This approach enables the model to interpret the impact of controlled process variables on device performance and, in turn, provide valuable insights into the optimal fabrication conditions through Bayesian network-based process optimization.

As another outlook, incorporating ion effects could enhance our work in the future. On top of our model that has assumed ion migration to be insignificant, several considerations are needed for accounting the ions. First, ion migration depends on the perovskite's chemical composition, with different ions having varying activation energies [19]. Since perovskites contain mixed cations, halogen anions, and impurities, a mixed-integer approach or classification strategy may be necessary for data preparation and training. Second, outputs should account for $i - V$ hysteresis by including open-circuit voltages and fill factors from both forward- and backward-scan curves. Third, applied potential influences ion migration, requiring more data points or gradients from $i - V$ curves. If needed, a more computationally intensive convolutional neural network could replace the current model. Ultimately, experimental data with clear hysteresis effects will be crucial for calibration and validation, ensuring a meaningful extension of this work.

5.4. Fast model calibration and device characterization

In solar cell applications, comparing to full $i - V$ curves, i_{sc} , V_{oc} , FF, and PCE are more commonly used to compare the performances across various devices. Here, neural network A is validated with experimental results from reported perovskite solar cells. Subsequently, it can be used for rapid characterization of physical parameters and estimation of recombination losses in conjunction with the neural networks B and C.

For model calibration, we take four reported perovskite solar cells adopted in tandem applications with varies band gaps—stand-alone top sub cell A1 [29], stand-alone bottom sub cell A2 [29], stand-alone top sub cell A3 [30], and stand-alone bottom sub cell A4 [31], and source the corresponding material properties from literature. The remaining unknown parameters regarding device recombination rates are either assumed or fitted. In essence, we fit the parameters B_{rad} , v_{IV} , and v_V for each cell by minimizing the following calibration MSE:

$$\text{MSE} = \left[(\hat{V}_{oc} - \hat{V}_{oc}^{\exp})^2 + (\hat{\delta} - \hat{\delta}^{\exp})^2 + (\hat{\eta} - \hat{\eta}^{\exp})^2 \right] / 3 \quad (4)$$

where \hat{V}_{oc} , $\hat{\delta}$, and $\hat{\eta}$ are the normalized open-circuit voltage, FF, and PCE, respectively, and the superscript \exp indicates that the value is measured from experiment. The particle swarm optimization algorithm is employed for this minimization problem in MATLAB. Default options with a swarm size of two hundred are applied. Full lists of parameters are presented in Table S2 to S5 for Cells A1 to A4, together with their corresponding device architectures.

Here, certain assumptions have been made. The auger recombination coefficients are set at the lower bound, assuming that auger recombination is not the dominant loss mechanism. The charge carrier lifetimes for electrons and holes are assumed to be about 10^{-5} s [32], which can be tuned according to measurements [33]. With these assumptions, the number of fitting parameters is reduced to three, corresponding to the number of terms used to calculate the calibration MSE in Eq. 4.

As summarized in Table 5, the MSEs for calibrations of cells A1 to A4 range from 10^{-8} to 10^{-6} , and the V_{oc} , FF, and PCE agree with experiments well. The time taken for each calibration is about 24 s (wall-clock time), while the time taken with the optoelectronic model is around three days [14]. Calibration and characterization with the trained neural network model are therefore roughly more than one thousand times

Table 5

Cell performances from experiments and calibrated neural network models.

Cell index	Band gap/eV	Experiments			Predicted			MSE
		V_{oc}/V	$\tilde{\delta}$	$\eta/\%$	V_{oc}/V	$\tilde{\delta}$	$\eta/\%$	
A1	1.75	1.21	0.768	16.1	1.21	0.768	16.1	8.2×10^{-8}
A2	1.22	0.83	0.808	21.1	0.83	0.808	21.1	5.5×10^{-7}
A3	1.79	1.26	0.791	17.8	1.26	0.791	17.8	2.6×10^{-7}
A4	1.25	0.84	0.724	17.5	0.84	0.725	17.5	1.6×10^{-7}

faster than an optoelectronic model.

To further validate the model, we import the parameters from Table S2 to S5 back to COMSOL to generate the $i - V$ curves and compare them with experiments. Fig. 5(A) depicts the differences between simulated and experimental $i - V$ curves. We notice that while there are good agreements for most current-voltage points, some simulated points near the MPP show differences compared to the experimental data. This could be due to the difficulty in locating the MPP position with limited information from the $i - V$ curve. Nevertheless, considering the small MSEs, our model demonstrates good capability of predicting key device metrics.

Once calibrated, we can use the models for fast device characterization, predicting key device performances, and estimating recombination losses. Here, as an example, we conduct a loss analysis for the four devices with neural networks B and C. The percentage contributions of various recombination channels to the total recombination losses are shown in Fig. 5(B) and 5(C). The dominant loss channels are the surface recombinations at interfaces IV and V in Cells A1 and A3 with wide band gap perovskites. Radiative recombination is found to be influential in Cells A2 and A4 with narrow band gap perovskites. Hence, in order to enhance device performances, one should focus on suppressing the surface recombination at interface V in Cells A1 to A3, lowering the surface recombination at interface IV in Cells A1, A3, and A4, and having better perovskite film quality for lower bulk recombination losses in Cells A2 and A4.

5.5. Support in-house device characterization and optimization

In this section, we present a method for utilizing our neural networks to assist with in-house characterization of perovskite solar cells. We

fabricated three batches of perovskite solar cells with the structure FTO/self-assembled monolayer (SAM)/perovskite/C₆₀/SnO₂/Ag. Five champion cells are calibrated with neural network A and reported in this work. The experimental and model predicted V_{oc} , FF, and PCE values for these cells (Cells B1 to B5) are shown in Table 6. With the calibrated parameters, the simulated $i - V$ curves from COMSOL agree with the experiments well, as shown in Fig. 6(A) and 7(A). As such, our in-house experiments further validate the trained neural network, together with the numerical model. Lists of parameters used in the model can be found in Table S6 and S7. The accuracy of the recombination parameters extracted during calibration process is shown in Figure S2 to S6. Details on cell fabrications are provided in the Experimental Procedures section.

5.5.1. 1.56 eV band gap perovskite solar cells

For the first batch, Me-4PACz is adopted as the SAM, a perovskite layer with a band gap of 1.56 eV is used, and PMMA/PC₆₁BM and PEABr/MABr are chosen as two different passivators at the perovskite/C₆₀ surface (interface V). The champion devices are numbered as Cell B1 and Cell B2 with PMMA/PC₆₁BM and PEABr/MABr as passivators, respectively. We compare the performances of the two champion cells and their $i - V$ characteristics are shown in Fig. 6(A). By conducting a loss analysis with neural networks B and C, the model supports that PEABr/MABr has a better effect on reducing the recombination loss at interface V than PMMA/PC₆₁BM, as shown in Fig. 6(B) and 6(C). This improvement at the perovskite/C₆₀ interface has resulted in enhancements in V_{oc} from 1.11 to 1.15 V, FF from 0.834 to 0.867, and PCE from 22.9 % to 24.7 %.

In addition, we employ the neural network model to predict the theoretical optimal PCE if both surface recombination losses at interfaces IV and V are minimized. The optimization is conducted in

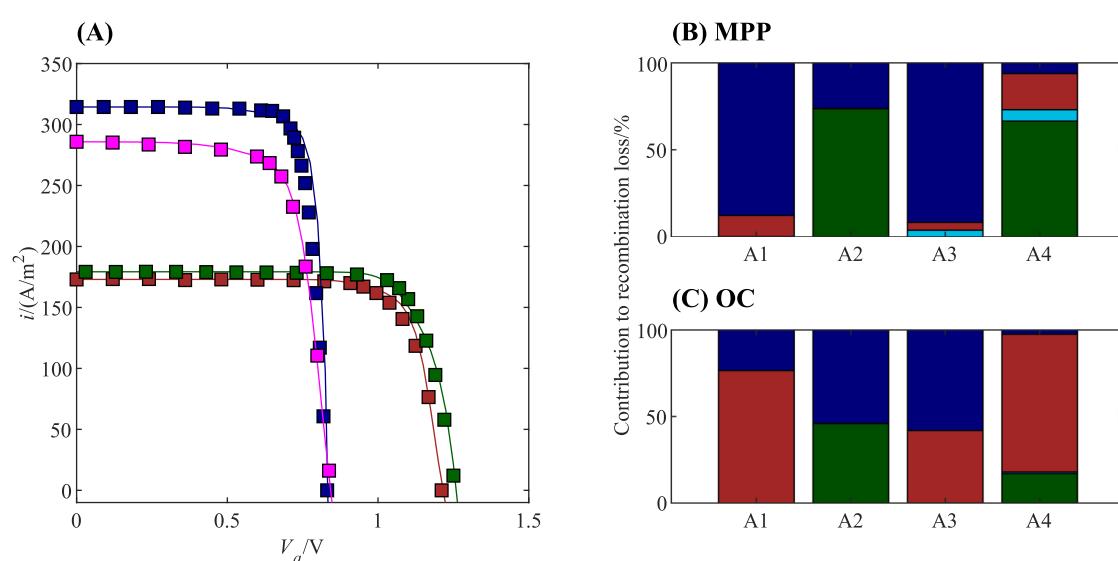


Fig. 5. (A) Simulated (lines) and experimental (symbols) $i - V$ curves for Cells A1 (red), A2 (blue), A3 (green) and A4 (magenta); together with percentage contributions to total recombination losses in cells A1 to A4 at (B) MPP and (C) OC due to radiative recombination (green), SRH recombination (cyan), auger recombination (magenta), surface recombination at interface IV (red) and at interface V (blue). Here, auger recombination losses are non-dominant and hence cannot be observed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6

Cell performances from in house characterizations and calibrated neural network models.

Cell index	Band gap/eV	SAM	Passivator	Experimental			Predicted			MSE
				V _{oc} /V	$\tilde{\delta}$	$\eta/\%$	V _{oc} /V	$\tilde{\delta}$	$\eta/\%$	
B1	1.56	Me-4PACz	PMMA/ PC ₆₁ BM	1.11	0.834	22.9	1.11	0.835	22.9	2.0×10^{-6}
B2	1.56	Me-4PACz	PEABr/ MABr	1.15	0.867	24.7	1.15	0.867	24.7	9.9×10^{-7}
B3	1.63	Me-4PACz	PMMA/ PC ₆₁ BM	1.12	0.812	20.7	1.12	0.813	20.7	1.9×10^{-7}
B4	1.63	Me-4PACz/ MeO-2PACz	PMMA/ PC ₆₁ BM	1.12	0.845	21.3	1.12	0.837	21.1	1.8×10^{-4}
B5	1.63	Me-4PACz/ MeO-2PACz	PEABr/ MABr	1.19	0.834	22.8	1.19	0.835	22.8	2.1×10^{-6}

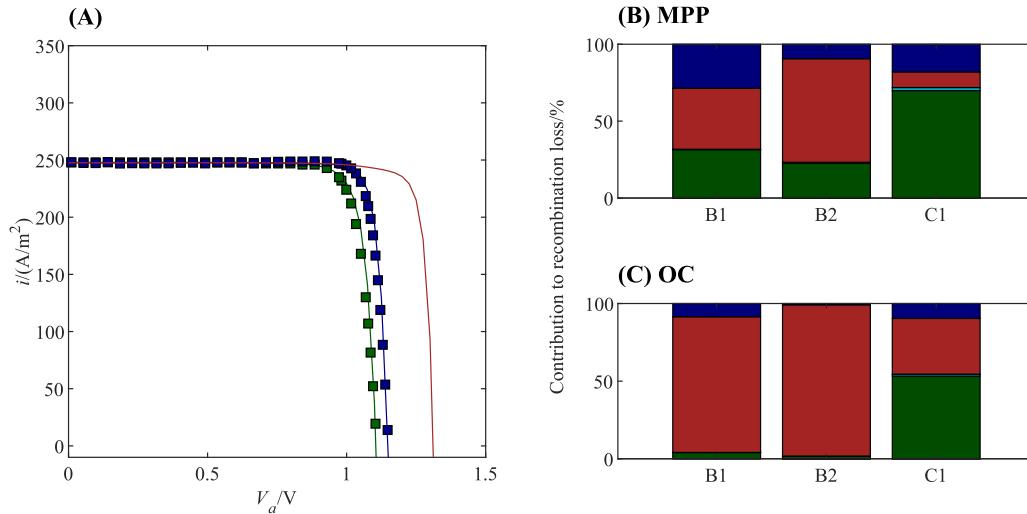


Fig. 6. (A) Simulated (lines) and experimental (symbols) i – V curves for Cells B1 (green), B2 (blue), and C1 (red); together with percentage contributions to total recombination losses at (B) MPP and (C) OC due to radiative recombination (green), SRH recombination (cyan), auger recombination (magenta), surface recombination at interface IV (red) and at interface V (blue). Here, auger recombination losses are non-dominant and hence cannot be observed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

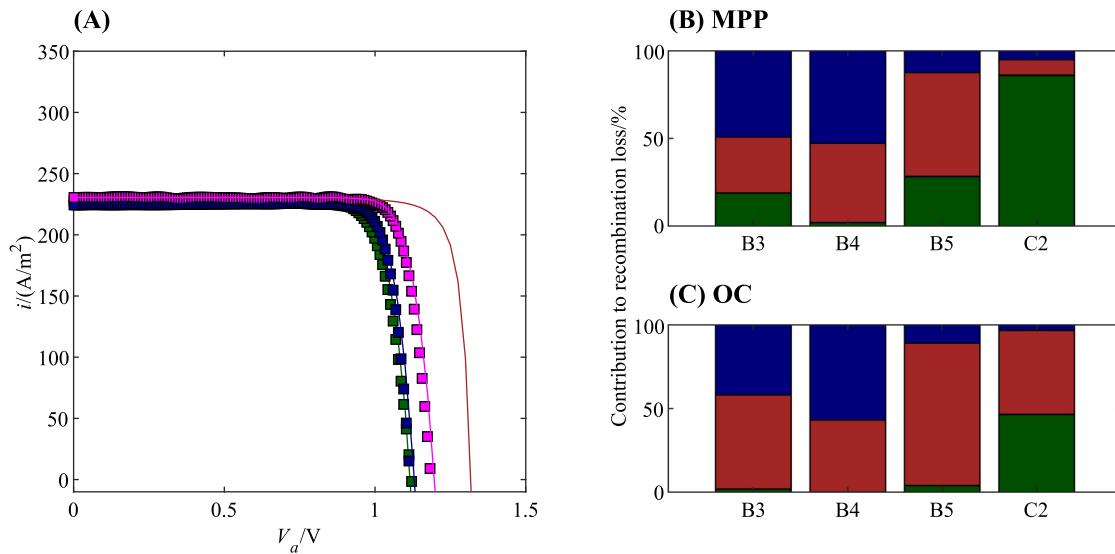


Fig. 7. (A) Simulated (lines) and experimental (symbols) i – V curves for Cells B3 (green), B4 (blue), B3 (magenta), and C2 (red); together with percentage contributions to total recombination losses at (B) MPP and (C) OC due to radiative recombination (green), SRH recombination (cyan), auger recombination (magenta), surface recombination at interface IV (red) and at interface V (blue). Here, auger recombination losses are non-dominant and hence cannot be observed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

MATLAB using particle swarm algorithm with default optimization options and a swarm size of 200. The time consumed for the optimization is about 28 s (wall-clock time). We choose Cell B2 as the baseline. The optimized device, Cell C1, has its i – V curve shown in Fig. 6(A). We

compare the recombination losses between Cell B1 and Cell C1 (see Fig. 6(B) and 6(C)) and notice a significant reduction in the dominant recombination loss at interface IV in Cell C1 after optimization; therefore, focusing on a better perovskite/SAM surface might be a priority on

improving the performances of the fabricated perovskite solar cells. For Cell C1, the theoretical optimal V_{oc} , FF, and PCE are predicted to be 1.31 V, 0.878, and 28.9 %, respectively.

5.5.2. 1.63 eV band gap perovskite solar cells

Next, we fabricated perovskite solar cells with a band gap of 1.63 eV. In this second batch, we report two champion devices, Cell B3 and Cell B4, with Me-4PACz and Me-4PACz/MeO-2PACz as SAM, respectively. Both cells employ PMMA/PC₆₁BM as passivator at interface V. It has been reported that a change in SAM might affect not only the interface, but also the energy alignment at the front contact [34]. Therefore, we also vary the work function of the front-contact layer and the valence band energy level of the SAM during model calibration for these devices (see Table S7). Our calibration results indicate that the mixed Me-4PACz/MeO-2PACz SAM has raised the work function of the front-contact layer from 4.9 eV to 5.0 eV, which is closer to the valence band energy level of the perovskite layer and hence facilitates a smoother hole transport. As a result, comparing to Cell B3 with an FF of 0.812 and a PCE of 20.7 %, Cell B4 achieves a higher FF and PCE of 0.845 and 21.3 %, respectively.

The loss analysis for Cell B4 (see Fig. 7(B) and 7(C)) suggests that the surface recombination at interface V is a significant loss channel. Based on our previous analysis of 1.56 eV band gap perovskite solar cells, we anticipated that switching to a PEABr/MABr passivator at interface V might also improve the performances of the 1.63 eV band gap devices. We hence fabricated another batch of similar perovskite solar cells with a band gap of 1.63 eV, Me-4PACz/MeO-2PACz as SAM, and PEABr/MABr as passivator. Comparing to Cell B4 in the second batch that has a V_{oc} of 1.12 V and a PCE of 21.3 %, this champion cell of the third batch, Cell B5, has achieved a much higher V_{oc} of 1.19 V and a PCE of 22.8 %. After calibrating this champion cell with our model (see Table 6 and Fig. 7(A)), we attribute the observed performance improvement in Cell B5 to a reduced surface recombination rate at interface V—percentage contributions to total recombination losses have decreased from above 50 % to around 10 %.

On the other hand, surface recombination at interface IV emerges as the main loss channel in Cell B5. We optimize this champion cell to estimate the theoretical PCE that could be reached by minimizing surface recombination at both interfaces IV and V. As shown in Fig. 7, the optimized device, Cell C3, is predicted to show a PCE of 25.5 %. The loss analysis shows that the percentage contributions to total losses from surface recombination have dropped from 71 % to 14 % at MPP and 96 % to 53 % at OC. Hence, further improvements to the SAM material or passivation at the perovskite/ETL interface could be helpful for fabricating high-efficiency 1.63 eV band gap perovskite solar cells according to our simulation result.

6. Conclusion

Through replacing the electronic part of the optoelectronic model with an artificial neural network for planar p-i-n structured perovskite solar cells, we develop a surrogate model that enables efficient predictions of device performance and recombination losses. During data preparation, we employ techniques such as even data sampling, logarithmic scaling, data normalization and early stopping for efficient training of neural networks. The hyperparameters during training have been optimized with Bayesian optimization. The optimized neural networks can accurately predict the results with testing MSEs below 4×10^{-4} . By adopting the neural network models, device characterization can be conducted more efficiently in around 24 s. The device architecture can then be improved based on sensitivity analysis, loss analysis, and optimizations.

The model has been applied to support the in-house characterization of perovskite solar cells and has provided insights into dominant recombination losses. By addressing key recombination losses, we have

observed experimental improvements in PCEs of approximately 1.5 % to 2 %. Furthermore, if the interface recombinations are minimized, we predict theoretical PCEs of 28.9 % and 25.5 % for perovskite solar cells with band gaps of 1.56 eV and 1.63 eV, respectively. These optimizations can be performed in under thirty seconds; hence, an iterative process that combines device fabrication, characterization, and optimization through modeling will allow continuous developments of the perovskite solar cells, which places our model a valuable tool towards digital twins applications.

7. Experimental procedures

7.1. Materials

Patterned FTO glass substrates were purchased from Suzhou Sunyo Technology Co., Ltd. [4-(3,6-Dimethyl-9H-carbazol-9-yl)butyl]phosphonic acid (Me-4PACz, 99 %) and [2-(3,6-Dimethoxy-9H-carbazol-9-yl)ethyl]phosphonic Acid (MeO-2PACz, >98 %) was purchased from Tokyo Chemical Industry Co., Ltd. Reagent ethanol (denatured, 99.5 %, anhydrous), 2-Propanol (IPA, 99.5 %, anhydrous), Chlorobenzene (CB, 99.8 %, anhydrous), N,N-Dimethylformamide (DMF, 99.8 %, anhydrous) and Dimethyl sulfoxide (DMSO, >99.9 %, anhydrous) were purchased from Sigma Aldrich. Cesium chloride (CsCl, 99.9 %), Cesium iodide (CsI, 99.999 %), Formamidinium iodide (FAI, 99.5 %), Methylammonium iodide (MAI, 99.5 %), Methylamine hydrochloride (MACl, 99.5 %), Methylammonium bromide (MABr, 99.5 %), Phenethylammonium bromide (PEABr, 99.5 %), Lead(II) bromide (PbBr₂, 99.99 %), and Lead(II) iodide (PbI₂, 99.999 %) were bought from Xi'an Yuri Solar Co., Ltd. Fullerene (C₆₀) and [6,6]-phenyl-C₆₁-butyric acid methyl ester (PC₆₁BM) were bought from Liaoning Youxuan. Polymethyl Methacrylate (PMMA) was purchased from J&K Scientific.

8. Methods

Patterned FTO glasses were washed by deionized (DI) water, acetone, ethanol sequentially, each for 15 min in an ultrasonic bath. The glass substrates were then blown to dry by a nitrogen gun and treated with ultraviolet ozone for 20 min before use. Me-4PACz with a concentration of 0.6 mg/mL in ethanol and a mixture of Me-4PACz and MeO-2PACz (mass ratio 1:1) with a concentration of 0.6 mg/mL in ethanol were used for the hole-transporting layer. For both depositions of the hole-transporting layer, the solution was spin-coated onto the pretreated FTO surface at 4000 rpm for 20 s, followed by annealing at 100 °C for 10 min.

For the 1.56 eV perovskite layer, 11.1 mg of MAI, 11.8 mg of CsCl, 18.9 mg of MACl, 228.7 mg of FAI, and 710 mg of PbI₂ were dissolved in 1 mL of DMF and DMSO solvent mixture (v/v = 4/1). The precursor was stirred for 6 h and filtered using 0.22 μm PTFE syringe filters before use. The precursor was then spin-coated onto the HTL at 4500 rpm for 40 s with an acceleration of 500 rpm/s. 200 μL CB was drop casted onto the surface as the antisolvent during the last 20 s of spinning. The substrate was then transferred to a hotplate and annealed at 100 °C for 30 min and 150 °C for 20 min.

For the 1.63 eV perovskite layer, 16.8 mg of CsI, 17.2 mg of FAI, 22.4 mg MABr, 73.4 mg of PbBr₂, and 507.1 mg of PbI₂ were mixed and dissolved in 1 mL of a DMF and DMSO solvent mixture (v/v = 4/1). The precursor was stirred for 6 h and filtered using 0.22 um PTFE syringe filters before use. The precursor was then spin-coated onto the hole-transporting layer at 3000 rpm for 40 s with an acceleration of 500 rpm/s. 250 μL CB was drop casted onto the surface as the antisolvent during the last 20 s of spinning. The substrate was then transferred to a hotplate and annealed at 100 °C for 20 min.

Two passivators for perovskite surface were applied in the experiments. The first passivator was a mixture of PEABr and MABr prepared by dissolving 1 mg of PEABr and 1 mg of MABr into 1 mL of IPA. The second passivator was a mixture of PMMA and PC₆₁BM prepared by

dissolving 0.5 mg of PMMA and 1.5 mg of PC₆₁BM into 1 mL of CB. For both passivation recipes, the precursor solution was spin-coated onto the perovskite at 4000 rpm for 20 s, followed by annealing at 100 °C for 5 min. After this, a 25 nm layer of C₆₀ was thermally evaporated at a rate of 0.2 Å/s as the electron-transporting layer, followed by the deposition of a 10 nm SnO₂ buffer layer using atomic layer deposition at 80 °C. Finally, a 100 nm layer of Ag was thermally evaporated as the back electrode.

CRediT authorship contribution statement

Xinhai Zhao: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Chaopeng Huang:** Writing – original draft, Investigation, Conceptualization. **Erik Bergersson:** Writing – review & editing, Software, Conceptualization. **Nikita Suprun:** Writing – original draft, Investigation. **Hu Quee Tan:** Writing – review & editing, Methodology. **Yurou Zhang:** Writing – review & editing, Visualization. **Yuxia Jiang:** Writing – review & editing, Validation. **Chunhui Shou:** Project administration, Funding acquisition. **Jingsong Sun:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jun Peng:** Writing – review & editing, Supervision, Resources. **Hansong Xue:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

X.Z., C.H., J.S., Y.Z., Y.J. and J.P. acknowledge the support of Key R&D Program of Zhejiang (2024SSYS0061).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.apenergy.2025.125922>.

Data availability

The data and code supporting this article have been included in the main paper, as part of the SI or can be found at <https://github.com/xuehansong-sutd/Accelerating-device-characterization>.

References

- [1] Feng S, Wang J. Prediction of Organic–Inorganic Hybrid Perovskite Band Gap by Multiple Machine Learning Algorithms. *Molecules* 2024;29:499.
- [2] Obada DO, Okafor E, Abolade SA, Ukpong AM, Dodoo-Arhin D, Akande A. Explainable machine learning for predicting the band gaps of ABX₃ perovskites. *Mater Sci Semicond Process* 2023;161:107427.
- [3] Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett* 2018;120: 145301.
- [4] Jarin S, Yuan Y, Zhang M, Hu M, Rana M, Wang S, et al. Predicting the crystal structure and lattice parameters of the perovskite materials via different machine learning models based on basic atom properties. *Crystals* 2022;12:1570.
- [5] Zhang Y, Xu X. Machine learning lattice constants for cubic perovskite compounds. *ChemistrySelect* 2020;5:9999–10009.
- [6] Talapatra A, Uberuaga BP, Stanek CR, Pilania G. A machine learning approach for the prediction of formability and thermodynamic stability of single and double perovskite oxides. *Chem Mater* 2021;33:845–58.
- [7] Zhang S, Lu T, Xu P, Tao Q, Li M, Lu W. Predicting the Formability of Hybrid Organic–Inorganic Perovskites via an Interpretable Machine Learning Strategy. *J Phys Chem Lett* 2021;12:7423–30.
- [8] Ichwani R, Price S, Oyewole OK, Neamtu R, Soboyejo WO. Using machine learning for prediction of spray coated perovskite solar cells efficiency: from experimental to theoretical models. *Materials & Design* 2023;233:112161.
- [9] Hu Y, Hu X, Zhang L, Zheng T, You J, Jia B, et al. Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells. *Adv Energy Mater* 2022;12.
- [10] Oboh IO, Offor UH, Okon ND. Artificial neural network modeling for potential performance enhancement of a planar perovskite solar cell with a novel TiO₂ electron transport bilayer using nonlinear programming. *Energy Rep* 2022;8: 973–88.
- [11] Sun S, Xie Z, Qin G, Xiao L. Light trapping nano structures with over 30% enhancement in perovskite solar cells. *Org Electron* 2019;75:105385.
- [12] Wu J, Li Y, Li Y, Xie W, Shi J, Li D, et al. Using hysteresis to predict the charge recombination properties of perovskite solar cells. *J Mater Chem A* 2021;9: 6382–92.
- [13] Tan HQ, Zhao X, Ambardekar A, Bergersson E, Xue H. Exploring the optimal design space of transparent perovskite solar cells for four-terminal tandem applications through Pareto front optimization. *APL Machine Learning* 2024;2.
- [14] Xue H, Bergersson E, Stangl R. Correlating variability of modeling parameters with photovoltaic performance: Monte Carlo simulation of a meso-structured perovskite solar cell. *Appl Energy* 2019;237:131–44.
- [15] Xue H, Stangl R, Bergersson E. Elucidating the functional form of the recombination losses in a planar perovskite solar cell: a scaling analysis. *J Appl Phys* 2020;128: 123102.
- [16] Guemard CA, Myers D, Emery K. Proposed reference irradiance spectra for solar energy systems testing. *Sol Energy* 2002;73:443–67.
- [17] Courtier NE, Cave JM, Foster JM, Walker AB, Richardson G. How transport layer properties affect perovskite solar cell performance: insights from a coupled charge transport/ion migration model. *Energy & Environ Sci* 2019;12:396–409.
- [18] Campa A, Krc J, Topic M. Two approaches for incoherent propagation of light in rigorous numerical simulations. *Prog Electromagn Res* 2013;137:187–202.
- [19] Zai H, Ma Y, Chen Q, Zhou H. Ion migration in halide perovskite solar cells: mechanism, characterization, impact and suppression. *J Energy Chem* 2021;63: 528–49.
- [20] Zhao X, Tan HQ, Bergersson E, Xue H. Elucidating the underlying physics in a two-terminal all-perovskite tandem solar cell: a guideline towards 30% power conversion efficiency. *Sol Energy* 2022;231:716–31.
- [21] Tan HQ, Zhao X, Bergersson E, Lin F, Xue H. Optoelectronic modeling and sensitivity analysis of a four-terminal all-perovskite tandem solar cell – identifying pathways to improve efficiency. *Sol Energy* 2021;216:589–600.
- [22] Xue H, Bergersson E, Stangl R. A spatially smoothed device model for meso-structured perovskite solar cells. *J Appl Phys* 2018;124:193103.
- [23] Xue H, Fu K, Wong LH, Bergersson E, Stangl R. Modelling and loss analysis of meso-structured perovskite solar cells. *J Appl Phys* 2017;122:083105.
- [24] Zhao X, Tan HQ, Bergersson E, Chen W, Hou Y, Xue H. Modeling and sensitivity analysis of a two-terminal perovskite on organic tandem solar cell. *Cell Rep Phys Sci* 2022;3:101038.
- [25] Haykin S. Neural networks. Prentice Hall; 2007. p. 842.
- [26] Burden F, Winkler D. Bayesian regularization of neural networks. In: Methods in molecular biology™. Humana Press; 2008. p. 23–42.
- [27] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; 15:1929–58.
- [28] Ren Z, Oviedo F, Thway M, Tian SIP, Wang Y, Xue H, et al. Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. *Npj Computational Mater* 2020;6.
- [29] Lin R, Xiao K, Qin Z, Han Q, Zhang C, Wei M, et al. Monolithic all-perovskite tandem solar cells with 24.8% efficiency exploiting comproportionation to suppress Sn (II) oxidation in precursor ink. *Nat Energy* 2019;4:864–73.
- [30] Chen W, Zhu Y, Xiu J, Chen G, Liang H, Liu S, et al. Monolithic perovskite/organic tandem solar cells with 23.6% efficiency enabled by reduced voltage losses and optimized interconnecting layer. *Nat Energy* 2022;7:229–37.
- [31] Zhao D, Wang C, Song Z, Yu Y, Chen C, Zhao X, et al. Four-terminal all-perovskite tandem solar cells achieving power conversion efficiencies exceeding 23%. *ACS Energy Let* 2018;3:305–6.
- [32] Caprioglio P, Wolff CM, Sandberg OJ, Armin A, Rech B, Albrecht S, et al. On the origin of the ideality factor in perovskite solar cells. *Adv Energy Mater* 2020;10: 2000502.
- [33] J. Tong, Z. Song, D. H. Kim, X. Chen, C. Chen, A. F. Palmstrom, P. F. Ndione, M. O. Reese, S. P. Dunfield, O. G. Reid and others, “Carrier lifetimes of > 1 μs in Sn-Pb perovskites enable efficient all-perovskite tandem solar cells,” *Science*, vol. 364, p. 475–479, 2019.
- [34] Li W, Martínez-Ferrero E, Palomares E. Self-assembled molecules as selective contacts for efficient and stable perovskite solar cells. *Mater Chem Front* 2024;8: 681–99.