Physics-Informed Convolutional Surrogates for Coupled Drift-Diffusion Equations: Reconstructing Full J-V Curves of Perovskite Solar Cells

Mehmet Ozdincer[*]
University of Toronto, Toronto, Canada

Prof. Erik Birgersson
National University of Singapore, Singapore
Solar Energy Research Institute of Singapore (SERIS)

[*]Corresponding author: memo@cs.toronto.edu

Abstract

We present a physics-constrained deep learning surrogate that reconstructs full current-voltage (J-V) curves of perovskite solar cells from 31 coupled drift-diffusion parameters spanning >20 orders of magnitude. On ∼15k held-out devices the model achieves median $R^2 = 0.9975$, median MAE of 2.96 A/m^2, and median V_{oc} error of 0.39 mV, which is an order of magnitude below experimental measurement uncertainty, while reducing inference from ∼4800 s per COMSOL simulation to milliseconds ($>10^4\times$ speedup), enabling million-device screening in under one hour on a single GPU. Physics knowledge informs every stage of the pipeline: 71 analytically derived drift-diffusion features are distilled to 5 based on the independent physical mechanisms governing J-V shape (photocurrent generation, band gap energetics, built-in field alignment, interface band alignment, recombination losses), while soft monotonicity, convexity, and curvature penalties encode qualitative PDE properties on the output. The architecture itself is physics-motivated: bidirectional dilated convolutions provide full receptive-field coverage of the compact 8-point J-V representation, matching the thermodynamic equilibrium nature of J-V responses and their monotonically decaying spatial correlation structure, which renders self-attention redundant and consistently harmful on these short sequences.

Code: github.com/memo-ozdincer/perovskite-jv-surrogate
Model: huggingface.co/memo-ozdincer/perovskite-jv-surrogate

1. Introduction

Perovskite solar cells now exceed 26% PCE in single-junction configurations (Green et al., 2014; Zhao, 2023), but optimizing them requires navigating a 31-dimensional parameter space of bandgaps, mobilities, thicknesses, doping densities, recombination coefficients, and contact work functions, whose interactions are governed by coupled nonlinear drift-diffusion physics.

The standard computational approach is coupled Poisson–drift-diffusion simulation via COMSOL finite-element methods (FEM), which faithfully resolves the transport physics but costs ∼4800 s per device (∼ 160 years for a typical 10^6 device campaign), making its use for high-throughput and manufacturing-side research at labs like SERIS infeasible.

Zhao et al. (2025) achieved $>10^3\times$ speedup for *scalar* metrics (V_{oc}, FF, PCE) via neural network surrogates. However, these metrics do not capture device characteristics that may be derived from full curves including: S-shaped kinks from interface barriers, fill-factor degradation modes, and the drift-to-recombination transition at the maximum power point. Two devices with identical PCE can have qualitatively different J-V curves and correspondingly different operational characteristics.

We address this gap with the following contributions:

1. **First full J-V curve surrogate from coupled drift-diffusion parameters** (§2): median $R^2 = 0.9975$, median V_{oc} error of 0.39 mV (below experimental measurement uncertainty), and $>10^4\times$ speedup over COMSOL FEM, which reduces a 10^6-device screen-

ing campaign from ∼160 years to under one hour on a single GPU.

2. **Physics-driven feature engineering** (§6.4–6.5): we derive 71 analytical drift-diffusion features and select 5 based on physical reasoning about the independent mechanisms governing J-V shape, photocurrent generation, band gap energetics, built-in field alignment, interface band alignment, and recombination losses, with multicollinearity analysis (175 pairs at $|r| > 0.85$) supporting the selection.

3. **Domain-matched architecture design** (§6.6–2.2): bidirectional dilated convolutions provide full receptive-field coverage of the 8-point J-V sequence, matching the physical structure of thermodynamic equilibrium responses where spatial correlations decay monotonically with voltage separation, which rendered self-attention redundant and consistently harmful.

4. **Physics-native representation pipeline** (§6.3–6.7): a co-designed pipeline where each component exploits specific physical structure of J-V curves—monotone PCHIP compression ($45 \rightarrow 8$ points) simultaneously reduces output dimensionality and encodes a shape prior, learnable Gaussian RBF voltage embeddings give the convolutional backbone a continuous voltage coordinate matching the non-uniform grid physics, J_{sc}-normalization decouples current magnitude from curve shape, and soft monotonicity/convexity/curvature penalties encode qualitative PDE properties without requiring PDE evaluation during training.

2. Results

Figure 1 displays J-V reconstructions for randomly sampled test devices spanning the full performance range. The model's 8-point PCHIP representation, interpolated back to a dense voltage grid via Fritsch–Carlson monotone cubic Hermite interpolation (Fritsch & Carlson, 1980), faithfully captures both the flat photocurrent plateau at low voltages and the sharp sigmoidal transition at the knee/MPP region. The vast majority of devices achieve $R^2 > 0.99$; the rare low-R^2 cases correspond to pathological device configurations discussed in §2.3.

2.1. Primary Performance

Table 1 reports performance of the primary model (bidirectional dilated conv, batch size 512, with physics features) on the held-out test set, averaged over 3 random seeds.

Table 1. Test-set performance (∼15k devices), averaged over 3 seeds (± seed-to-seed std). The mean–median gap in R^2 reflects a long left tail from pathological devices (§2.3), not systematic bias.

Metric	Mean	Median
R^2 (ΔV-wt.)	0.9858 ± 0.0009	$\mathbf{0.9975 \pm 0.0004}$
MAE (A/m^2)	4.75 ± 0.26	2.96 ± 0.34
RMSE (A/m^2)	6.66 ± 0.26	4.27 ± 0.34
$\lvert V_{oc}$ err.$\rvert$ (V)	0.0056 ± 0.0058	$\mathbf{0.00039 \pm 0.00016}$
$\lvert I_{sc}$ err.$\rvert$ (A/m^2)	0.59 ± 0.53	0.36 ± 0.62

Physical significance. The median MAE of 2.96 A/m^2 corresponds to <0.5% of a typical $J_{sc} \approx 200$ A/m^2, meaning the model faithfully resolves both the flat-band photocurrent magnitude and the sharp transition at MPP where most power extraction occurs. The I_{sc} median error of 0.36 A/m^2 (with seed-to-seed variability of ±0.62) confirms that the flat-band region ($V \ll V_{mpp}$) is trivially learned, consistent with the physics, since $J(V \approx 0) \approx J_{sc}$ is determined primarily by photogeneration and collection efficiency, which are well-captured by the input features.

The median V_{oc} error of 0.39 mV is an order of magnitude below typical experimental measurement uncertainty (±5 mV for perovskite devices with hysteresis), meaning the model's voltage-axis localization of the open-circuit point is more precise than what current fabrication and measurement technology can resolve. This sub-millivolt precision in V_{oc} reconstruction is important because V_{oc} determines the upper voltage bound of the J-V curve, and errors in V_{oc} propagate into the reconstructed current values throughout the knee region where dJ/dV is large.

The mean–median gap. The gap between mean $R^2 = 0.986$ and median $R^2 = 0.998$ indicates that fewer than 1% of devices account for the majority of aggregate error. This asymmetry is by design: the physics constraints *correctly refuse* to fit S-shaped or concave-up curves that violate the monotone-sigmoid assumption (§2.3). A model that achieved higher mean R^2 by fitting these pathological devices would necessarily sacrifice physical plausibility on the remaining 99%.

Computational performance. Total wall time: ∼10 min (preprocessing 24 s, training ∼9.5 min, inference ∼10 s on ∼15k test samples). Training throughput: ∼12.7k samples/s (batch size 512); inference throughput: ∼1.4k samples/s. This represents $>10^4\times$ speedup over COMSOL FEM per device, enabling a 10^6-device Latin Hypercube screening sweep in <1 hour.

2.2. Architecture Design Rationale

Each architectural decision is motivated by the physics of J-V curves and validated empirically (full 30-run ablation in Appendix C).

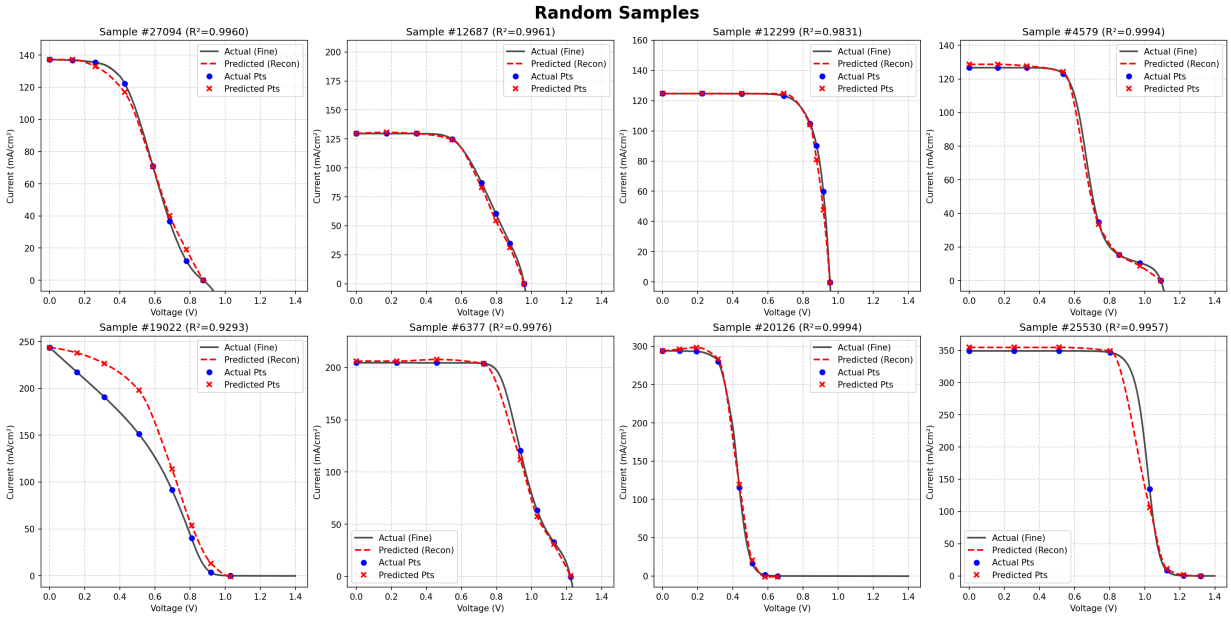


Figure 1. **Random test-set reconstructions** ($R^2 = 0.58$–0.999). Blue circles: 8 PCHIP query points predicted by the dilated conv model; dashed: PCHIP reconstruction; solid grey: COMSOL finite-element reference. The model captures flat-band plateaus, sharp knees, and the approach to V_{oc} across devices with widely varying material parameters. The rare low-R^2 cases (bottom row) correspond to S-shaped or barrier-dominated devices where physics constraints *correctly refuse* unphysical fits (§2.3).

Spatial mixing: inter-voltage coupling is physically required. J-V curves are smooth, monotone responses governed by coupled transport equations, which state that the current at any voltage depends on the carrier distribution across the *entire* device. A pointwise (1×1) baseline that predicts each voltage independently achieves only $R^2_{med} = 0.989$, compared to 0.997 for the dilated conv, confirming that the architecture must exchange information across voltage points to capture the interdependence encoded in the drift-diffusion equations.

Bidirectional context: equilibrium, not causality. J-V curves describe *thermodynamic equilibrium* responses, not temporal sequences: at any applied voltage, the current depends on the full band diagram ($V_{bi} - V$ across all layers), recombination at both interfaces, and extraction at both contacts. Causal (unidirectional) masking discards this bidirectional physical dependence, and indeed the causal TCN shows $3.9\times$ higher V_{oc} error variance than the bidirectional conv (1.51 ± 1.68 mV vs. 0.39 ± 0.16 mV), with instability concentrated near V_{oc} where the causal architecture lacks access to high-voltage context.

Dilated convolutions: full receptive field without attention. Block 3's dilation d=2 with kernel k=5 yields a single-layer receptive field of 9 positions, already exceeding the 8-point sequence length. This architectural property has a direct physical consequence: because every output position already has bidirectional context over all input positions, the $\mathcal{O}(n^2)$ pairwise interactions of self-attention (Vaswani et al., 2017) are redundant. Empirically, adding 4-head attention degrades median R^2 from $\sim$0.997 to $\sim$0.991 across both Conv and TCN backbones (Appendix C, Figure 11). The underlying reason is structural: J-V spatial correlations decay monotonically with voltage separation, matching convolution's local inductive bias, while attention treats all voltage pairs uniformly and amplifies noise from weakly coupled points.

Batch size 512: gradient noise meets the loss landscape. Increasing batch size from 128 to 512 yields the largest single-factor improvement (R^2_{med}: $0.9964 \rightarrow 0.9975$; MAE: $3.74 \rightarrow 2.96$ A/m^2) while halving training time to $\sim$10 min. The physics-constrained loss (Eq. 25) combines five terms with different gradient scales; larger batches reduce per-term gradient variance, enabling more stable optimization of the multi-objective landscape.

2.3. Failure Mode Analysis

The left tail of the R^2 distribution ($R^2_{mean} = 0.986$ vs. $R^2_{med} = 0.998$, affecting $<$1% of devices) arises from three physically distinct failure modes, each traceable to a specific violation of the assumptions encoded in the model's architecture and loss:

1. **S-shaped / concave-up curves**: Large interface band offsets ($|\Delta E_v^{HP}|$ or $|\Delta E_c^{PE}| \gg k_B T$) produce barrier-

dominated transport where the J-V curve exhibits a local current *increase* with voltage (the "S-kink"), violating the monotone-sigmoid assumption inherent in the PCHIP representation. Physically, S-kinks arise when extraction barriers at one interface force carriers to accumulate, creating a voltage-dependent barrier height that produces non-monotone current response. The monotonicity constraint *correctly refuses* to fit these curves: the model produces a smooth sigmoid approximation rather than an unphysical reconstruction. This is a deliberate trade-off: capturing S-shaped behavior would require abandoning the monotone PCHIP representation and significantly increasing output dimensionality, at the cost of less constrained predictions for the 99%+ well-behaved devices. Notably, this failure mode correlates strongly with $|\Delta E_v^{HP}| \gg k_B T$ and large $|E_g^{offset}|$, both of which are among the selected physics features, meaning the model has the information to *detect* these devices even though it cannot faithfully *reconstruct* them.

2. **Mislocated knee**: When V_{oc} falls far from the training-set center (typically $V_{oc} < 0.5$ V or $V_{oc} > 1.2$ V), the sigmoid transition occurs outside the dense query-point region, and the 8-point budget lacks resolution where it is most needed. This is a consequence of the fixed query-point grid: the points are concentrated near the *typical* MPP voltage, but extreme-V_{oc} devices have their steepest curvature elsewhere. Adaptive query-point placement conditioned on predicted V_{oc} would address this failure mode while preserving the compact 8-point representation.

3. **Extreme series resistance**: $R_s^{total} > 10^3\ \Omega{\cdot}\text{cm}^2$ produces nearly linear J-V curves ($J(V) \approx J_{sc} - V/R_s$) that are qualitatively unlike the sigmoid majority. The convexity and curvature penalties are designed for sigmoidal shapes and actively resist linear reconstructions. These devices have FF $\ll 0.5$ and represent impractical configurations that would never be fabricated. Consistent with this, R_s^{total} achieved $|r| < 0.001$ with target variables during feature selection (§6.5), confirming that resistive effects are negligible for the vast majority of the parameter space.

In all three cases, the failure mode is *physically interpretable* and corresponds to devices outside the intended operating regime. The physics features themselves enable a lightweight anomaly detector: a simple threshold on $|E_g^{offset}|$, $|\Delta E_v^{HP}|$, and R_s^{total} can flag pathological devices at negligible computational cost before running inference, providing a built-in reliability indicator.

Figures 2 and 3 contrast the best and worst reconstructions, making the failure modes visually concrete.

3. Background: Drift-Diffusion Model

The ground-truth J-V curves are generated by a coupled Poisson–drift-diffusion simulator (Zhao, 2023; Zhao et al., 2025; Xue et al., 2019; 2020) modeling a planar p-i-n perovskite stack: Glass / Front-contact / HTL / Perovskite absorber / ETL / Back-contact. Optical absorption follows a vectorized transfer-matrix formulation (Appendix A) yielding the spatially averaged photogeneration rate $G_{avg} = l_P^{-1} \int_0^{l_P} G(x)\, dx$, which enters as one of the 31 inputs.

The electronic transport is governed by three coupled PDEs:

$$\nabla \cdot (\varepsilon \nabla \psi) = -q(p - n + N_D^+ - N_A^-) \tag{1}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{J}_n - R + G \tag{2}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \mathbf{J}_p - R + G \tag{3}$$

where the current densities follow drift-diffusion:

$$\mathbf{J}_n = qn\mu_n \mathbf{E} + qD_n \nabla n, \qquad \mathbf{J}_p = qp\mu_p \mathbf{E} - qD_p \nabla p \tag{4}$$

with diffusion coefficients related to mobilities through the Einstein relation $D = \mu k_B T/q$.

Recombination combines radiative, Shockley-Read-Hall (SRH), Auger, and interface terms:

$$R_{rad} = B_{rad}(np - n_i^2) \tag{5}$$

$$R_{SRH} = \frac{np - n_i^2}{\tau_h(n+n_1) + \tau_e(p+p_1)} \tag{6}$$

$$R_{Aug} = (A_e n + A_h p)(np - n_i^2) \tag{7}$$

with intrinsic carrier concentration $n_i = \sqrt{N_c N_v} \exp(-E_g/2k_B T)$, built-in voltage $V_{bi} = W_{anode} - W_{cathode}$, and interface recombination velocity $j_s = v_s \Delta c$.

The system is solved in COMSOL at each applied voltage V to generate the $J(V)$ curve. The 31 input parameters (Table 2) span >20 orders of magnitude, from carrier mobilities ($\sim 10^{-4}$–10^2 cm^2/Vs) to doping densities ($\sim 10^{17}$–10^{21} cm^{-3}) to recombination coefficients ($\sim 10^{-12}$–10^{-6} cm^3/s), necessitating careful preprocessing (§6.3).

4. Two-Stage Surrogate Pipeline

The framework operates as a modular two-stage pipeline (Figure 4):

Stage 1: prediction (Zhao et al., 2025). Three-layer ANNs with Bayesian regularization, trained on 100k LHS samples, predict scalar device metrics from the 31 raw parameters:

$$\mathbf{x} \in \mathbb{R}^{31} \xrightarrow[\text{Zhao et al.}]{\text{3-layer ANN}} \{V_{oc}^{pred}, V_{mpp}^{pred}, J_{sc}^{pred}\} \tag{8}$$

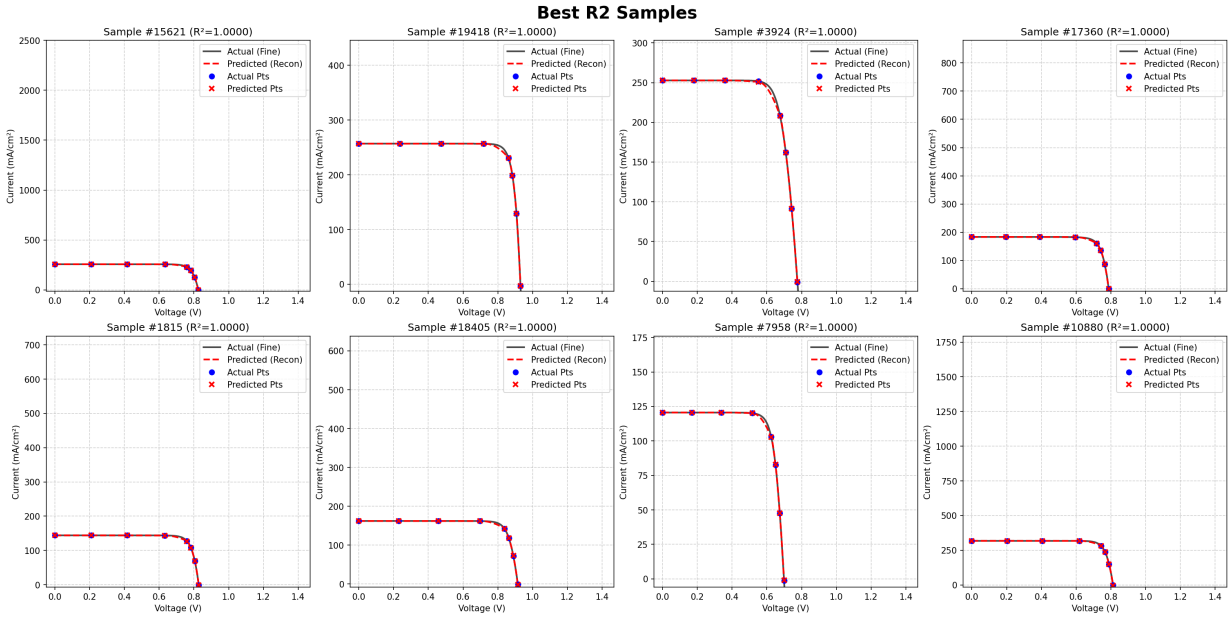


Figure 2. **Best R^2 reconstructions** (≈ 1.000): near-perfect reconstruction for well-behaved sigmoid curves with clear knee regions. These devices have favorable energetic alignment ($E_g^{offset} \approx 0$), moderate barrier heights, and high J_{max}, placing them squarely within the model's physics-constrained operating regime.

Table 2. The 31 COMSOL parameters grouped by physical origin. Parameters span >20 orders of magnitude, with six per-layer quantities (band edges, mobilities), three layer thicknesses, and global recombination/generation parameters.

Group	n	**Parameters**
Band edges	6	$\chi_{e,h}$ for ETL, perovskite, HTL (eV)
Mobilities	6	$\mu_{n,p}$ per layer (cm^2/Vs)
Thicknesses	3	l_H, l_P, l_E (nm)
SRH lifetimes	2	τ_e, τ_h (s)
Density of states	4	$N_{c,v}$ for perovskite and ETL (cm^{-3})
Recombination	4	$v_{IV,V}$ (cm/s), B_{rad} (cm^3/s), C_{Aug} (cm^6/s)
Permittivities	3	ε per layer (relative)
Contacts	2	$W_{anode}, W_{cathode}$ (eV)
Generation	1	G_{avg} (cm^{-3}s^{-1})

with MSE $< 3.1 \times 10^{-4}$ for V_{oc}, validated against 9 fabricated devices spanning $E_g \in \{1.56, 1.63\}$ eV (calibration MSE 8.2×10^{-8}–1.8×10^{-4}) (Zhao et al., 2025).

Stage 2: Curve reconstruction (this work). Precomputed scalar values[1] condition the curve model alongside 31 scaled parameters and m selected physics features:

$$\mathbf{x}_{in} = [\underbrace{\mathbf{x}^{sc}}_{31}\,;\,\underbrace{V_{oc}^{pred}, V_{mpp}^{pred}}_{\text{Stage 1}}\,;\,\underbrace{\boldsymbol{\phi}(\mathbf{x})}_{m}] \xrightarrow{f_\theta} \hat{\mathbf{J}} \in \mathbb{R}^8 \quad (9)$$

where $m = 5$ after train-only feature selection (§6.5).

This design prevents inter-stage gradient leakage, enables swapping ground-truth for predicted scalars during ablation, and allows independent training of each stage.

5. Related Work

ML surrogates for perovskites. Prior machine learning surrogates for perovskite solar cells predict scalar properties: band gap from composition (Feng et al., 2024; Djeradi et al., 2024), formability classification (Talapatra et al., 2021), lattice constants (Zhang & Xu, 2020), and device-level metrics

[1]To isolate the intrinsic accuracy of the curve reconstruction stage, the primary results in this paper use ground-truth V_{oc} and V_{mpp} from COMSOL rather than Stage 1 predictions. Zhao et al. (2025) demonstrated that V_{oc} can be predicted from the 31 raw parameters with $R^2 = 0.996$, and we verified that V_{mpp} can be derived from Zhao's predicted scalars with $R^2 = 0.99$ using gradient-boosted tree regressors. The ablation removing scalar conditioning entirely (§7) shows only $\Delta R^2_{med} < 0.001$, indicating that end-to-end deployment with predicted scalars would yield negligibly lower curve reconstruction performance.

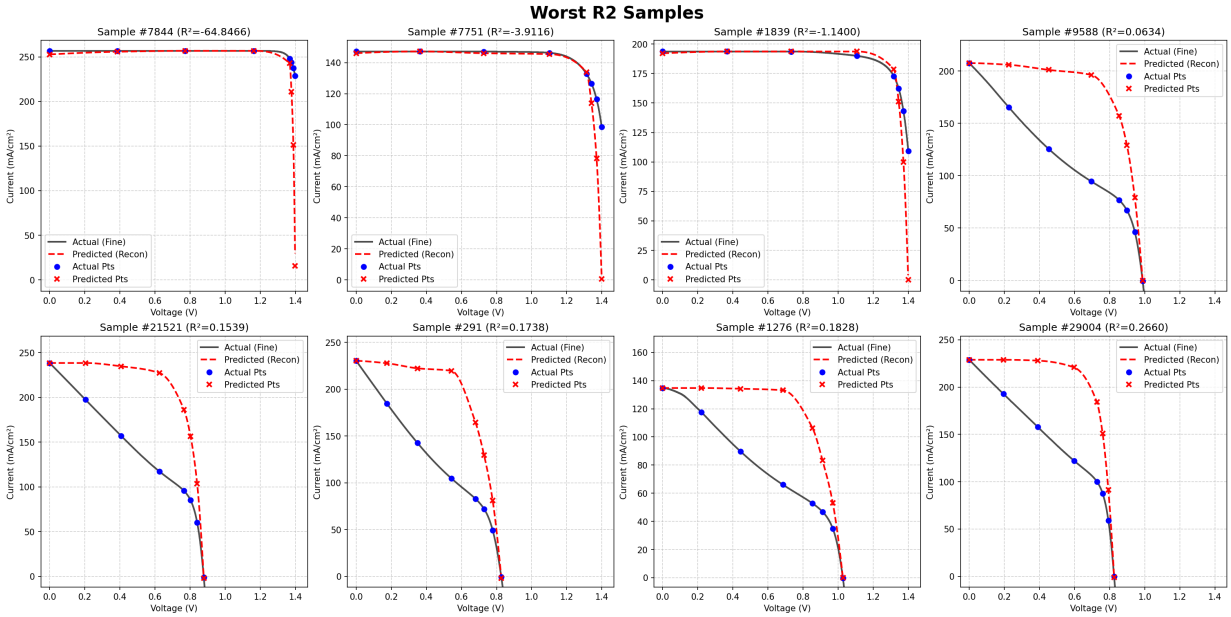


Figure 3. **The very few outliers in R^2 reconstructions** (< 0): S-shaped kinks (barrier-dominated transport), concave-up curves, and extreme-resistance configurations. In each case, the physics constraints *correctly prevent* unphysical fits, producing smooth sigmoid approximations instead. These devices have large $|E_g^{offset}|$, extreme $|\chi_P^e|$ values, or $|\Delta E_v^{HP}| \gg k_B T$, violating the monotone-sigmoid assumption inherent in the PCHIP representation.

(PCE, FF) (Hu et al., 2022; Oboh et al., 2023; Ren et al., 2019; Suprun, 2024). None predict full J-V curves from material parameters.

J-V curve prediction. For non-perovskite photovoltaics, Li et al. (2020) and Porowski et al. (2024) predicted I-V curves from environmental conditions using CNNs and MLPs, respectively, but for fixed device geometries rather than variable material parameters. Two concurrent works address J-V reconstruction for perovskite devices (Table 3). Toprak (2025) mapped irradiance and voltage (G, V) to current density J for a *single fixed device geometry* (2 input dimensions, $R^2 = 0.994$, 8k training samples). Zbinden et al. (2026) used conditional variational autoencoders for forward and inverse J-V modeling from 6 fabrication parameters ($R^2 = 0.996$, 50k training samples, validated on 4 fabricated devices). Our work generalizes across 31 coupled drift-diffusion dimensions with $\sim$150k training samples and explicit physics constraints, and contributes the architectural finding that bidirectional convolutions outperform attention-based and point-wise architectures on short physics-governed sequences.

Physics-informed neural networks (PINNs). Classical PINNs (Raissi et al., 2019) embed PDE residuals as loss terms, requiring PDE evaluation during training. Recent work has applied this paradigm directly to semiconductor drift-diffusion: DDNet (Riganti et al., 2025) embeds the Poisson and continuity-equation residuals into a neural network loss to solve the same class of coupled PDEs that govern our devices, and Baelongandi et al. (2024) applied PINNs to solar cell modeling and prediction. However, such residual-based PINNs operate as *forward solvers*: they require retraining or extensive iterative optimization for every new device configuration, making them unsuitable for the high-throughput screening ($>10^6$ devices) targeted here. Our approach instead uses physics to *inform inputs* (71 derived features) and *constrain outputs* (monotonicity, convexity, curvature), avoiding PDE evaluation entirely and enabling instantaneous amortized inference across the full 31-dimensional parameter space. This is closer to the "physics-informed feature engineering" paradigm than to residual-based PINNs (Raissi et al., 2019).

6. Method

6.1. Problem Formulation

Given device parameters $\mathbf{x} \in \mathbb{R}^{31}$ and upstream scalar predictions $\{V_{oc}^{pred}, V_{mpp}^{pred}\}$, predict a compact J-V representation:

$$\hat{\mathbf{J}} = f_\theta(\mathbf{x}_{in}; \mathbf{V}_{query}) \in \mathbb{R}^8 \tag{10}$$

where $\mathbf{V}_{query}$ are 8 voltage query points concentrated near the MPP. Convention: positive photocurrent ($J > 0$ for $V < V_{oc}$). The full 45-point J-V curve is recovered via PCHIP interpolation from the 8 predicted points.

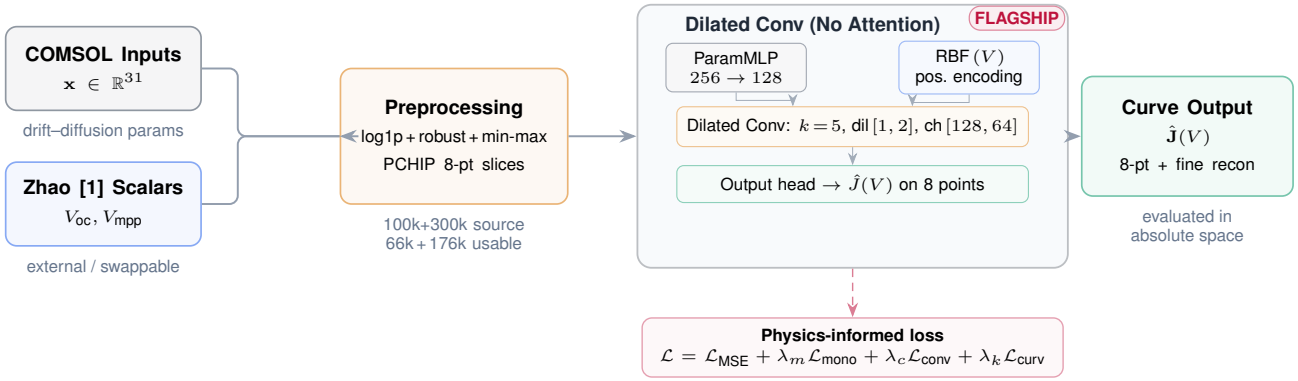


Figure 4. **Two-stage surrogate pipeline.** Stage 1: ANNs (Zhao et al., 2025) predict scalar metrics from 31 parameters. Stage 2: dilated conv model consumes scaled parameters, Stage 1 scalars, and m=5 physics features to produce 8-point J-V curves under physics-constrained loss.

Table 3. Comparison with concurrent J-V curve reconstruction approaches. Our method operates in the highest-dimensional parameter space with the largest training set and strongest physics constraints.

	Toprak (2025)	Zbinden et al. (2026)	**Ours**
Input dimensions	2	6	$31 + m$
Architecture	MLP	Conditional AE	Dilated Conv
Training data	8k	50k	150k
Best R^2	0.994	0.996	**0.998**
Physics constraints	–	Latent regularization	Mono+conv+curv
Inverse design	No	Encoder-based	Gradient-based
Experimental validation	No	4 devices	Via Zhao et al. (2025): 9 devices

6.2. Data

Two LHS campaigns generate the ground-truth dataset:

- **Campaign 1**: 100k samples → usable subset after COMSOL convergence filtering
- **Campaign 2**: 300k samples → usable subset after filtering
- **Combined**: ∼150k devices after convergence and quality filtering, split 80/10/10 train/val/test (120k/15k/15k)

Filtering removes only devices where COMSOL failed to converge or produced grossly unphysical curves (e.g., negative J_{sc}, $V_{oc} < 0$); no curvature-based or performance-based filtering is applied, preserving the full LHS distribution including pathological devices.

The voltage grid is **non-uniform** (Figure 5), with resolution concentrated where device-performance sensitivity is highest:

- **Low voltage** ($0 \le V \le 0.4$ V): 5 coarse points, $\Delta V = 0.1$ V
- **Knee/MPP region** ($0.425 \le V \le 1.4$ V): 40 dense points, $\Delta V = 0.025$ V
- **Total**: $N_V = 45$ points

This design reflects the physics: the flat-band region ($V \ll V_{mpp}$) has $J \approx J_{sc}$ with near-zero curvature, while the knee region contains the MPP, the drift-to-recombination transition, and the approach to V_{oc}, all critical for fill-factor evaluation and device optimization. The non-uniform spacing necessitates ΔV-weighted evaluation metrics (§6.8).

Dataset statistics. Figure 6 summarizes the ∼150k-device dataset. The LHS design achieves near-orthogonal parameter coverage (most pairwise correlations < 0.2), ensuring that the training data uniformly spans the design space. Figure 7 shows the per-voltage-point J distributions: spread is minimal at low V (flat-band: $J \approx J_{sc}$) and peaks near the knee/V_{mpp}, directly motivating the dense grid and curvature-based sample weighting.

6.3. Preprocessing

Parameter scaling. The >20 orders-of-magnitude dynamic range requires multi-stage normalization. Parameters are grouped by physical type, with group-specific transformations:

$$x \xrightarrow{\log_{1p}} x' \xrightarrow{\text{Robust}} \frac{x' - \tilde{x}'}{\text{IQR}} \xrightarrow{\text{MinMax}} x''' \in [-1, 1] \qquad (11)$$

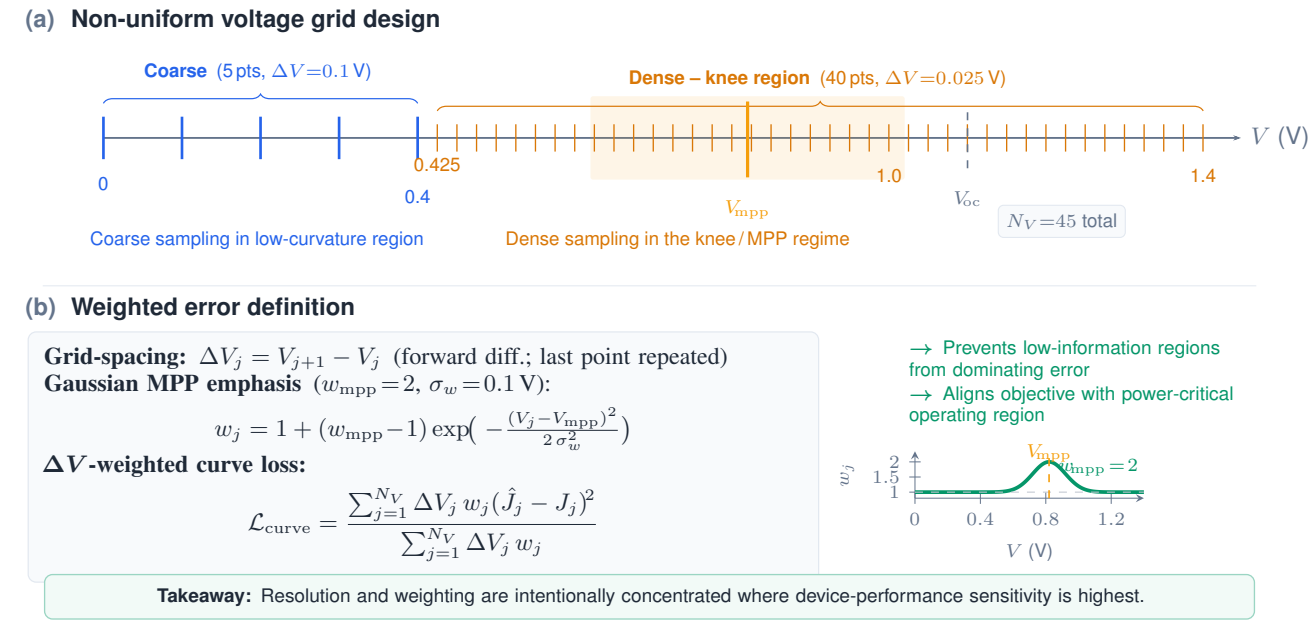


Figure 5. **Non-uniform voltage grid and ΔV-weighted loss.** (a) Coarse/dense grid design: 5 points at $\Delta V = 0.1$ V in the flat-band region, 40 points at $\Delta V = 0.025$ V in the knee/MPP regime, for $N_V = 45$ total. (b) ΔV-weighted curve loss with Gaussian emphasis w_j centered at V_{mpp} ($w_{mpp} = 2$, $\sigma_w = 0.1$ V), preventing low-information flat-band points from dominating the objective while concentrating fitting effort where power extraction and device-performance sensitivity are highest.

The four groups are:

- **Thicknesses** (3): $l_H, l_P, l_E \rightarrow$ RobustScaler $\rightarrow$ MinMax$[-1, 1]$
- **Material properties** (19): mobilities, DOS, energies, permittivities $\rightarrow \log_{1p} \rightarrow$ RobustScaler $\rightarrow$ MinMax$[-1, 1]$
- **Contacts** (2): $W_{anode}, W_{cathode} \rightarrow$ RobustScaler $\rightarrow$ MinMax$[-1, 1]$
- **Recombination/generation** (7): G_{avg}, Auger, B_{rad}, lifetimes, surface velocities $\rightarrow$ RobustScaler $\rightarrow$ MinMax$[-1, 1]$

RobustScaler uses median and interquartile range (IQR) rather than mean and standard deviation, providing resilience to outliers from the LHS tails. The $\log_{1p}$ transform compresses the multi-order-of-magnitude range of material properties before robust scaling. All transformer parameters are fit exclusively on the training split and stored via `joblib` for reproducible deployment. Figure 8 confirms that the multi-stage pipeline compresses the >20 orders-of-magnitude dynamic range into a well-behaved $[-1, 1]$ interval with outliers contained within $[-3, 3]$.

Compact curve representation. Rather than predicting all 45 grid points directly, each J-V curve is represented by **8 query points** concentrated near the MPP. The full 45-point curve is recovered by PCHIP interpolation (Fritsch & Carlson, 1980) with Fritsch–Carlson slopes, which preserves monotonicity by construction, simultaneously achieving dimensionality reduction ($45 \rightarrow 8$) and encoding a physics prior (monotone decreasing $J(V)$).

Figure 9 quantifies this approximation. The 8-point PCHIP reconstruction achieves $R^2 = 0.994$ against the full COMSOL reference; by $\sim$10 points the representation error saturates at $\sim$0.999. The 8-point budget is Pareto-optimal: sufficient fidelity for practical applications while keeping the sequence short enough for efficient convolutional processing.

Current normalization. Curves are J_{sc}-normalized to decouple magnitude from shape, enabling the network to learn shape independently of overall current scale:

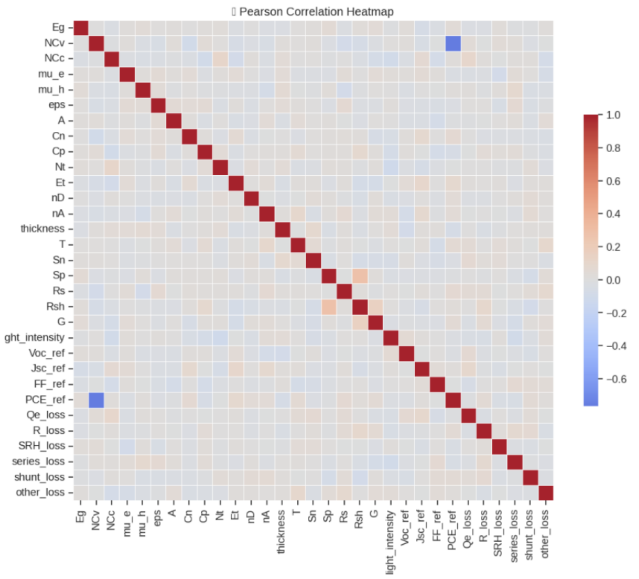
$$\tilde{J}(V) = 2\,J(V)/J_{sc} - 1 \in [-1, 1] \tag{12}$$

so that $\tilde{J}(0) = +1$ and $\tilde{J}(V_{oc}) = -1$. Denormalization recovers absolute current: $J(V) = (\tilde{J}(V) + 1) \cdot J_{sc}/2$.

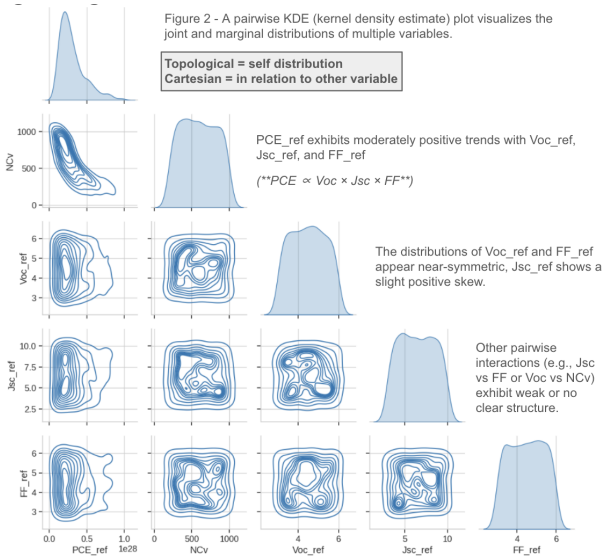
Training samples are additionally weighted by curvature:

$$w_i = 1 + 4.0\,(\kappa_i/\kappa_{max})^{1.5} \tag{13}$$

where κ_i is the maximum curvature of sample i. This emphasizes sharp-knee devices that are most informative for fill-factor prediction and most challenging to reconstruct.



(a) Pearson correlation heatmap of the 31 input parameters. Most pairwise correlations are < 0.2, confirming that the LHS design achieves near-orthogonal coverage. PCE shows negative correlation with N_{CV} via the chain $R \propto n_i^2 \propto N_c N_v \exp(-E_g/k_B T)$.



(b) Pairwise KDE of V_{oc}, J_{sc}, FF, PCE. Near-symmetric V_{oc} and FF distributions reflect the LHS coverage; right-skewed J_{sc} reflects the log-uniform distribution of generation-related parameters.

Figure 6. **Dataset statistics.** Near-orthogonal 31-parameter design space (a) and scalar metric distributions (b) from the combined LHS campaigns (∼150k usable devices after COMSOL convergence and quality filtering).

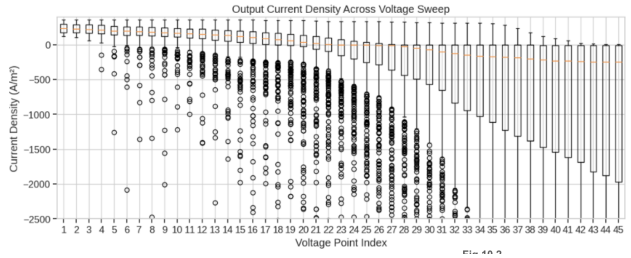


Figure 7. $J(V_j)$ **distributions per voltage point.** Spread is minimal at low V (flat-band region: $J \approx J_{sc}$ for all devices) and peaks near the knee/V_{mpp} region, motivating both the dense voltage grid and the curvature-based sample weighting that emphasizes sharp-knee devices.

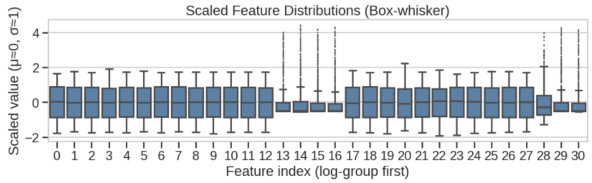


Figure 8. **All 31 parameters after normalization** ($\log_{1p} \rightarrow$ RobustScaler $\rightarrow$ MinMax$[-1, 1]$). Outliers are visible but contained within $[-3, 3]$, confirming that the multi-stage normalization handles the $>$20 orders-of-magnitude dynamic range effectively.

6.4. Physics Feature Engineering

From the raw 31 parameters, we derive 71 features encoding closed-form approximations of the drift-diffusion equations, the *same physics* that COMSOL solves numerically, expressed analytically (Table 4; full definitions in Appendix B). These features are computed once during preprocessing and cached as part of the input tensor.

Physically notable features include the **Hecht extraction figures of merit** (Courtier et al., 2019), which quantify the competition between charge extraction (drift) and recombination:

$$\Theta_e = \frac{\mu_n^P \tau_e V_{bi}}{l_P^2}, \qquad \Theta_h = \frac{\mu_p^P \tau_h V_{bi}}{l_P^2}, \qquad \Theta_{min} = \min(\Theta_e, \Theta_h) \tag{14}$$

$\Theta \gg 1$ indicates extraction-dominated transport (high fill factor, efficient charge collection); $\Theta \ll 1$ signals recombination-limited operation (low fill factor, S-shaped risk). Although Θ_{min} is an important physical concept, it does not survive train-only feature selection (§6.5) because its predictive information is already encoded by the combination of upstream features (E_g^{offset}, J_{max}^{log}) that capture the same V_{bi}, l_P, and transport physics through less redundant coordinates.

Composite predictors combine multiple physical mecha-

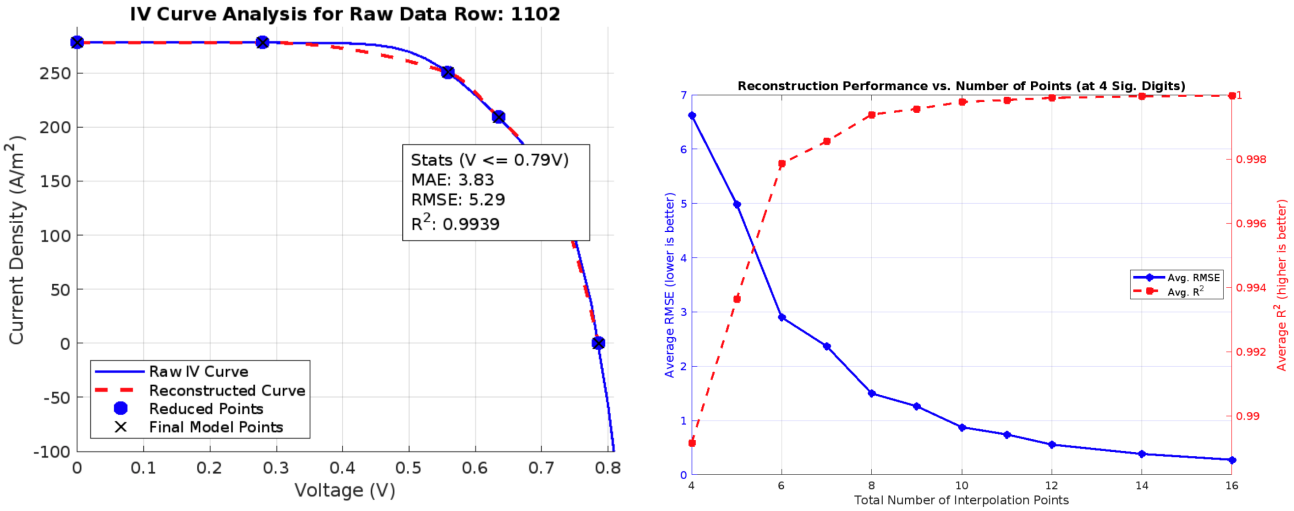


Figure 9. **PCHIP reconstruction fidelity. Left**: 8-point reconstruction (dashed orange) vs. COMSOL reference (solid blue), $R^2 = 0.994$. The representation faithfully captures both the flat photocurrent plateau and the sharp knee transition. **Right**: R^2 vs. number of query points; performance saturates at $\sim$0.999 by 10 points and $\sim$0.9997 by 12 points. The 8-point budget is Pareto-optimal, balancing fidelity against model output dimensionality and sequence length for the convolutional backbone.

Table 4. Physics feature taxonomy (71 total). Each category derives from a specific aspect of the drift-diffusion equations. Features encode the same physics COMSOL solves numerically, providing the network with physically meaningful coordinates that shortcut nonlinear representation learning.

Category	**Physical origin**	n	**Key features**
Energetics	Band theory	7	E_g, V_{bi}, E_g^{offset}
Barriers	Band alignment	9	ΔE_v^{HP}, $\Phi_{h,ETL}$, $\Phi_{e,HTL}$
Diffusion lengths	Einstein + SRH	8	$L_{n,p}$, $\eta_{coll,e/h}$
$\mu\tau$ products	Transport quality	4	$\mu\tau_{e,h}$, $\mu\tau_{balance}$
Extraction FOMs	Hecht equation	6	$\Theta_{e,h}$, Θ_{min}
Series resistance	Ohm's law	5	$R_s^{HTL,ETL,total}$
Generation	Beer–Lambert	4	$J_{max} = qG_{avg}l_P$
Recombination	SRH/Auger theory	8	τ_{eff}, V_{oc}^{loss}
Surface	Interface recombination	4	v_s, surface/bulk ratio
Geometry	Electrostatics	9	ε-ratios, Debye λ_D
Composite	Multi-physics combinations	7	FF_{pred}, Quality, J_{sc}^{pred}

nisms into quantities that approximate device-level metrics:

$$\text{FF}_{pred} = \log \Theta_{min} - \log R_s^{total} \tag{15}$$

$$V_{oc}^{pred} = E_g + \tfrac{1}{2}(\Phi_{h,ETL} + \Phi_{e,HTL}) - k_B T \log R_{SRH} \tag{16}$$

$$J_{sc}^{pred} = \log(J_{max}) + \log(\eta_{coll,min}) \tag{17}$$

$$\text{Quality} = \log(\Theta_{min}) + \text{SRH}_{str} - \log(R_s^{total}) \tag{18}$$

These provide the network with physically meaningful coordinates, shortcutting the nonlinear combinations (products, ratios, exponentials of 6 or more raw parameters) that it would otherwise need to discover.

Analytical ceiling functions bound scalar predictions:

$$J_{sc}^{ceil} = qG_{avg}l_P \qquad \text{(Beer–Lambert absorption limit)} \tag{19}$$

$$V_{oc}^{ceil} = \min(E_g, V_{bi}) \qquad \text{(thermodynamic + electrostatic bound)} \tag{20}$$

$$V_{mpp}^{est} = V_{oc} - k_B T \ln(1 + V_{oc}/k_B T) \qquad \text{(Shockley approximation)} \tag{21}$$

$$\text{FF}^{est} = \frac{v_{oc} - \ln(v_{oc} + 0.72)}{v_{oc} + 1}, \quad v_{oc} = V_{oc}/(k_B T) \qquad \text{(Green formula)} \tag{22}$$

6.5. Physics-Guided Feature Selection

While 71 candidate features encode the full analytical structure of the drift-diffusion equations, J-V curve shape is

fundamentally governed by a small number of independent physical mechanisms. We selected five features based on physical reasoning about which mechanisms independently govern J-V behavior, then used multicollinearity and relevance analysis to support this selection.

Physics-based selection. The drift-diffusion equations (Eqs. 1–4) govern J-V behavior through five largely independent mechanisms: (i) *photocurrent generation*, setting the absolute current scale; (ii) *band gap energetics*, determining the absorption edge and thermodynamic voltage limit; (iii) *built-in field alignment*, controlling charge extraction efficiency; (iv) *interface band alignment*, governing carrier transfer at heterojunctions; and (v) *non-radiative recombination*, dictating the open-circuit voltage deficit from the Shockley–Queisser limit. Each mechanism maps to a single dominant feature:

1. $\boldsymbol{J_{max}^{log} = \log(qG_{avg}l_P)}$, the Beer–Lambert photocurrent ceiling. This is the strongest single predictor of J-V curve shape ($\max|r| = 0.973$ with target current values), as it sets the absolute photocurrent scale from which all other device-level metrics derive. J_{max} combines absorption (G_{avg}) and geometry (l_P), capturing the dominant source of inter-device variability in the current axis.

2. $\boldsymbol{E_g}$, the perovskite band gap ($\max|r| = 0.432$). E_g directly determines the absorption edge and the thermodynamic upper bound for V_{oc} via the Shockley–Queisser limit. Changes in E_g shift the entire J-V curve along the voltage axis and modulate J_{sc} through the solar spectrum overlap.

3. $\boldsymbol{E_g^{offset} = E_g - V_{bi}}$, the mismatch between band gap and built-in potential ($\max|r| = 0.324$). When $E_g^{offset} > 0$, the internal electric field (V_{bi}/l_P) is insufficient to sweep carriers across the full energy range, causing voltage-dependent extraction losses that reshape the knee region and degrade fill factor. This feature encodes the competition between the thermodynamic voltage window (E_g) and the electrostatic driving force ($V_{bi} = W_{anode} - W_{cathode}$).

4. $\boldsymbol{\chi_P^e}$, the perovskite electron affinity ($\max|r| = 0.334$). χ_P^e determines the conduction band offset at the perovskite/ETL interface ($\Delta E_c^{PE} = \chi_{ETL}^e - \chi_P^e$), governing whether electrons face an extraction barrier. Large $|\Delta E_c^{PE}|$ produces S-shaped kinks in the J-V curve; moderate alignment enables efficient electron collection.

5. $\boldsymbol{V_{oc}^{loss} = -E_g/(k_BT) + \log(R_{SRH})}$, which is a composite recombination metric ($\max|r| = 0.432$). This feature encodes the competition between the thermodynamic V_{oc} limit (set by E_g) and non-radiative SRH recombination losses, directly predicting the voltage at which the J-V curve crosses zero. The logarithmic form linearizes the exponential dependence of J_0 on recombination rates.

Together, these five features span the primary determinants of J-V shape: photocurrent magnitude (J_{max}), voltage limits (E_g, V_{oc}^{loss}), fill factor (E_g^{offset}), and carrier extraction (χ_P^e).

Statistical support. A two-stage statistical analysis supports the physics-based selection:

1. **Multicollinearity analysis**: among all 71 candidate features, 175 pairs exhibit Pearson $|r| > 0.85$, revealing extensive redundancy. For example, the three generation-related features (J_{max}^{log}, $G_{per\,nm}^{log}$, G_{avg}^{log}) are mutually correlated ($|r| > 0.9$) because they differ only by factors of q and l_P; similarly, all Hecht extraction figures of merit (Θ_e, Θ_h, Θ_{min}) and diffusion length ratios ($\eta_{coll,e}$, $\eta_{coll,h}$) are collinear with l_P-dependent features ($|r| > 0.93$). Dropping the less informative member of each correlated pair eliminates 34 features (71→37).

2. **Relevance filter**: of the remaining 37 features, only those with $\max_{y\in\text{targets}}|r(\phi_i, y)| \geq 0.3$ are retained, removing 32 features with negligible independent predictive power, including all series resistance features (R_s^{HTL}, R_s^{ETL}, R_s^{total}; $|r| < 0.001$), all permittivity-related and thickness-ratio features, and all Auger and radiative recombination terms ($|r| = 0.0$).

The statistical filters independently converge on the same $m = 5$ features (71→37→5), supporting the physics-based selection. The multicollinearity structure reveals *why* the other 66 features are redundant: they encode overlapping aspects of the same underlying transport physics. The drift-diffusion equations impose strong inter-feature correlations (e.g., Θ_{min} is algebraically related to L_n, $\mu\tau$, and η_{coll} through shared dependence on μ, τ, V_{bi}, and l_P), making most intermediate quantities redundant once the five governing mechanisms are directly represented.

Why not Θ_{min} or R_s^{total}? Although the Hecht extraction parameter Θ_{min} and total series resistance R_s^{total} are physically important concepts (Courtier et al., 2019), they do not survive the selection procedure. Θ_{min} achieves only $\max|r| = 0.189$ with target variables, below the relevance threshold, because its predictive information is already captured by the combination of E_g^{offset} (which encodes the V_{bi}/l_P^2 dependence) and J_{max}^{log} (which encodes l_P and generation effects). R_s^{total} achieves $|r| < 0.001$, reflecting the fact that in the LHS parameter space, most devices have

sufficiently low series resistance that R_s does not limit J-V shape; resistive effects only dominate in the extreme tail of the distribution (§2.3).

Interpretation: low-dimensional physics. The compression from 71 to 5 reveals that the physics feature space is highly structured. While the generalized Shockley diode equation relates J-V shape to 3–4 effective parameters (J_{sc}, J_0, n, R_{sh}), it cannot be directly fit to the 31-dimensional coupled drift-diffusion parameter space because it assumes idealized single-junction behavior without interface barriers, spatially varying fields, or multi-layer transport. Our five selected features provide a physically interpretable low-dimensional coordinate system that goes beyond the Shockley framework: J_{max}^{log} captures photocurrent generation, E_g and V_{oc}^{loss} encode the voltage-axis physics, and E_g^{offset} and χ_P^e capture the multi-layer transport effects (field alignment, interface barriers) that the diode equation cannot represent.

The feature mask is computed exclusively on the training split and stored alongside the model checkpoint; no target information from validation or test data influences feature selection. A hash of the training indices is stored for reproducibility verification.

6.6. Architecture

The full forward pass maps device parameters to 8 current-density predictions:

$$\underbrace{\mathbf{x}_{in} \in \mathbb{R}^{d_{in}}}_{31+2+m} \xrightarrow{\text{ParamMLP}} \mathbf{h} \in \mathbb{R}^{128}$$
$$\xrightarrow[\text{concat}]{\oplus\,\text{RBF}(\mathbf{V})} \mathbf{S} \in \mathbb{R}^{8\times 256} \xrightarrow{\text{DilConv}\times 3} \xrightarrow{\text{Lin}} \hat{\mathbf{J}} \in \mathbb{R}^{8} \qquad (23)$$

ParamMLP. A three-layer MLP embeds the concatenated input ($\mathbf{x}^{sc}$, scalars, selected features) into a fixed-size representation: $d_{in} \to 256 \to 128 \to 128$, with GELU activation, BatchNorm, and Dropout(p=0.036) after each layer. The output $\mathbf{h} \in \mathbb{R}^{128}$ is a device-level embedding shared across all voltage query points.

Gaussian RBF positional encoding. Inspired by Fourier feature encodings that enable MLPs to learn high-frequency functions in low-dimensional domains (Tancik et al., 2020), each voltage query point V_j is encoded via learnable Gaussian radial basis functions:

$$\text{RBF}(V_j) = \left[\exp\!\left(-(V_j - \mu_k)^2/2\sigma_k^2\right)\right]_{k=1}^{128} \qquad (24)$$

where centers μ_k are initialized uniformly over $[0, V_{max}]$ and bandwidths σ_k are learnable. The 128-dimensional device embedding $\mathbf{h}$ is broadcast across 8 voltage positions and concatenated with the 128-dim RBF encoding, producing $\mathbf{S} \in \mathbb{R}^{8\times 256}$, a sequence representation where each position encodes both device properties and voltage location.

Dilated 1D conv backbone. Drawing on the temporal convolutional network (TCN) paradigm (Bai et al., 2018), three residual blocks with symmetric (non-causal) padding process the sequence:

Block	Ch.	k	d	Details
1	256→128	5	1	Conv-BN-GELU-Drop + skip
2	128→128	5	1	Residual connection
3	128→64	5	2	RF>8: full bidir. coverage

Output: pointwise Linear($64 \to 1$) $\to \hat{\mathbf{J}} \in \mathbb{R}^8$.

Why dilation suffices. Block 3's dilation d=2 with kernel k=5 gives a single-layer receptive field of 9 positions ($= k + (k-1)(d-1)$), already exceeding the 8-point sequence length. Combined with Blocks 1-2, every output position has *bidirectional* context over all 8 input positions. This is the architectural mechanism that makes self-attention unnecessary (§2.2): global context is already provided by the convolutional architecture, and the local inductive bias of convolution better matches the monotonically decaying spatial correlation structure of J-V curves.

6.7. Physics-Constrained Loss

The total loss combines reconstruction accuracy with four physics-motivated penalties:

$$\mathcal{L} = \underbrace{0.98 \cdot \mathcal{L}_{MSE}}_{\text{reconstruction}} + \underbrace{0.005 \cdot \mathcal{L}_{mono}}_{\text{monotonicity: } dJ/dV \le 0} + \underbrace{0.005 \cdot \mathcal{L}_{conv}}_{\text{convexity: } d^2J/dV^2 \le 0}$$
$$+ \underbrace{0.01 \cdot \mathcal{L}_{curv}}_{\text{curvature bound}} + \underbrace{\lambda_{jac} \cdot \mathcal{L}_{jac}}_{\text{Jacobian smoothness}} \qquad (25)$$

Each penalty encodes a qualitative property of physical J-V curves:

Reconstruction loss. Weighted mean squared error on the 8 predicted current values:

$$\mathcal{L}_{MSE} = \frac{1}{8B} \sum_{i=1}^{B} \sum_{j=1}^{8} w_{ij} (\hat{J}_{ij} - J_{ij})^2 \qquad (26)$$

where w_{ij} incorporates sample curvature weighting (Eq. 13).

Monotonicity penalty. Current density must decrease with voltage in the power-producing quadrant ($dJ/dV \le 0$):

$$\mathcal{L}_{mono} = \frac{1}{7B} \sum_{i,j} [\text{ReLU}(\hat{J}_{i,j+1} - \hat{J}_{i,j})]^2 \qquad (27)$$

Penalizes any predicted increase in current with voltage, which is a direct encoding of the fundamental physics that net photocurrent decreases as the applied voltage reduces the built-in field and enhances recombination.

Convexity penalty. In the knee region, J-V curves are concave (negative second derivative):

$$\mathcal{L}_{conv} = \frac{1}{6B}\sum_{i,j}[\text{ReLU}(2\hat{J}_{i,j} - \hat{J}_{i,j-1} - \hat{J}_{i,j+1})]^2 \quad (28)$$

Excess curvature penalty. Bounds the magnitude of the second finite difference to prevent oscillatory artifacts:

$$\mathcal{L}_{curv} = \frac{1}{6B}\sum_{i,j}[\text{ReLU}(|\Delta^2\hat{J}_{ij}| - 0.8)]^2 \quad (29)$$

Jacobian regularization. A smoothness prior on the parameter-to-curve mapping:

$$\mathcal{L}_{jac} = \frac{1}{d_{in}}\left\|\frac{\partial\hat{\mathbf{J}}}{\partial\mathbf{x}_{in}}\right\|_F^2 \quad (30)$$

estimated via Hutchinson's trace estimator with K=4 random projections in fp32 (bf16 causes overflow due to the squared norm computation). Default $\lambda_{jac} = 0$; recommended range $[10^{-4}, 10^{-2}]$ when Jacobian smoothness is desired for inverse design (§8).

6.8. ΔV-Weighted Evaluation Metrics

The non-uniform voltage grid necessitates ΔV-weighted metrics to ensure that evaluation reflects physical significance rather than grid density. Without weighting, the 40 dense knee-region points ($\Delta V = 0.025$ V) would dominate the 5 coarse flat-band points ($\Delta V = 0.1$ V), making errors in the flat-band region invisible.

The ΔV-weighted per-curve R^2:

$$R_i^2 = 1 - \frac{\sum_j \Delta V_j\big(\hat{J}_i(V_j) - J_i(V_j)\big)^2}{\sum_j \Delta V_j\big(J_i(V_j) - \bar{J}_i\big)^2} \quad (31)$$

where $\bar{J}_i = \sum_j \Delta V_j J_i(V_j)/\sum_j \Delta V_j$ and $\Delta V_j = V_{j+1} - V_j$ (last point repeated).

The ΔV-weighted curve loss used during training:

$$\mathcal{L}_{curve} = \frac{\sum_{j=1}^{N_V}\Delta V_j\, w_j\,(\hat{J}_j - J_j)^2}{\sum_{j=1}^{N_V}\Delta V_j\, w_j} \quad (32)$$

with Gaussian MPP emphasis ($w_{mpp} = 2$, $\sigma_w = 0.1$ V):

$$w_j = 1 + (w_{mpp} - 1)\exp\left(-\frac{(V_j - V_{mpp})^2}{2\sigma_w^2}\right) \quad (33)$$

We report mean, median, 5th percentile, and 95th percentile R_i^2 aggregates across the test set.

Table 5. Training configuration. Hyperparameters selected via Optuna TPE sampler (50 trials with median pruning; Table 6).

Component	Parameter	Value
Optimizer	AdamW	β_1=0.9, β_2=0.999
Learning rate	OneCycleLR	η_{pk}=5.5×10^{-3}
Weight decay		5.4×10^{-5}
Batch size		512 (primary); 128 (abl.)
Precision		bf16 (fp32 for Jacobian)
Max epochs		100
Early stopping		20 ep. patience on val loss
Grad. clipping		$\|\nabla\|_{max} = 1.0$

Table 6. HPO search space and optimal values (Optuna TPE, 50 trials, median pruning).

Hyperparameter	Range	Best
Learning rate η	$[10^{-3}, 2\times10^{-2}]$	5.5×10^{-3}
Weight decay	$[10^{-6}, 10^{-4}]$	5.4×10^{-5}
Dropout	$[0, 0.2]$	0.036
Kernel size k	$\{3, 5, 7, 9\}$	5
Conv blocks	$[1, 3]$	3
Channel dims	$\{32, 64, 128\}$	128→128→64
RBF bands	$[12, 28]$	19
RBF σ_{init}	$[0.05, 0.15]$	0.128
MLP depth	$[1, 4]$	3
MLP width	$\{64, \ldots, 512\}$	256
w_{mono}	$[10^{-4}, 0.05]$	0.005
w_{conv}	$[10^{-4}, 0.05]$	0.005

6.9. Training

Training uses AdamW with OneCycleLR scheduling (peak $\eta = 5.5 \times 10^{-3}$, cosine annealing to $\eta_{peak}/150$), bf16 mixed precision, and gradient clipping at $\|\nabla\|_{max} = 1.0$ to stabilize early epochs. The primary model uses batch size 512; ablation studies use batch size 128. Early stopping with 20-epoch patience on validation loss prevents overfitting. Hyperparameters are selected via Optuna TPE sampler with 50 trials and median pruning; Table 6 reports the search space and optimal values. Total training time: ∼10 min on a single GPU (batch 512); ∼23 min for batch 128 configurations.

7. Analysis and Discussion

Physics features as PDE summaries. Without the derived features, the MLP must discover nonlinear combinations such as:

$$J_{max} = qG_{avg}l_P \quad \text{(3 raw params)}$$
$$V_{oc}^{loss} = -E_g/(k_BT) + \log(R_{SRH}) \quad \text{(4+ raw params)}$$
$$E_g^{offset} = (E_g^h - E_g^e) - (W_{anode} - W_{cathode}) \quad \text{(4 raw params)}$$

spanning 6 orders of magnitude. With the derived features, the MLP learns only the residual mapping from precomputed physics to J-V shape. The most impactful fea-

ture is J_{max}^{log} (photocurrent ceiling, $|r| = 0.973$), followed by G_{avg}^{log} (generation rate, $|r| = 0.582$), E_g (band gap, $|r| = 0.432$), and V_{oc}^{loss} (recombination loss, $|r| = 0.432$).

Upstream vs. downstream features. A notable pattern in the feature selection (§6.5) is that all five surviving features are "upstream" quantities, which are fundamental material properties and their direct combinations (J_{max}^{log}, E_g, E_g^{offset}, χ_P^e, V_{oc}^{loss}), instead of "downstream" derived transport metrics (Θ_{min}, L_n, R_s^{total}). This suggests that the network benefits most from physically interpretable coordinates that are minimally redundant, rather than from pre-computed transport summaries that share extensive mutual information through their common dependence on the raw parameters.

Scalar-to-curve error propagation. The two-stage pipeline's modular design raises a natural question: how sensitive is curve quality to upstream scalar accuracy? As noted in §4, the primary results use ground-truth V_{oc} and V_{mpp} to characterize the curve stage's best-case performance. The ablation removing scalar conditioning entirely (Table 7, row 8) shows that curve R_{med}^2 changes from 0.9975 to 0.9966, a marginal difference ($\Delta R_{med}^2 < 0.001$) confirming that the curve model is only weakly sensitive to its scalar inputs. Since Zhao et al. (2025)'s neural network surrogates achieve $R^2 = 0.996$ for V_{oc} prediction and tree-based regressors recover V_{mpp} at $R^2 = 0.99$, the small additional error introduced by replacing ground-truth with predicted scalars is expected to fall well within this already-marginal sensitivity band. The dominant source of reconstruction error is therefore intrinsic to the curve model, not propagated from Stage 1.

Comparison with related approaches. Our model's median $R^2 = 0.998$ exceeds both Toprak ($R^2 = 0.994$ on a 2-input single-device problem) (Toprak, 2025) and Zbinden et al. (2026) ($R^2 = 0.996$ on a 6-parameter problem), while operating in a 31-dimensional parameter space with explicit physics constraints. The upstream scalar predictors have been experimentally validated against 9 fabricated perovskite devices spanning $E_g \in \{1.56, 1.63\}$ eV (Zhao et al., 2025), with calibration MSE ranging from 8.2×10^{-8} to 1.8×10^{-4}. Zbinden et al. (2026) provide complementary validation on 4 devices, supporting drift-diffusion–trained surrogates as viable for real-world deployment.

8. Applications

High-throughput screening. A 10^6-device Latin Hypercube sweep (Chen et al., 2025; Wen et al., 2025), ∼160 years via COMSOL at ∼4800 s/device, completes in <1 hour with the surrogate. Full-curve evaluation captures fill-factor degradation mechanisms and S-shaped kinks that scalar-only screening (Hu et al., 2022) misses entirely, enabling discovery of devices with high nominal PCE but poor operational stability signatures. The curve representation also enables voltage-dependent analysis: identifying which devices maintain current at high voltages (desirable for tandem applications) vs. which exhibit premature roll-off.

Inverse design via Jacobian. The differentiable surrogate provides analytic Jacobian $\partial\hat{\mathbf{J}}/\partial\mathbf{x}$, enabling gradient-based optimization toward target J-V properties (Wang et al., 2022; Lu et al., 2022; Lee et al., 2020) (Figure 10):

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta\, \nabla_{\mathbf{x}}\mathcal{L}_{target}, \qquad \nabla_{\mathbf{x}}\mathcal{L}_{target} = \frac{\partial \mathcal{L}_{target}}{\partial \hat{\mathbf{J}}} \cdot \frac{\partial \hat{\mathbf{J}}}{\partial \mathbf{x}} \tag{34}$$

Unlike scalar optimization ("maximize PCE"), curve-level inverse design can target specific J-V shape properties: maximum fill factor, smooth knee curvature, or matching a reference profile from a champion device. Combined with the 24 s experimental calibration protocol of Zhao et al. (2025), this enables a closed-loop *characterize* → *predict* → *optimize* workflow within a single experimental session.

Sensitivity analysis. The Jacobian $\partial\hat{\mathbf{J}}/\partial\mathbf{x}$ also enables systematic sensitivity analysis: which parameters most strongly influence each region of the J-V curve. This produces physics-interpretable results, e.g., mobilities and lifetimes dominate the knee region, while contact work functions primarily affect V_{oc}, that can guide experimental efforts toward the parameters with highest leverage for improving specific aspects of device performance.

9. Conclusion

We presented a physics-constrained dilated convolutional network for J-V curve reconstruction in perovskite solar cells, achieving median $R^2 = 0.9975$ across ∼15k test devices spanning 31 coupled drift-diffusion parameters with >20 orders of magnitude in dynamic range. The model reconstructs V_{oc} to 0.39 mV median error (below experimental measurement uncertainty) and median MAE of 2.96 A/m^2 (<0.5% of typical J_{sc}), completing training in ∼10 min and enabling $>10^4\times$ speedup over COMSOL FEM.

Physics knowledge is embedded at every stage of the pipeline: analytically derived drift-diffusion features reduce the 71-candidate input space to 5 features spanning independent physical mechanisms, soft monotonicity, convexity, and curvature penalties enforce qualitative PDE properties on the output, and a modular two-stage design decouples scalar prediction from curve reconstruction. The architectural choice of bidirectional dilated convolutions is grounded in the physics of J-V responses: thermodynamic equilibria (not temporal sequences) with monotonically decaying

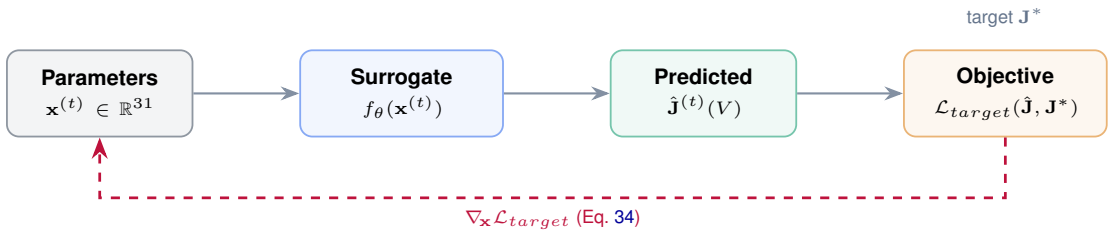


Figure 10. **Inverse design loop.** The frozen surrogate f_θ provides analytic Jacobian gradients for iterative parameter updates toward a target J-V profile $\mathbf{J}^*$. Unlike scalar optimization, curve-level objectives can target specific shape properties (fill factor, knee smoothness, reference matching).

spatial correlations that match convolution's local inductive bias, rendering self-attention redundant and consistently harmful on these short sequences. The differentiable surrogate additionally enables gradient-based inverse design via analytic Jacobians, targeting specific J-V shape properties rather than scalar metrics alone.

Future directions include: deep-ensemble uncertainty quantification for deployment safety, Bayesian optimization for closed-loop experimental design, extension to tandem perovskite/silicon architectures (Zhao, 2023), end-to-end joint training of both pipeline stages, and out-of-distribution detection to flag devices outside the training manifold.

References

Baelongandi, A. T., Kambale, W. V., Nzung, A. M., Bampende, M. N., Kadurha, D. K., and Kyamakya, K. Towards a physics informed neural network approach for modeling and prediction of renewable energy production: Use case of solar cells. In *2024 IEEE PES/IAS PowerAfrica*, pp. 01–05, 2024. doi: 10.1109/PowerAfrica61624.2024.10759416.

Bai, S., Kolter, J. Z., and Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Chen, C., Xiao, J., and Wang, Z. Multidimensional high-throughput screening for mixed perovskite materials with machine learning. *The Journal of Chemical Physics*, 162 (11):114101, 2025. doi: 10.1063/5.0251300.

Courtier, N. E. et al. How transport layer properties affect perovskite solar cell performance. *Energy & Environmental Science*, 12:396–409, 2019.

Djeradi, S., Dahame, T., Fadla, M. A., Bentria, B., Kanoun, M. B., and Goumri-Said, S. High-throughput ensemble-learning-driven band gap prediction of double perovskites solar cells absorber. *Machine Learning and Knowledge Extraction*, 6(1):435–447, 2024. doi: 10.3390/make6010022.

Feng, S. et al. Prediction of organic-inorganic hybrid perovskite band gap by multiple ML algorithms. *Molecules*, 29:499, 2024.

Fritsch, F. N. and Carlson, R. E. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17 (2):238–246, 1980.

Green, M. A., Ho-Baillie, A., and Snaith, H. J. The emergence of perovskite solar cells. *Nature Photonics*, 8(7): 506–514, 2014. doi: 10.1038/nphoton.2014.134.

Hu, Y. et al. Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells. *Advanced Energy Materials*, 12, 2022.

Lee, X. Y., Waite, J., Yang, C.-H., Pokuri, B., Joshi, A., Balu, A., Hegde, C., Ganapathysubramanian, B., and Sarkar, S. *Fast Inverse Design of Microstructures via Generative Invariance Networks*. 2020. doi: 10.21203/rs.3.rs-88996/v1.

Li, J., Li, R., Jia, Y., and Zhang, Z. Prediction of I–V characteristic curve for photovoltaic modules based on convolutional neural network. *Sensors*, 20(7):2119, 2020. doi: 10.3390/s20072119.

Lu, T., Li, H., Li, M., Wang, S., and Lu, W. Inverse design of hybrid organic–inorganic perovskites with suitable bandgaps via proactive searching progress. *ACS Omega*, 7(25):21583–21594, 2022. doi: 10.1021/acsomega.2c01380.

Oboh, I. O. et al. Explainable ML for predicting band gaps of ABX3 perovskites. *Materials Science in Semiconductor Processing*, 161:107427, 2023.

Porowski, R., Kowalik, R., Gorzelnik, T., et al. Application of machine learning to predict I-V characteristics of PV modules based on steady-state solar simulator. *Preprints*, 2024. doi: 10.20944/preprints202406.0748.v1.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear

partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

Ren, Z., Tian, S., Heumueller, T., Birgersson, E., et al. Physics-guided characterization and optimization of solar cells using surrogate machine learning model. pp. 3054–3058, 2019. doi: 10.1109/PVSC40753.2019.8980715.

Riganti, R. et al. DDNet: A unified physics-informed deep learning solver for the drift-diffusion equations of semiconductor devices. *arXiv preprint arXiv:2509.08073*, 2025.

Suprun, N. *Predictive Modelling For Enhanced Performance: Exploring Machine Learning Techniques To Optimize Perovskite Solar Cell Design*. PhD thesis, National University of Singapore, April 2024.

Talapatra, A. et al. A machine learning approach for formability and stability of perovskite oxides. *Chemistry of Materials*, 33:845–858, 2021.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*, 2020.

Toprak, A. High-accuracy ML approach for predicting J-V characteristics of perovskite solar cells. *Scientific Reports*, 15:41304, 2025.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, pp. 6000–6010, 2017.

Wang, J., Wang, Y., and Chen, Y. Inverse design of materials by machine learning. *Materials*, 15(5):1811, 2022. doi: 10.3390/ma15051811.

Wen, J., Yang, S., Jiang, L., Shi, Y., Huang, Z., Li, P., Xiong, H., Yu, Z., Zhao, X., Xu, B., Wu, B., Sa, B., and Qiu, Y. Accelerated discovery of high-performance small-molecule hole transport materials via molecular splicing, high-throughput screening, and machine learning. *Journal of Materials Informatics*, 5(3):30, 2025. doi: 10.20517/jmi.2024.102.

Xue, H., Birgersson, E., and Stangl, R. Correlating variability of modeling parameters with photovoltaic performance: Monte Carlo simulation of a meso-structured perovskite solar cell. *Applied Energy*, 237:131–144, 2019. doi: 10.1016/j.apenergy.2018.12.066.

Xue, H., Stangl, R., and Birgersson, E. Elucidating the functional form of the recombination losses in a planar perovskite solar cell: A scaling analysis. *Journal of Applied Physics*, 128(12):123102, 2020. doi: 10.1063/5.0013741.

Zbinden, O., Comi, E. L., Knapp, E., and Tress, W. Autoencoder for parameter estimation and J-V simulation of perovskite solar cells. *npj Computational Materials*, 12:7, 2026.

Zhang, Y. and Xu, X. Machine learning lattice constants for cubic perovskite compounds. *ChemistrySelect*, 5:9999–10009, 2020.

Zhao, X. *Mathematical Modeling of Two-terminal Perovskite-Based Thin-film Tandem Solar Cells*. PhD thesis, National University of Singapore, 2023.

Zhao, X., Huang, C., Birgersson, E., Suprun, N., Tan, H. Q., Zhang, Y., Jiang, Y., Shou, C., Sun, J., Peng, J., and Xue, H. Accelerating device characterization in perovskite solar cells via neural network approach. *Applied Energy*, 392:125922, 2025.

A. Optical Model Details

Light propagation follows the vectorized transfer-matrix formulation (Zhao, 2023). The electric field in layer n:

$$E_z(x) = A_n e^{-ik_n(x-L_{n-1})} + B_n e^{ik_n(x-L_{n-1})} \tag{35}$$

where $k_n = k_0(n_n - i\kappa_n)$. Photogeneration rate in the absorber:

$$G(x) = \frac{1}{hc}\int_{\lambda_l}^{\lambda_u} \lambda\, Q(x,\lambda)\, d\lambda \tag{36}$$

$$Q = \frac{1}{2} c\varepsilon_0 \alpha n\, |E_z|^2, \quad \alpha = \frac{4\pi\kappa}{\lambda} \tag{37}$$

A phase-elimination method removes incoherent substrate interference. $G_{avg} = l_P^{-1}\int_0^{l_P} G(x)\, dx$ enters as one of the 31 inputs.

B. Full Physics Feature Definitions

Energetics (7)

$$E_g = \chi_P^h - \chi_P^e, \quad V_{bi} = W_{anode} - W_{cathode}, \quad E_g^{offset} = E_g - V_{bi} \tag{38}$$

Interface Barriers (9)

$$\Delta E_v^{HP} = \chi_{HTL}^h - \chi_P^h, \quad \Delta E_c^{PE} = \chi_{ETL}^e - \chi_P^e \tag{39}$$

$$\Phi_{h,ETL} = \chi_{ETL}^h - \chi_P^h, \quad \Phi_{e,HTL} = \chi_{HTL}^e - \chi_P^e \tag{40}$$

Diffusion Lengths (8)

$$L_n = \sqrt{\mu_n^P \tau_e k_B T/q}, \quad L_p = \sqrt{\mu_p^P \tau_h k_B T/q} \tag{41}$$

$$\eta_{coll,e} = L_n/l_P, \quad \eta_{coll,h} = L_p/l_P \tag{42}$$

$\mu\tau$ Products (4)

$$\mu\tau_e = \mu_n^P \cdot \tau_e, \quad \mu\tau_h = \mu_p^P \cdot \tau_h, \quad \mu\tau_{bal} = \mu\tau_e/\mu\tau_h \tag{43}$$

Extraction Figures of Merit (6)

$$\mathcal{E}_{bi} = V_{bi}/l_P, \quad \Theta_e = \mu_n^P \tau_e V_{bi}/l_P^2 \tag{44}$$

$$\Theta_h = \mu_p^P \tau_h V_{bi}/l_P^2, \quad \Theta_{min} = \min(\Theta_e, \Theta_h) \tag{45}$$

Series Resistance (5)

$$\sigma_{HTL} = q\mu_p^{HTL} N_v^{HTL}, \quad R_s^{HTL} = l_{HTL}/\sigma_{HTL} \tag{46}$$

$$R_s^{total} = R_s^{ETL} + R_s^P + R_s^{HTL} \tag{47}$$

Generation (4)

$$J_{max} = q \cdot G_{avg} \cdot l_P, \quad G_{per\,nm} = G_{avg}/l_P \tag{48}$$

Recombination (8)

$$\tau_{eff} = (\tau_e^{-1} + \tau_h^{-1})^{-1}, \quad V_{oc}^{loss} = -E_g/(k_B T) + \log(R_{SRH}) \tag{49}$$

Composite Predictors (7)

$$\text{FF}_{pred} = \log(\Theta_{min}) - \log(R_s^{total}) \tag{50}$$

$$V_{oc}^{pred} = E_g + \tfrac{1}{2}(\Phi_{h,ETL} + \Phi_{e,HTL}) - k_B T \log(R_{SRH}) \tag{51}$$

$$J_{sc}^{pred} = \log(J_{max}) + \log(\eta_{coll,min}) \tag{52}$$

$$\text{Quality} = \log(\Theta_{min}) + \text{SRH}_{str} - \log(R_s^{total}) \tag{53}$$

C. Full Architecture Ablation

Table 7 reports the complete ablation spanning 10 configurations × 3 seeds = 30 runs. All values are averaged over 3 seeds with ± indicating seed-to-seed standard deviation.

Self-attention analysis. Adding 4-head self-attention ($d_{model} = 64$) consistently degrades all metrics (rows 6–7 vs. rows 2, 4; Figure 11). The Conv+Attn configuration's extreme V_{oc} instability (23.7 ± 39.9 mV, driven by a single catastrophic seed) further demonstrates that attention introduces harmful optimization landscape pathologies on these short sequences.

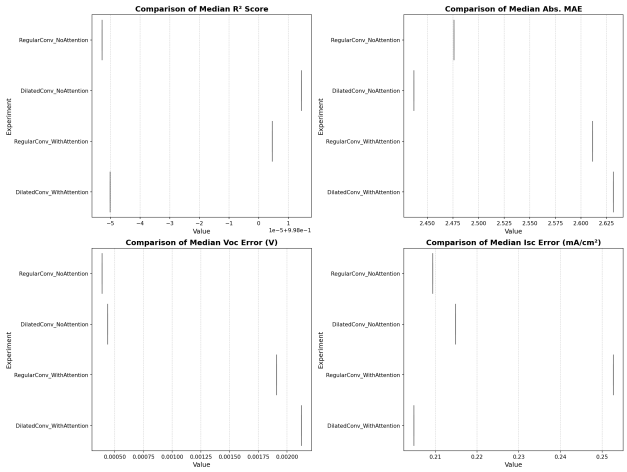


Figure 11. **Self-attention consistently degrades performance.** Paired comparisons across Conv±Attn and TCN±Attn backbones. Adding attention worsens median R^2, MAE, V_{oc} error, and I_{sc} error in every case.

Scalar conditioning. Removing upstream scalar conditioning (V_{oc}^{pred}, V_{mpp}^{pred}; row 8) barely affects median R^2 (0.9966 vs. 0.9967), confirming that the 31 raw parameters and 5 physics features already encode sufficient information for curve reconstruction.

Data scaling and training duration. Using only 100k training samples (row 9) instead of the full ∼120k degrades R^2_{med} from 0.9967 to 0.9948. Extending training to 200 epochs (row 10) provides no improvement over 100 epochs, indicating convergence within the early-stopping patience window.

Robustness. Figure 12 confirms tight clustering of key metrics across all 30 runs, with all non-attention conv variants achieving median $R^2 > 0.996$ regardless of seed.

D. Parameter Distributions

Figures 13–**??** show the raw parameter distributions grouped by physical type, confirming that the LHS design achieves near-uniform coverage. Long upper tails in Group 2 (mobilities, permittivities) motivate the $\log_{1p}$ transform applied to material properties during preprocessing (§6.3).

Table 7. Architecture ablation (30 runs = 10 configurations × 3 seeds). All bidirectional conv variants (rows 1–3) achieve $R^2_{med} \geq 0.996$; attention (rows 6–7) and pointwise architectures (row 5) are consistently worse. The champion configuration (row 1, batch 512) achieves the highest median R^2 and lowest median MAE.

	Configuration	R^2 (med.)	MAE$_{med}$ (A/m^2)	$\|V_{oc}\|_{med}$ (mV)	$\|I_{sc}\|_{med}$ (A/m^2)
Architecture variants (bidir. conv, batch 128, 100 ep.)					
1	**Conv-Dilated (BS512)**	**.9975±.0004**	**2.96±.34**	**0.39±.16**	0.36±.62
2	Conv-Dilated	.9964±.0007	3.74±.63	1.04±.30	0.0±0.0
3	Conv-NoDilation	.9967±.0004	3.60±.36	0.60±.39	0.66±.63
4	TCN-Dilated (causal)	.9965±.0007	3.68±.67	1.51±1.68	0.94±1.63
5	Pointwise (1×1)	.9890±.0004	6.48±.04	6.11±1.35	1.31±1.18
Self-attention variants					
6	Conv-Dilated + Attn	.9914±.0006	5.94±.30	23.7±39.9	1.59±.65
7	TCN-Dilated + Attn	.9917±.0005	5.89±.61	2.12±2.53	1.49±2.59
Training & data variants					
8	No scalar cond.	.9966±.0003	3.63±.34	0.86±.33	1.20±1.51
9	100k data only	.9948±.0001	4.57±.26	0.86±.58	1.02±.98
10	200 epochs	.9965±.0005	3.69±.29	1.43±1.50	0.83±.75

Figure 12. **Architecture comparison across all 30 runs.** Non-attention conv variants (blue shades) cluster tightly at median $R^2 > 0.996$, while attention variants (red) and pointwise (orange) are clearly separated.

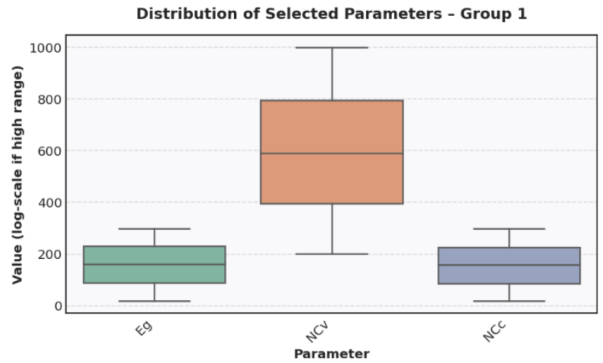


Figure 13. **Group 1**: E_g, N_{CV}, N_{CC}. Density-of-states parameters span ∼3 orders of magnitude. Near-uniform marginals confirm the LHS design.

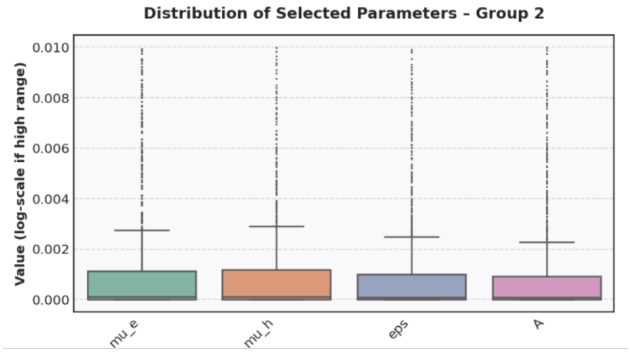


Figure 14. **Group 2**: μ_e, μ_h, ε, A. Long upper tails from log-uniform sampling, motivating the $\log_{1p}$ transform in preprocessing.

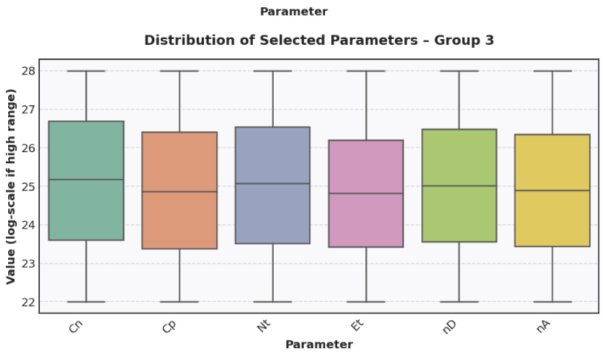


Figure 15. **Group 3**: C_n, C_p, N_t, E_t, n_D, n_A. Recombination and doping parameters.