ACS APPLIED ENERGY MATERIALS

www.acsaem.org Article

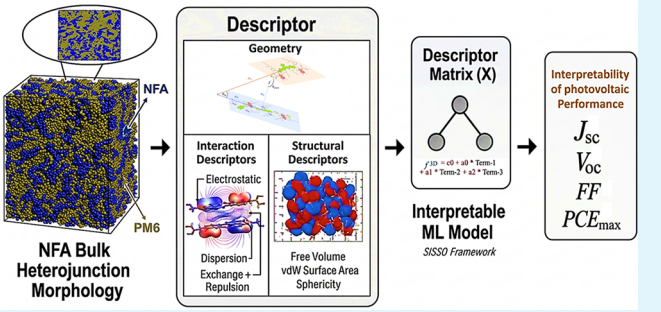
Atomistic Simulation and Symbolic Regression Reveal Design Rules Linking Blend Morphology to Photovoltaic Performance in PM6:NFA Solar Cells

Bibhas Das, Ram Sewak, and Anirban Mondal*

Cite This: https://doi.org/10.1021/acsaem.6c01289 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The performance of non-fullerene-acceptor (NFA) organic solar cells is governed by the nanoscale donor–acceptor morphology of the active layer, yet quantitative design rules linking molecular packing to device metrics remain limited. Here, we integrate fully atomistic molecular dynamics simulations with symbolic-regression-based machine learning to derive interpretable morphology–performance relationships for PM6:NFA blends. Ten representative morphologies were characterized using geometric, packing, and interaction descriptors capturing density, orientational order, interfacial spacing, aggregation, and free volume. After variance inflation factor–based descriptor screening, SISSO (sure independence screening and sparsifying operator) symbolic regression was employed to construct compact analytic models for short-circuit current (J_{SC}), open-circuit voltage (V_{OC}), fill factor (FF), and power conversion efficiency (PCE_{max}). The resulting low-dimensional expressions accurately reproduce experimental performance trends (r = 0.97–1.00) while retaining physical interpretability. Three primary morphology axes emerge: (i) a density–anisotropy–free-volume axis governing percolation and charge collection, (ii) an orientation-mismatch axis influencing interfacial charge separation and transport directionality, and (iii) a clustering/compactness axis associated with energetic disorder and voltage losses. These results translate complex atomistic morphology into actionable, physically grounded design principles for polymer:NFA solar cells and provide a generalizable framework for morphology-driven optimization of next-generation organic photovoltaic materials.



KEYWORDS: *organic solar cells (OSCs), non-fullerene acceptors (NFAs), morphology–performance relationships, symbolic regression (SISSO), molecular dynamics simulations, interpretable machine learning*

1. INTRODUCTION

Organic solar cells (OSCs), distinguished by their lightweight, flexible, and solution-processable nature, continue to attract attention as promising candidates for next-generation photovoltaic technologies.[1] A major breakthrough in their performance has come from the development of non-fullerene acceptors (NFAs), which have pushed power conversion efficiencies (PCEs) beyond 18% and toward the 19–20% range.[2,3] NFAs provide tunable energy levels, strong and broad absorption, improved morphological stability, and higher open circuit voltages, advantages that have placed them at the center of contemporary OSC research.[4,5]

Despite these advances, the nanoscale morphology of the donor–acceptor (D/A) blend remains a key bottleneck.[6,7] Photovoltaic performance originates from a complex multiscale structure in the active layer, where interfacial characteristics dictate exciton dissociation, charge separation, transport, and recombination.[8] Achieving optimal morphology requires balancing interfacial area, percolation pathways, and minimizing trap states. However, blend structure is notoriously sensitive to molecular design, intermolecular interactions, solvent and additive choice, and thermal or solvent-vapor annealing.[9,10] Too much phase separation disrupts charge transport, whereas too little increases recombination, underscoring the narrow processing window for high-performance devices.[9]

Experimental techniques such as AFM, TEM, and GIWAXS have enhanced our understanding of OSC morphology; however, they often lack the necessary resolution, throughput, or direct structure–performance correlation required for predictive design.[8,9] Complementary theoretical approaches also face limitations: density functional theory (DFT) accurately captures

Received: April 16, 2026
Revised: June 8, 2026
Accepted: June 9, 2026

electronic structure but struggles with long-range interactions, while molecular dynamics (MD) simulations provide detailed structural evolution but rely on careful force-field parameterization and produce vast, high-dimensional datasets.[11,12] Critically, neither method alone reveals clear, low-dimensional physical relationships that connect molecular chemistry, morphology, and device metrics.[13]

Machine learning (ML) has therefore emerged as a powerful complement, enabling rapid prediction of OSC characteristics using random forests, gradient-boosted trees, neural networks, and symbolic models.[13−16] These studies leverage large molecular databases and physicochemical descriptors such as HOMO/LUMO levels, band gaps, and SMILES strings, and demonstrate improved accuracy compared to conventional regression.[13,14,17−19] Nevertheless, the vast majority of ML approaches bypass the morphological complexity of OSCs, mapping simple molecular descriptors directly to PCE or related metrics without capturing the critical intermediate step: how nanoscale donor–acceptor structure governs device behavior.[14,18,19] As highlighted in recent reviews,[15,18,19] interpretability remains a key challenge, with many models functioning as black boxes that offer limited physical insight or mechanistic guidance for interface engineering. Moreover, descriptor sets are often fragmented and biased toward what is computationally easy, and they are rarely transferable across different donor–acceptor combinations.[14,17−19]

These limitations create three persistent gaps in the field, namely a lack of mechanistic, physically grounded descriptors, insufficient exploration of morphology-derived parameters, and poor generalization across chemistries and device architectures. Addressing these gaps requires frameworks that can translate the rich structural information contained in MD-simulated morphologies into compact, interpretable, and predictive models of OSC performance.

In this work, we build such a framework by integrating fully atomistic MD simulations with interpretable machine learning. We generate a comprehensive set of geometry-, interaction-, and packing-based descriptors directly from D/A interfacial morphologies, an approach that fundamentally differs from conventional models, which are dominated by quantum-chemical descriptors alone.[15,20] Leveraging the SISSO (sure independence screening and sparsifying operator)[21] methodology, we identify sparse, analytic relationships between morphology and all key device metrics [short-circuit current (J_{SC}), open-circuit voltage (V_{OC}), fill factor (FF), and power conversion efficiency (PCE_{max})]. Our results reveal that the seemingly complex morphological landscape can be reduced to three physically meaningful axes, namely density–anisotropy– spacing; orientation mismatch; and clustering, which govern photovoltaic behavior across diverse NFA chemistries. These low-dimensional descriptors not only achieve high predictive accuracy but also provide mechanistically transparent design rules, enabling inverse design of optimized morphologies.

Overall, this study bridges a long-standing gap between predictive modeling and physical understanding in OSC research. By fusing MD-derived morphological descriptors with symbolic learning, we establish a generalizable and interpretable blueprint for engineering donor–acceptor interfaces. The resulting insights offer a path toward data-driven, morphology-centric design of next-generation organic solar cells, with broad implications for scalable and high-efficiency solar energy technologies.

2. METHODS

2.1. Molecular Dynamics Simulation

To establish morphology–performance relationships across chemically diverse donor–acceptor blends, we performed fully atomistic MD simulations for ten distinct PM6:NFA combinations. Figures S1 and S2 show the molecular structures of the ten NFAs along with the PM6 dimer model used in this work. Although PM6 forms long polymer chains experimentally, a dimeric repeat-unit model was adopted because prior atomistic and coarse-grained studies have shown that dimer- and short-oligomer representations reliably capture the key short-range donor–acceptor interactions governing miscibility and interfacial packing in conjugated polymer:NFA blends.[22,23] Each NFA contains 189–220 atoms, and the PM6 dimer consists of 314 atoms.

Initial geometries of all NFAs and the PM6 dimer were constructed in GaussView[24] and optimized using DFT at the B3LYP/def2-SVP level as implemented in ORCA.[25] Bonded and nonbonded force-field parameters were generated using the Sobtop package,[26] which derives parameters from the DFT Hessian. The Sobtop-derived force field, combined with Lennard–Jones nonbonded potentials, has been previously validated to accurately capture dispersion-dominated donor–acceptor interactions and weak intermolecular forces governing OSC morphologies,[27] with results consistent with experimental observations.[28] Accordingly, the same parametrization and simulation protocol employed here are expected to reliably describe relative trends in intermolecular interactions and morphology evolution across PM6:NFA systems. The partial atomic charges were obtained using DDEC analysis[29] via the ChargeMol program.[30] Although the chosen DFT level does not explicitly include dispersion corrections, intermolecular van der Waals interactions in MD simulations were handled through Lennard–Jones pair potentials with a 13 Å cutoff. It should be noted that the role of DFT in the present workflow is limited to isolated-molecule geometry optimization, bonded-parameter generation, and partial-charge derivation, whereas intermolecular interactions governing morphology are treated entirely at the MD level. Accordingly, although the chosen DFT level does not include an explicit dispersion correction, van der Waals interactions are not omitted; they are explicitly accounted for in the MD simulations via Lennard–Jones potentials, with long-range electrostatics treated using the particle-mesh Ewald method.[31] Thus, DFT and MD serve complementary roles, with DFT providing intramolecular parameters and MD capturing intermolecular packing and dispersion effects. Explicit dispersion correction at the DFT stage would be essential only if intermolecular binding or packing were computed directly at the quantum-chemical level,[32] which is not the case here.

Blend morphologies were assembled by mixing 500 PM6 dimers with 600 NFA monomers using GROMACS.[33] The donor:acceptor ratio of 1:1.2 was chosen to match experimentally reported PM6:NFA formulations[34] and was kept fixed across all systems to isolate morphology and chemistry effects. Such a slightly acceptor-rich composition is known to promote balanced phase separation, adequate interfacial area, and efficient electron percolation in high-performing blends.[6,35] Random packing was employed to generate unbiased initial configurations. Each system was first energy-minimized (steepest descent) and then annealed in the NPT ensemble

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX−XXX

from 100 to 800 K over 1.5 ns. Temperature control was achieved using a velocity-rescaling thermostat,[36] pressure control using the Berendsen barostat,[37] and long-range electrostatics were treated with smooth PME. The annealed systems were subsequently cooled in two stages, from 800 to 600 K over 0.2 ns and from 600 to 300 K over 0.6 ns, within a 1.5 ns of NPT dynamics. Simulations employed the leapfrog Verlet integrator,[38] a 1 fs timestep, and a 1.3 nm cutoff for nonbonded interactions. Equilibration was confirmed by monitoring the stabilization of thermodynamic quantities (temperature and volume) and potential energy (See Figure S3). Structural analysis based on aromatic ring centroids indicates homogeneous amorphous packing (see Figure S4). A similar homogeneous amorphous packing was observed across all 10 systems, as well as in the external validation system. After equilibration, each system was simulated for an additional 5 ns in the NVT ensemble to produce trajectories used for computing geometric, interaction-based, and packing descriptors at the donor–acceptor interface. This workflow follows established protocols for generating amorphous OSC morphologies.[39,40] The same simulation protocol was applied to all ten PM6:NFA systems.

While longer simulations or multiple independent trajectories could further enhance sampling, the present protocol captures the relevant local and mesoscale structural features governing morphology–property relationships. We note that PM6 is represented as a dimer, a common approximation that balances computational efficiency with physical fidelity. This representation preserves the essential conjugated backbone and key interaction motifs while improving configurational sampling and avoiding kinetic trapping or artificial entanglement associated with long polymer chains.[22] Although such coarse-graining may limit the description of long-range polymer connectivity, it remains suitable for capturing intermolecular packing and donor–acceptor interactions, which are central to the present study.

In this context, DFT is employed as a computationally efficient approach for obtaining ground-state geometries and force-field parameters, with accuracy dependent on the chosen exchange–correlation functional. We note that the variational principle applies strictly to the exact functional, while approximate Kohn–Sham functionals do not guarantee energies above the exact ground state.[41] Classical MD, in turn, is limited by force-field fidelity, the absence of explicit electronic polarization and bond rearrangement, and finite time and length scales, which may restrict sampling of slow morphological relaxation processes.[42] Accordingly, MD is used here to generate equilibrated amorphous morphologies and extract local structural descriptors, rather than to fully reproduce experimental film formation. The employed annealing–quenching protocol, followed by a 5 ns production run, yields thermodynamically stable, homogeneous amorphous morphologies, as determined by centroid-level analysis of conjugated units (Figure S4). While additional validation, such as simulations with multiple trajectories, or direct comparison with experimental probes (e.g., GIWAXS) would further strengthen morphology assessment,[43] these aspects represent known limitations and directions for future work without affecting the primary objective of extracting physically meaningful morphology descriptors for structure–performance analysis.[44]

It should be clarified that the classical, atomistically resolved molecular dynamics configurations generated in this work are intended to serve as comparative, protocol-level morphology models that capture relative structural trends across the PM6:NFA series, rather than complete physical reconstructions of experimentally processed thin films. Direct validation via local characterization techniques (e.g., GIWAXS, RSoXS, TEM, or AFM) for each individual blend is currently constrained by the lack of comprehensive literature datasets. Nevertheless, the physical relevance and predictive reliability of the employed simulation and morphology analysis protocol have been explicitly established via protocol-level benchmarking against experimental thermodynamic data. Specifically, this exact workflow was previously utilized to model the miscibility and morphological stability of C-shaped (CF) and S-shaped (SF) non-fullerene acceptor isomers blended with the donor polymer D18.[27] By implementing *k*-means clustering analysis in conjunction with Voronoi tessellation, those simulations successfully quantified cluster-size distributions and local molecular packing, revealing that the C-shaped NFA forms smaller, more uniformly dispersed clusters that enhance phase mixing, whereas the S-shaped NFA aggregates into large, phase-separated domains. Importantly, these computational configurations directly reproduced the macroscopic experimental trends reported by Bai et al.,[28] where the SF:D18 blend exhibits a significantly higher experimental Flory–Huggins interaction parameter ($\chi = 2.70$) compared with the highly miscible CF:D18 blend ($\chi = 0.57$).

Because this identical, experimentally validated protocol is applied uniformly across the PM6:NFA series here, it provides a dependable basis for extracting relative structural metrics including local packing, aggregation scales, donor–acceptor contact configurations, and phase-mixing behavior. To further verify that the extracted parameters are invariant to timescale sampling limits, the production trajectory for the unseen external validation system (PM6:Y6-BO) was extended to 10 ns. Doubling the production window yielded no statistically significant deviation in the descriptors or subsequent SISSO predictions, confirming that the structural configurations are thoroughly sampled. Complete thermal, mechanical, and structural relaxation was verified by monitoring the strict convergence of the instantaneous temperature, system volume, and potential energy (Figure S3).

2.2. Computation of Morphological Descriptors

To translate the equilibrated MD morphologies into quantitative inputs for machine learning, we extracted a comprehensive set of descriptors that capture the geometric arrangement, intermolecular interactions, and packing motifs at the donor–acceptor interface. These descriptors serve as physically meaningful, low-dimensional representations of the high-dimensional atomistic configurations and were computed uniformly across all ten PM6:NFA systems to ensure methodological consistency.

2.2.1. Geometry-based Descriptors. From the equilibrated amorphous morphologies, three geometry-based descriptors, namely centroid separation, molecular plane angle, and backbone alignment angle, were computed using an identical analysis pipeline, and their definitions are schematically illustrated in Figure S5. For every saved trajectory frame, atomic coordinates were unwrapped under periodic boundary conditions using the minimum-image convention. For each donor (PM6) and acceptor (NFA) molecule, we calculated the molecular centroid, a plane normal defined from three reference atoms on the conjugated core, and a backbone direction vector obtained from two atoms aligned with the molecular long axis.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

For all donor–acceptor pairs within 10 Å, we collected (i) the centroid distance r_{DA}, (ii) the interplanar angle

$$\theta_{plane} = \arccos(\mathbf{n}_D \cdot \mathbf{n}_A) \quad (1)$$

and (iii) the backbone-alignment angle

$$\theta_{bb} = \arccos(|\mathbf{u}_D \cdot \mathbf{u}_A|) \quad (2)$$

Here $\mathbf{n}_D$ and $\mathbf{n}_A$ are the unit normal vectors to the plane of the donor and acceptor molecule, respectively. While $\mathbf{u}_D$ and $\mathbf{u}_A$ are the donor and acceptor backbone unit vectors (the absolute value folds antiparallel orientations, giving a 0–90° alignment metric). Distances were histogrammed in 0.1 Å bins (0–10 Å); θ_{plane} in 10° bins (0 – 180°); and θ_{bb} in 10° bins (0–90°). For each descriptor, we extracted the full distribution, the arithmetic mean (averaged over bin centers), and the mode (the bin with the highest occupancy).

2.2.2. Intermolecular Interaction and Packing Descriptors. To quantify the weak noncovalent interactions that shape donor–acceptor miscibility and interfacial packing, we performed an energy decomposition analysis using a classical force-field scheme (EDA-FF) implemented in Multiwfn.[45] For each PM6:NFA system, 200 donor–acceptor dimers within ~6 Å were randomly sampled from the 5 ns NVT production trajectories. For each dimer, electrostatic (ΔE_{ele}), exchange–repulsion (ΔE_{xrep}), dispersion (ΔE_{disp}), and total interaction energies (ΔE_{tot}) were computed. Together, these terms distinguish preferred binding regions, steric constraints, and π–π or other dispersive interactions that collectively determine interfacial stability.[46,47]

In addition to interaction energies, several molecular-shape descriptors were obtained from Multiwfn, including van der Waals (vdW) surface area, sphericity, and dimer density. The vdW surface area reflects the extent of molecular contact; sphericity probes shape anisotropy relevant to packing; and density was calculated from molecular mass and vdW volume.[48,49] To evaluate bulk segregation and packing heterogeneity, free-volume fractions were calculated using a grid-based occupancy scheme in which grid points falling inside smooth vdW boundary functions were assigned as occupied. The free-volume fraction for each blend was obtained from the equilibrated structure prior to descriptor extraction. Finally, molecular packing motifs were characterized by identifying centroids of terminal–terminal (TT), terminal–core (TC), and core–core (CC) units and performing *k*-means clustering on the pairwise distance matrices. Following procedures used previously for CF and SF blend analyses,[27] silhouette scoring was used to determine the optimal number of clusters, yielding fourteen distinct packing motifs consistently across all ten PM6:NFA morphologies.

2.3. Multicollinearity Analysis and Descriptor Refinement

To ensure that the structure–property models were trained on a statistically robust and interpretable feature set, we first evaluated the degree of multicollinearity among all 14 numerical descriptors. A complete list of descriptors is tabulated in Table 1. As we aimed to diagnose correlations within the descriptor space itself, the device performance metrics (J_{SC}, V_{OC}, FF, and PCE_{max}) were excluded from this analysis. For each descriptor (X_j), we regressed it against the remaining descriptors and computed the corresponding coefficient of determination (R_j^2); the variance inflation factor was then evaluated as $VIF_j = 1/(1 - R_j^2)$. This quantity reflects the factor by which the variance of the regression coefficient of (X_j) is increased due to correlations with other predictors, and therefore provides a direct measure of multicollinearity.[50,51] Large VIF values indicate that a descriptor is close to a linear combination of the others, which can inflate standard errors, widen confidence intervals, and produce unstable or non interpretable regression coefficients, even when the overall model accuracy appears satisfactory.[50,51] To mitigate these effects, we adopted a conservative cutoff of VIF = 5, consistent with widely used statistical guidelines, and implemented an iterative elimination protocol in which, at each iteration, the descriptor with the highest VIF was removed, followed by recomputation of all VIFs for the reduced set. The process continued until every remaining descriptor satisfied VIF < 5. This systematic refinement produced a compact and approximately orthogonal subset of descriptors that preserved the essential structural, energetic, and morphological information while promoting numerical stability and interpretability in the subsequent regression analyses.

2.4. SISSO-Based Modelling

To translate the refined set of morphological descriptors into transparent and predictive structure–property relationships, we employed the SISSO++ framework.[21] SISSO++ is a high-performance C++ implementation of the SISSO methodology, which integrates symbolic regression with compressed sensing to identify low-dimensional, interpretable models from high-dimensional feature spaces efficiently. This approach is particularly suited to our problem, where the goal is to uncover concise analytic expressions that link interfacial morphology to photovoltaic performance.

Beginning with the selected primary descriptors, SISSO++ systematically expands the feature space by applying a predefined set of mathematical operators (e.g., arithmetic operations, exponentiation, logarithms), thereby generating an extensive pool of nonlinear candidate features. Relevance filtering is then performed through sure-independence screening, and the final predictive expressions are obtained by solving a sparsity-promoting regression problem with l_0 regularization. This procedure yields models that strike a balance between predictive accuracy, robustness, and physical interpretability. The SISSO framework is particularly advantageous for materials property prediction as it combines predictive accuracy with interpretability by identifying compact analytical descriptors from physically meaningful variables. This is especially valuable in materials science, where uncovering structure–property relationships is as important as prediction itself. SISSO is well-suited to high-dimensional, correlated descriptor spaces and performs reliably even in small-data regimes, yielding explicit symbolic expressions that can be directly linked to underlying chemical and structural features.[52,53] However, its performance depends on the completeness of the initial descriptor pool and chosen operator space, and it may be susceptible to overfitting if model complexity is not carefully controlled. Additionally, extrapolation beyond the training domain and the lack of intrinsic uncertainty quantification remain general limitations.[52,54] In this work, these challenges are mitigated by employing a physically motivated, morphology-derived descriptor set combined with VIF-based screening, enabling the extraction of low-

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

Table 1. Morphological Descriptors Used in This Study, including Their Symbols, Definitions, and the Geometric, Interaction, or Structural Property They Represent

descriptor	symbol	definition
mean distance	r_{DA}	donor–acceptor centroid distance (Å)
plane angle	θ_{plane}	angle between donor/acceptor molecular plane normals
backbone angle	θ_{bb}	relative donor–acceptor backbone orientation angle
repulsion (exchange–repulsion)	ΔE_{xrep}	Pauli exchange–repulsion energy
inertia	I	structural inertia of donor–acceptor dimer
sphericity	Φ	shape descriptor measuring deviation from spherical geometry
density	ρ	mass per vdW volume of the dimer
free volume	V_{free}	fraction of unoccupied space in the morphology
silhouette score	S_{sil}	cluster quality metric from k-means analysis

dimensional, mechanistically interpretable relationships between nanoscale morphology and device performance.

In this study, we focused on feature spaces up to three dimensions (3D), enabling the construction of compact analytic models, denoted f^{3D}, that map the morphology descriptors to device-relevant targets (J_{SC}, V_{OC}, FF, and PCE_{max}). The resulting expressions provide direct insight into which geometric and energetic features of the amorphous interface most strongly influence photovoltaic performance, thereby bridging molecular-scale morphology with macroscopic device behavior.

2.5. Statistical Validation and Robustness of SISSO Models

Because the present study is based on a limited dataset ($n = 10$), we performed explicit statistical validation to assess model generalization, stability, and the risk of overfitting. Two complementary strategies were employed: Leave-One-Out Cross-Validation (LOOCV), which assesses prediction on unseen samples, and bootstrap resampling, which evaluates model robustness to perturbations in the training set composition.

LOOCV was carried out independently for each target property (J_{SC}, V_{OC}, FF, and PCE_{max}) using $k = 10$ deterministic folds. In each fold, one sample was excluded from model construction and used exclusively for testing, while the remaining nine samples were used for training. Thus, for fold i, the training and test sets satisfied $|\text{Train}_i| = 9$, $|\text{Test}_i| = 1$, and $\text{Train}_i \cap \text{Test}_i = \varnothing$.

For every fold, SISSO++ was rerun from the beginning using only the corresponding 9-sample training subset, including feature construction, descriptor identification, and symbolic regression under the same model complexity constraints used in the main analysis. The fold-specific model was then used to predict the held-out sample, yielding a LOOCV prediction $\hat{y}_i^{LOOCV}$. Repeating this procedure over all ten folds ensured that each sample was evaluated exactly once under strict train–test separation, providing a leakage-free estimate of out-of-sample predictive performance.

To further evaluate model stability, bootstrap resampling was performed independently for each target using $B = 30$ iterations. In each iteration, a bootstrap training set of size $n = 10$ was generated by sampling with replacement from the original dataset. Samples not selected in a given iteration formed the corresponding out-of-bag (OOB) set, with an expected OOB fraction of ~36.8%. A new SISSO++ model was then trained on each bootstrap sample using the same hyperparameter settings and descriptor dimensionality as in the main analysis.

Because symbolic regression can generate a different descriptor expression in each bootstrap iteration, the resulting models are not identical across resamples. Therefore, for each sample i, a bootstrap prediction was obtained by averaging only the predictions from those iterations in which the sample appeared in the OOB set:

$$\hat{y}_i^{\text{Bootstrap}} = \frac{1}{|B_i|}\sum_{b\in B_i}\hat{y}_{i,b}$$

where B_i denotes the set of bootstrap iterations in which sample i was out-of-bag. In total, this required $4 \times 30 = 120$ SISSO++ runs, of which 111 converged successfully (92.5%), and each device was evaluated approximately 10–11 times across OOB sets. A fixed random seed was used to ensure reproducibility of the resampling procedure.

3. RESULTS AND DISCUSSION

3.1. VIF Screening and Descriptor Curation

The initial VIF analysis revealed severe multicollinearity within the complete set of 14 descriptors: all descriptors exhibited VIF values well above 10 (mean $\approx 2.2 \times 10^5$), indicating that many could be expressed as near-linear combinations of others. Iterative removal of the descriptor with the highest VIF progressively reduced the maximum VIF from $\sim 10^6$ to 2.55 over nine iterations, yielding a statistically well conditioned core of six descriptors, namely plane angle (θ_{plane}), density (ρ), backbone angle (θ_{bb}), inertia (I), repulsion energy (ΔE_{xrep}), and sphericity (Φ), with final VIFs ranging from 1.43 to 2.55. (Figures 1 and S6). By the VIF criterion alone, these six descriptors would form the optimal low-collinearity subset.

For subsequent analyses, however, we deliberately retained a total of nine descriptors by including Silhouette score (S_{sil}), free volume (V_{free}), and mean distance (r_{DA}). Although these three exhibit moderate multicollinearity, they encode complementary and physically meaningful aspects of OSC morphology. r_{DA} represents the donor–acceptor charge-transfer distance, which correlates with the spatial separation of photogenerated electron–hole pairs. Larger charge-transfer distances have been linked to enhanced charge separation and external quantum efficiency in NFA-based OSCs.[2] S_{sil} quantifies clustering quality by comparing intra-cluster cohesion with inter-cluster separation, capturing the emergence of distinct morphological regimes in the amorphous blend. V_{free} denotes the fractional unoccupied space within the morphology; according to free-volume theory, larger and better-connected voids facilitate molecular mobility, diffusion, and charge transport, whereas low free volume reflects dense, rigid packing.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

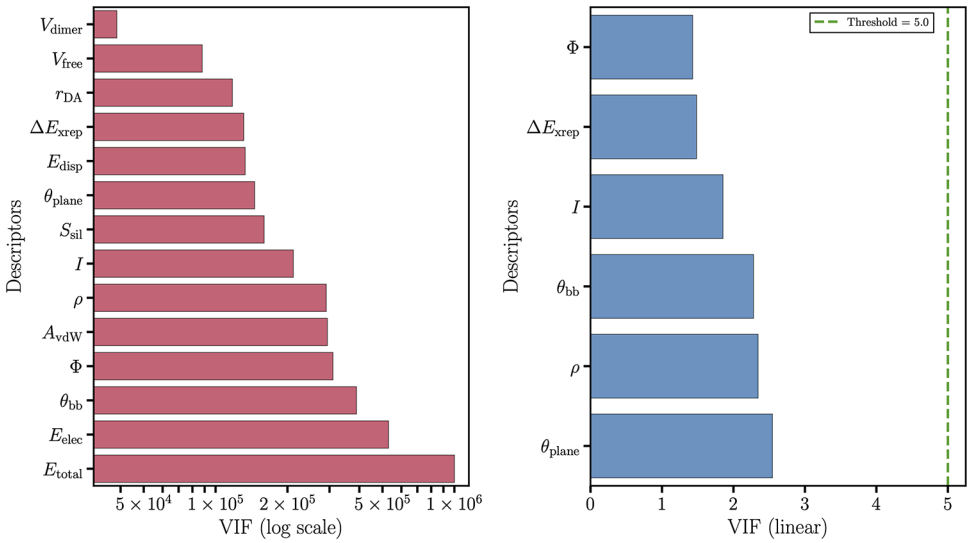


Figure 1. Variance inflation factor values for the complete set of 14 morphological descriptors (left) and after iterative pruning, highlighting the reduction of multicollinearity and the selection of six descriptors (right).

Thus, our final working set of nine descriptors strikes a balance between statistical parsimony and physical relevance. By accepting moderate VIFs for Silhouette score, free volume, and mean distance, we preserve mechanistically informative features that describe clustering, packing free space, and donor acceptor separation, which are essential factors for interpreting and predicting charge separation and transport in the studied OSC morphologies. The nine VIF-screened descriptors, along with their abbreviations and definitions, are summarized in Table 1.

3.2. Interpretable SISSO Models for Photovoltaic Parameters

Using the SISSO-derived models, we assessed the predictive power of the three-dimensional (3D) analytic expressions constructed from the morphology-dependent geometric, energetic, and structural descriptors. Table S1 summarizes the performance of these models in terms of RMSE, MaxAE, and Pearson's correlation coefficient (r). Across all four photovoltaic metrics, namely J_{SC}, V_{OC}, FF, and PCE_{max}, the 3D SISSO equations exhibit remarkably high accuracy. The J_{SC}, V_{OC}, and PCE_{max} models achieve correlation coefficients of 0.97, 0.97, and 0.99, respectively, demonstrating the strong discriminatory capability of the refined descriptor set. Even for the fill factor, which is typically more challenging to capture due to its sensitivity to coupled transport and recombination phenomena,[15,16] the 3D model achieves an r value of 0.99, demonstrating that morphology-derived descriptors alone can reproduce the governing trends with high fidelity.

Figure 2 further illustrates the strong agreement between SISSO-predicted and experimental values. The tight clustering around the diagonal across all panels confirms that the 3D models successfully encode the dominant structure–property relationships. Extending the SISSO search to higher-dimensional expressions did not yield significant improvements, indicating that the essential physics is already captured within the 3D feature space.

The effectiveness of the 3D models stems from the complementary roles of the selected descriptors, which collectively represent the key geometric arrangements and intermolecular interactions that shape the donor–acceptor morphology. Their combined influence enables SISSO to construct compact, interpretable expressions that robustly predict device performance. The identified SISSO descriptors can also be interpreted in terms of donor–acceptor charge-transfer physics rather than as purely statistical variables. In particular, r_{DA} is directly relevant because donor–acceptor separation governs wavefunction overlap, Coulomb binding, and the balance between localized charge-transfer states and charge separation.[55] The orientation descriptors θ_{plane} and θ_{bb} further modulate how conjugated cores align at the interface, thereby influencing electronic coupling and charge-transfer rates.[56] Packing-related quantities such as density (ρ) and free volume (V_{free}) regulate intermolecular contact and transport continuity,[57] while the anisotropy descriptor (Φ) reflects the degree of fibrillar organization that favors percolation and long-range charge transport.[12] The inertia term (I) provides a measure of mass distribution and local structural coherence, contributing to mesoscale connectivity relevant for charge migration.[58] In addition, ΔE_{xrep} and S_{sil} provide information on steric proximity and domain segregation, both of which affect interfacial area, exciton dissociation, and recombination.[55,59] Taken together, these descriptors provide a physically grounded representation of the donor–acceptor interfacial environment, and the SISSO expressions should therefore be viewed as compact morphology-based encodings of the coupled structural factors that govern charge generation, separation, transport, and recombination in PM6:NFA blends.

Overall, the 3D SISSO models deliver near-perfect predictions of the primary OSC parameters, underscoring the robustness of the curated descriptor set and confirming that physically grounded, morphology-based features are sufficient for accurate and computationally efficient prediction of photovoltaic behavior. It should be clarified that the present framework does not explicitly simulate excited-state charge-transfer dynamics or their deviations, but instead establishes morphology–performance relationships using experimentally measured device outputs and morphology descriptors extracted from atomistically simulated PM6:NFA blends. Each data point in the training

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

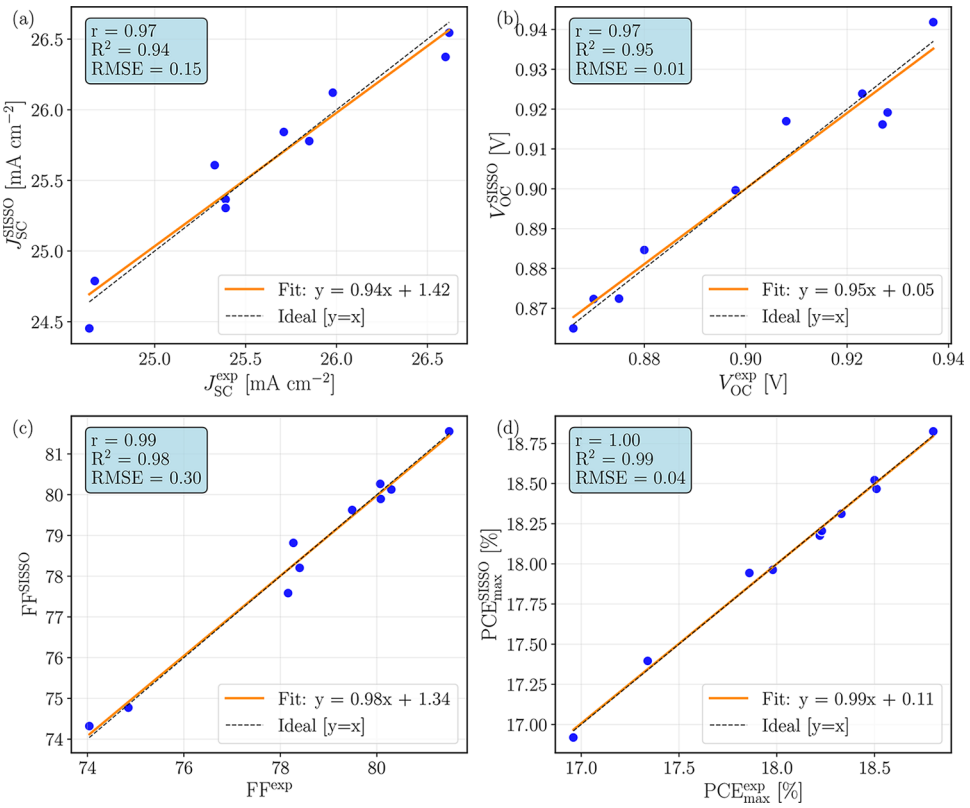


Figure 2. Parity plots comparing SISSO-predicted values from the 3D models (f^{3D}) with experimental measurements for (a) J_{SC}, (b) V_{OC}, (c) FF, and (d) PCE_{max}, demonstrating the high predictive accuracy of the SISSO-derived structure–property relationships.

set, therefore, corresponds to a distinct experimentally reported device paired with its simulated morphology, providing a direct morphology-to-experiment mapping. Within this framework, the morphological descriptors should be understood as physically meaningful proxies for interfacial charge-transfer behavior, packing, percolation, and transport, rather than as explicit descriptors of individual excited-state pathways.[7,58] Accordingly, the present results are intended as an interpretable morphology–performance model within the studied chemical space, not as a direct treatment of nonadiabatic excited-state processes.[60]

3.2.1. Leave-One-Out Cross-Validation. The resulting LOOCV metrics are summarized in Table 2, and the corresponding predicted-versus-experimental correlations are shown in Figure 3. All four targets retained strong predictive agreement under cross-validation, with Pearson correlation coefficients above 0.96. In particular, J_{SC}, V_{OC}, and FF showed very high correlations ($r > 0.98$), while PCE_{max} remained robust with $r = 0.9648$. The fitted validation lines also showed slopes close to unity and small intercepts, indicating that the models remained quantitatively well calibrated even when each prediction was generated for a sample excluded from training. The RMSE values were small relative to the scale of each target, further supporting that the learned relationships are not merely artifacts of interpolation.

3.2.2. Bootstrap Resampling Analysis. Bootstrap metrics are reported in Table 3, and the corresponding correlations are shown in Figure 4. The bootstrap analysis yielded uniformly strong agreement between predicted and experimental values for all targets. As expected, bootstrap correlations were slightly higher than LOOCV values because OOB predictions are averaged across multiple resampled models, thereby reducing variance. Thus, bootstrap here should be interpreted primarily as a measure of model stability under changes in training composition rather than as a stricter test of generalization than LOOCV.

Taken together, LOOCV and bootstrap provide complementary evidence that the SISSO models are not dominated by trivial interpolation, even with a limited sample size. LOOCV directly tests predictive performance on fully unseen samples, whereas bootstrap probes robustness to variations in the training data and reveals the stability of the learned structure–property relationships. The consistency between these two validation strategies indicates that the descriptors identified by SISSO capture meaningful and reproducible morphology–performance trends rather than unstable, sample-specific correlations. Although the small dataset necessarily limits the statistical scope of the study, the combined validation results

Table 2. LOOCV Validation Metrics for SISSO Models ($n = 10$, 10 Folds)[a]

property	Pearson r	R^2	RMSE	fit equation
J_{SC} [mA cm^{-2}]	0.997	0.994	0.050	$y = 1.017x - 0.433$
V_{OC} [V]	0.992	0.984	0.003	$y = 1.006x - 0.005$
FF [%]	0.988	0.976	0.351	$y = 0.967x + 2.633$
PCE_{max} [%]	0.965	0.919	0.151	$y = 0.960x + 0.775$

[a]Each metric reflects a prediction on samples excluded from training.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

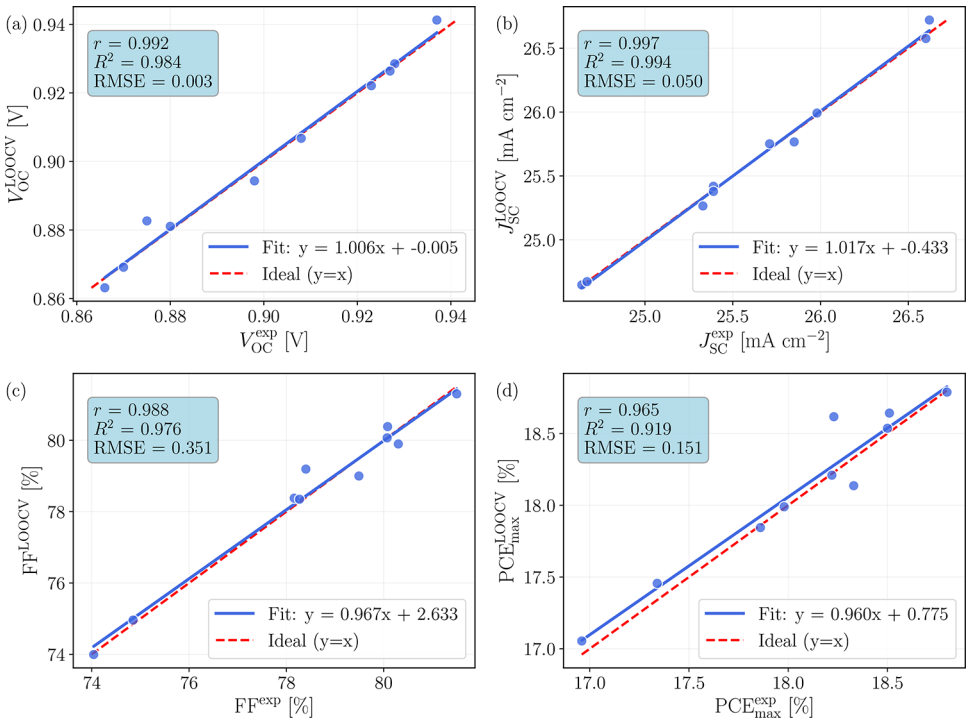


Figure 3. Correlation between experimental values and LOOCV predictions for (a) V_{OC}, (b) J_{SC}, (c) FF, and (d) PCE_{max}. Each point corresponds to a sample excluded from model training in its respective fold. Solid lines denote linear fits to the LOOCV predictions.

Table 3. Bootstrap Validation Metrics ($B = 30$ Iterations per Property) Based on Aggregated OOB Predictions

property	Pearson r	R^2	RMSE	avg OOB/device	fit equation
J_{SC} [mA cm^{-2}]	1.000	1.000	0.007	10.500	$y = 0.999x + 0.025$
V_{OC} [V]	1.000	0.999	0.001	10.600	$y = 0.980x + 0.019$
FF [%]	1.000	0.999	0.054	10.200	$y = 0.994x + 0.521$
PCE_{max} [%]	0.999	0.998	0.023	10.700	$y = 0.983x + 0.305$

support the reliability of the reported models within the chemically constrained PM6:NFA space considered here.

3.3. Descriptor-Level Mechanistic Insights into Morphological Control of J_{SC}

To elucidate how nanoscale morphology governs J_{SC}, we integrate insights from complementary statistical and machine-learning analyses, including Pearson correlation, mutual information (MI), PLS/SPCA latent-space learning, SHAP interpretability, and SISSO symbolic regression. Each method probes a different aspect of the structure–property landscape, with Pearson identifying dominant linear trends and MI isolating nonlinear predictor content; PLS/SPCA exposes the low-dimensional manifolds that organize correlated variables; SHAP quantifies the descriptor contributions within the final SISSO model; and SISSO itself yields an explicit analytic structure–property relation. Together, these techniques allow us to construct a layered interpretation of how morphology regulates charge extraction.

Pearson correlations (Figure 5) provide the initial view of descriptor sensitivities. Sphericity (Φ) exhibits the strongest linear signal with a pronounced negative correlation ($r \approx -0.56$), while r_{DA} and S_{sil} also correlate negatively with J_{SC}. These trends indicate that anisotropic, elongated, and tightly packed domains enhance J_{SC} by promoting extended percolation pathways and efficient charge separation. In contrast, density (ρ), θ_{plane}, and V_{free} correlate positively with J_{SC}, highlighting that densely packed local environments and moderate accessible voids can assist exciton dissociation and subsequent transport.

The MI analysis (Figure 6) refines this picture by quantifying nonlinear predictive content. Here, the geometric descriptor r_{DA} dominates, and the repulsion energy also exhibits a nonzero MI, with r_{DA} exhibiting MI $\sim$ 0.10. Conversely, ρ, Φ, V_{free}, S_{sil}, and I possess negligible standalone MI, suggesting that their influence arises not through isolated effects but rather via coupled, higher-order interactions that are later captured by the symbolic regression. This is consistent with the morphological complexity of organic blends, where descriptors often act synergistically rather than independently.

The latent-space models, PLS and SPCA (Table S2), reveal the structural axes that consolidate these correlations. Both analyses converge on a dominant "shape–spacing– packing" mode, featuring strong positive loadings from Φ and donor–acceptor spacing and negative loadings from ρ and V_{free}. This axis captures the primary morphological mechanism that drives J_{SC}, consistent with the Pearson and MI findings. A secondary axis couples clustering (S_{sil}), V_{free}, and inertia, reflecting the role of mesoscale connectivity and local fluctuations. Higher components encode joint orientation and energetic terms, further

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

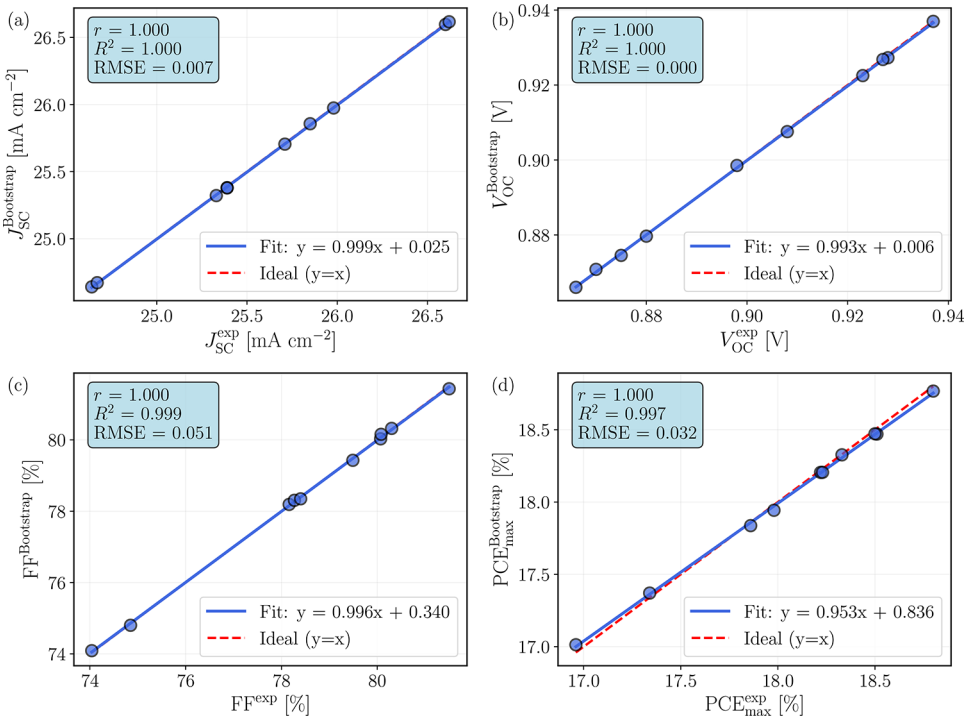


Figure 4. Correlation between experimental values and bootstrap-based predictions for (a) J_{SC}, (b) V_{OC}, (c) FF, and (d) PCE_{max}. Predictions were obtained by averaging OOB estimates across $B = 30$ bootstrap iterations.

supporting the view that J_{SC} is dictated by intertwined geometric and packing effects.

Interpreting these insights through the final SISSO model, SHAP analysis (Figure 7a) reveals a clear hierarchy of contributions. Sphericity is the overwhelmingly dominant factor: Φ produces large negative SHAP values (−130 to −220) because it appears as Φ^6 in the denominator of the principal SISSO term (eq 6), scaled by a significant positive coefficient ($a_2 = 2.14 \times 10^4$). This sixth-order dependence amplifies the effect of even small reductions in Φ, meaning that already anisotropic morphologies exert a disproportionate influence on enhancing J_{SC}. Density, in contrast, contributes modest positive SHAP values, consistent with the Pearson trends. S_{sil} contributes weakly negative values through its logarithmic dependence, while θ_{plane}, θ_{bb}, I, and V_{free} have near-zero SHAP magnitudes. Their limited contribution stems from their presence only in low-weight composite terms ($a_0 \sim 10^{-9}$, $a_1 \sim 10^{-2}$), confirming that they play secondary roles relative to the geometric descriptors dominating the SISSO manifold.

3.4. SISSO-Derived Axes Governing J_{SC}

The SISSO 3D model provides a compact, physically interpretable framework that connects nanoscale morphology to short-circuit current. In its analytic form,

$$J_{sc} = c_0 + T_0^{J_{SC}} + T_1^{J_{SC}} + T_2^{J_{SC}} \tag{3}$$

$$T_0^{J_{SC}} = a_0 \frac{I\, V_{free}}{\ln(S_{sil})} \tag{4}$$

$$T_1^{J_{SC}} = a_1 \left| I\, \Phi - \left(\theta_{plane} + \theta_{bb}\right) \right| \tag{5}$$

$$T_2^{J_{SC}} = a_2 \frac{\rho / V_{free}}{\Phi^6}, \qquad (a_0, a_1 > 0, a_2 >> 0) \tag{6}$$

the model predicts current densities with high fidelity to experiment (RMSE ∼ 0.15; Pearson $r = 0.97$). These three contributions, clustering (T_0^{JSC}), orientation mismatch (T_1^{JSC}), and the density–free-volume–shape axis (T_2^{JSC}), span the essential low-dimensional manifold governing J_{SC}. Among them, T_2^{JSC} is overwhelmingly dominant. This term couples density, free volume, and shape anisotropy in a manner that reflects the primary physical mechanism of charge transport. Dense packing (ρ) and moderate free volume promote efficient exciton separation and hopping, while strong anisotropy (low Φ) dramatically amplifies this effect through the $1/\Phi^6$ dependence. Even modest reductions in sphericity thus translate into substantial increases in J_{SC}, consistent with the expectation that elongated donor–acceptor domains establish long, continuous percolation pathways.

The second contribution, T_1^{JSC}, encodes an orientation-mismatch effect between an intrinsic mass–shape scale ($I\Phi$) and the combined molecular orientation ($\theta_{plane} + \theta_{bb}$). When these quantities differ, the resulting alignment mismatch strengthens the interfacial dipole. It promotes both exciton dissociation and directional charge transport, in agreement with earlier findings on orientation-controlled OPV performance.[56,61] Although smaller in magnitude than T_2^{JSC}, this axis provides a meaningful correction that fine-tunes transport efficiency across devices. The weakest contribution, T_0^{JSC}, acts as a clustering penalty. Because $0.58 < S_{sil} < 0.67$ ensures $\ln(S_{sil}) < 0$, an increase in S_{sil} deepens the penalty, particularly when combined with large inertia (I) and free volume. This behavior reflects the detrimental impact of excessive phase

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX−XXX

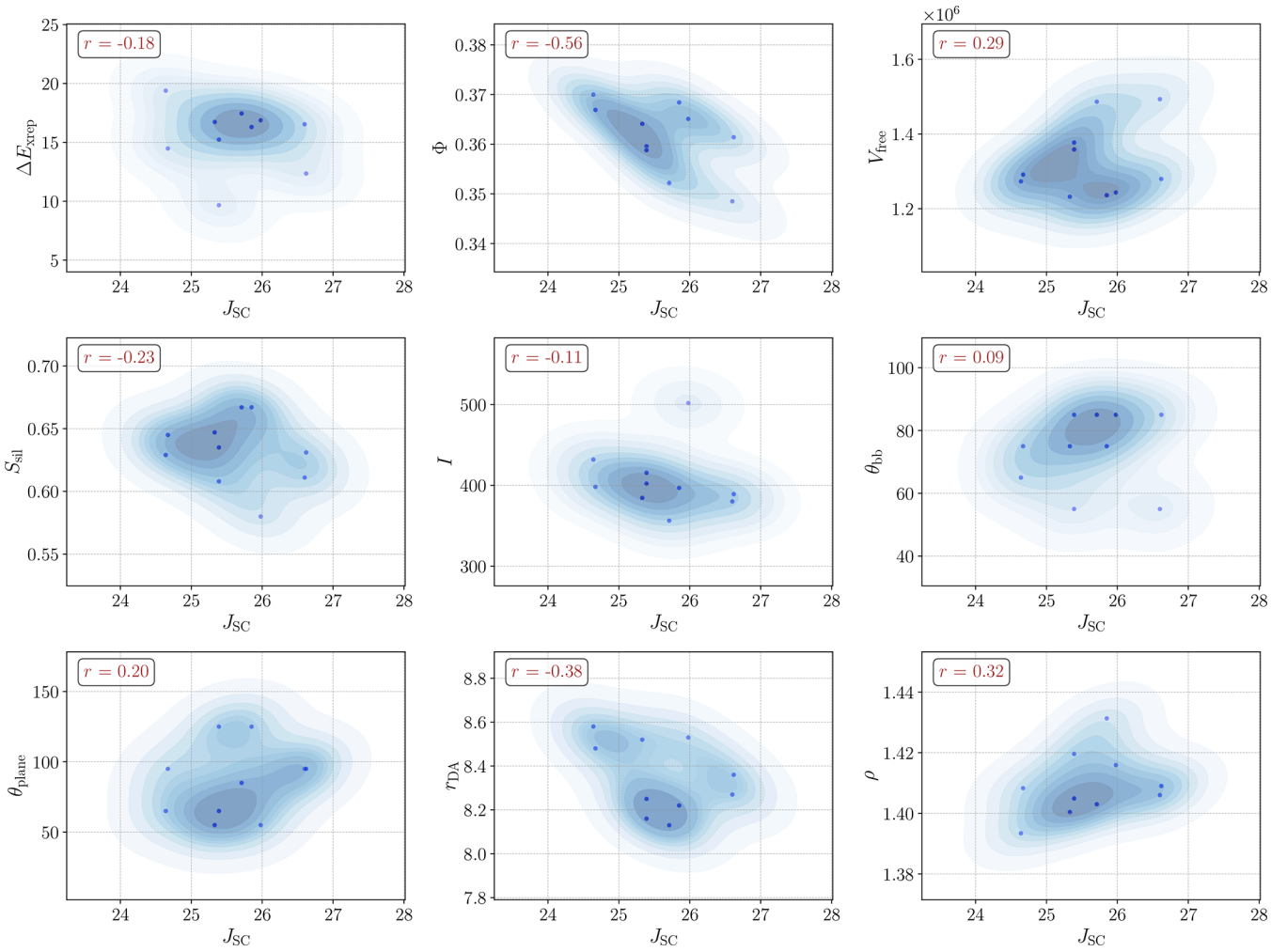


Figure 5. Pearson correlation analysis for J_{SC}: Pairwise linear correlations between morphological descriptors and J_{SC}, highlighting dominant linear trends governing short-circuit current.

separation: as domains coarsen and interfacial area decreases, exciton harvesting effectiveness is reduced.

Device-to-device variations (Table S3) map directly onto these three axes. High-performing devices (e.g., Devices 6 and 9) excel along the dominant T_2^{JSC} axis, as they possess low Φ, high ρ, and sufficient V_{free}, resulting in the largest density–anisotropy enhancement. They maintain moderate orientation mismatch (favourable T_1^{JSC}) and incur minimal clustering penalties (T_0^{JSC}). Devices 6, 8, 9, and 4 exhibit the lowest Φ and highest ρ, explaining their strong T_2^{JSC} contributions, while Devices 6 and 9 stand out with the highest J_{SC}. Orientation mismatch further boosts Devices 5, 8, 9, and 10 via elevated θ_{plane} and θ_{bb}. In contrast, Devices 2, 5, 7, and 10 suffer substantial negative T_0^{JSC} values due to high S_{sil}, whereas Devices 4, 6, and 8 incur negligible clustering penalties. As a result, each morphology occupies a distinct position in the 3D space spanned by (T_2^{JSC}, T_1^{JSC}, T_0^{JSC}), with Devices 6 and 9 positioned near the optimum along all axes.

Taken together, the statistical, latent space, and symbolic analyses converge on a consistent mechanistic picture in which J_{SC} is regulated not by isolated descriptors but by three emergent morphological axes. The primary driver is the density–anisotropy axis (T_2^{JSC}), which governs long-range percolation through dense, elongated domains. The secondary axis (T_1^{JSC}) captures orientation mismatch, refining exciton dissociation and directional hopping. The third axis (T_0^{JSC}) penalizes excessive clustering, which reduces the donor–acceptor interfacial area. Other variables, such as mean distance and repulsion energy, primarily act as correlated surrogates for these axes and therefore become redundant once the appropriate low-dimensional manifold is identified. This SISSO-derived decomposition reveals a hierarchical structure property relationship in which morphology enhances J_{SC} through dense, anisotropic packing, while orientation further modulates transport, and clustering suppresses interfacial charge generation.

Understanding how nanoscale morphology governs the remaining key photovoltaic metrics, namely open circuit voltage, fill factor (FF), and maximum power conversion efficiency, is critical for rationalizing energetic disorder, charge collection, and device efficiency in donor–NFA organic solar cells. We employed a unified analytical framework to dissect how individual descriptors and their nonlinear interactions influence device performance. Detailed statistical analyses are provided in the Supporting Information (Section S1).

3.5. SISSO-Derived Axes Governing V_{OC}

The SISSO 3D model for V_{OC} reveals that voltage is primarily governed by a density–shape compaction axis, with secondary

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

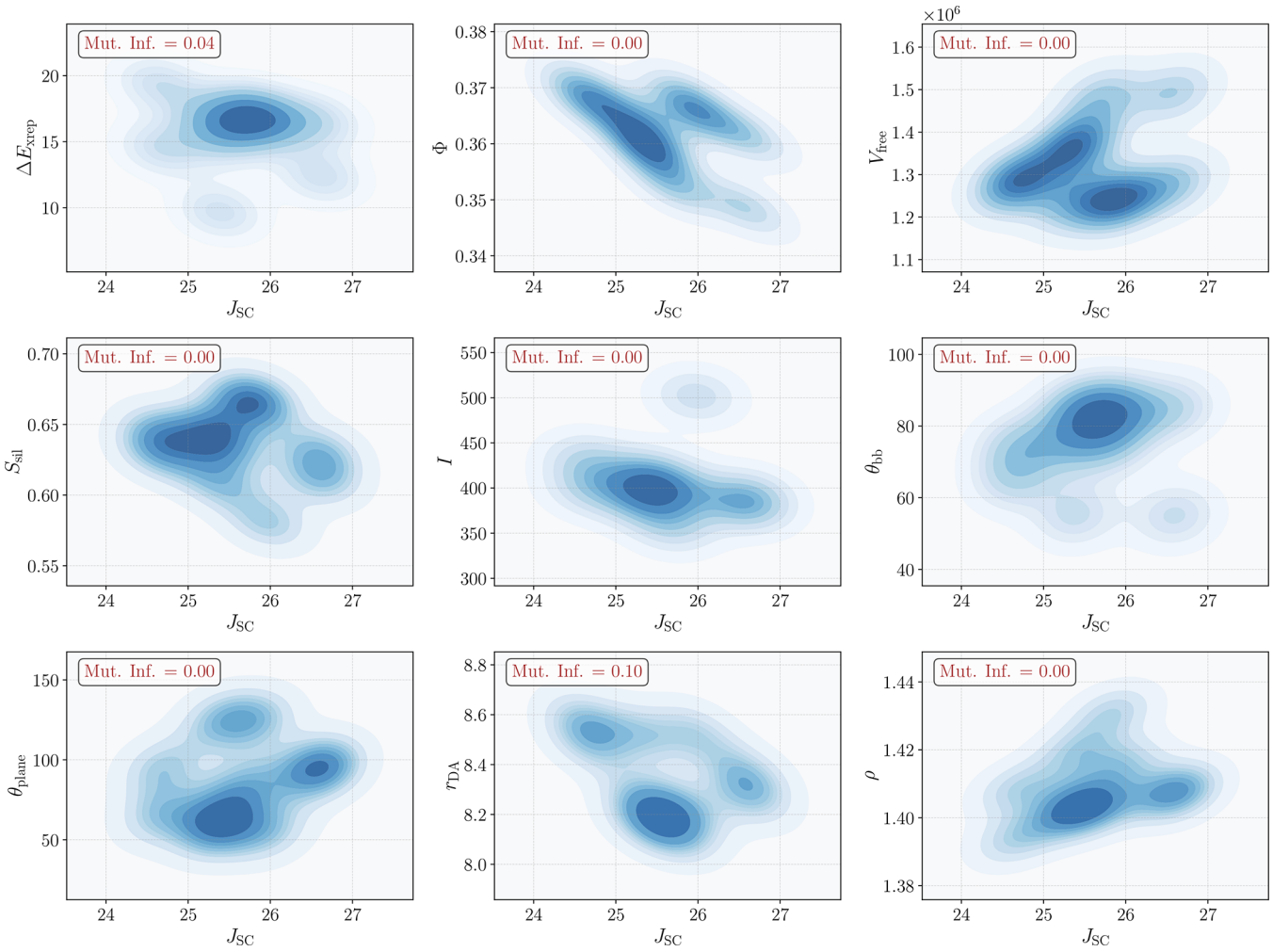


Figure 6. Mutual information analysis for J_{SC}. Nonlinear descriptor contributions to J_{SC} identified via MI, highlighting geometric features with significant nonlinear predictive content.

contributions from orientation-coherence descriptors. The explicit SISSO expression reads:

$$V_{OC} = c_0 + T_0^{V_{OC}} + T_1^{V_{OC}} + T_2^{V_{OC}} \tag{7}$$

$$T_0^{V_{OC}} = a_0 \left| I\,\Phi - \left(\theta_{plane} + \theta_{bb}\right) \right| \tag{8}$$

$$T_1^{V_{OC}} = a_1 \left| \left(\theta_{plane}/\theta_{bb}\right) - \rho \right| \tag{9}$$

$$T_2^{V_{OC}} = a_2 \frac{\rho/V_{free}}{\Phi^6}, \qquad (a_0<0,\ a_1>0,\ a_2<0) \tag{10}$$

which reproduces experimental voltages with RMSE (≈0.0058) and Pearson ($r \approx 0.974$). The dominant contribution (T_2^{VOC}) corresponds to the density–free-volume–shape axis (≈ −0.36 to −0.45). It indicates that high V_{OC} is favored in morphologies with spherical domains (high Φ), moderate density (low ρ), and low free volume. The negative coefficient (a_2) implies that increasing compaction or anisotropy enhances energetic-disorder penalties, consistent with Pearson trends. The coherence term (T_0^{VOC}) penalizes misalignment between domain-scale geometry ($I\Phi$) and molecular orientation ($\theta_{plane} + \theta_{bb}$), while the secondary asymmetry term (T_1^{VOC}) provides modest positive correction when orientation offsets density.

Device-level variations (Table S3) follow naturally from the interplay of these three terms. High-voltage devices (e.g., Device 1, $V_{OC} = 0.937$ V) combine high Φ, low ρ, and moderate V_{free}, minimizing T_2^{VOC} penalties, while intermediate and low-voltage devices show progressively more negative T_2^{VOC} and/or coherence mismatches. Collectively, V_{OC} emerges from three latent morphological axes, namely a dominant compaction axis, a secondary orientation coherence axis, and a minor orientation density asymmetry axis.

3.6. SISSO-Derived Axes Governing FF

The 3D model for FF captures the geometric trade-offs governing charge collection:

$$FF = c_0 + T_0^{FF} + T_1^{FF} + T_2^{FF} \tag{11}$$

with the composite terms (using the descriptor symbols from Table 1)

$$T_0^{FF} = a_0 \left| \frac{r_{DA}}{\theta_{plane}} - \ln(r_{DA}) \right| \tag{12}$$

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX−XXX

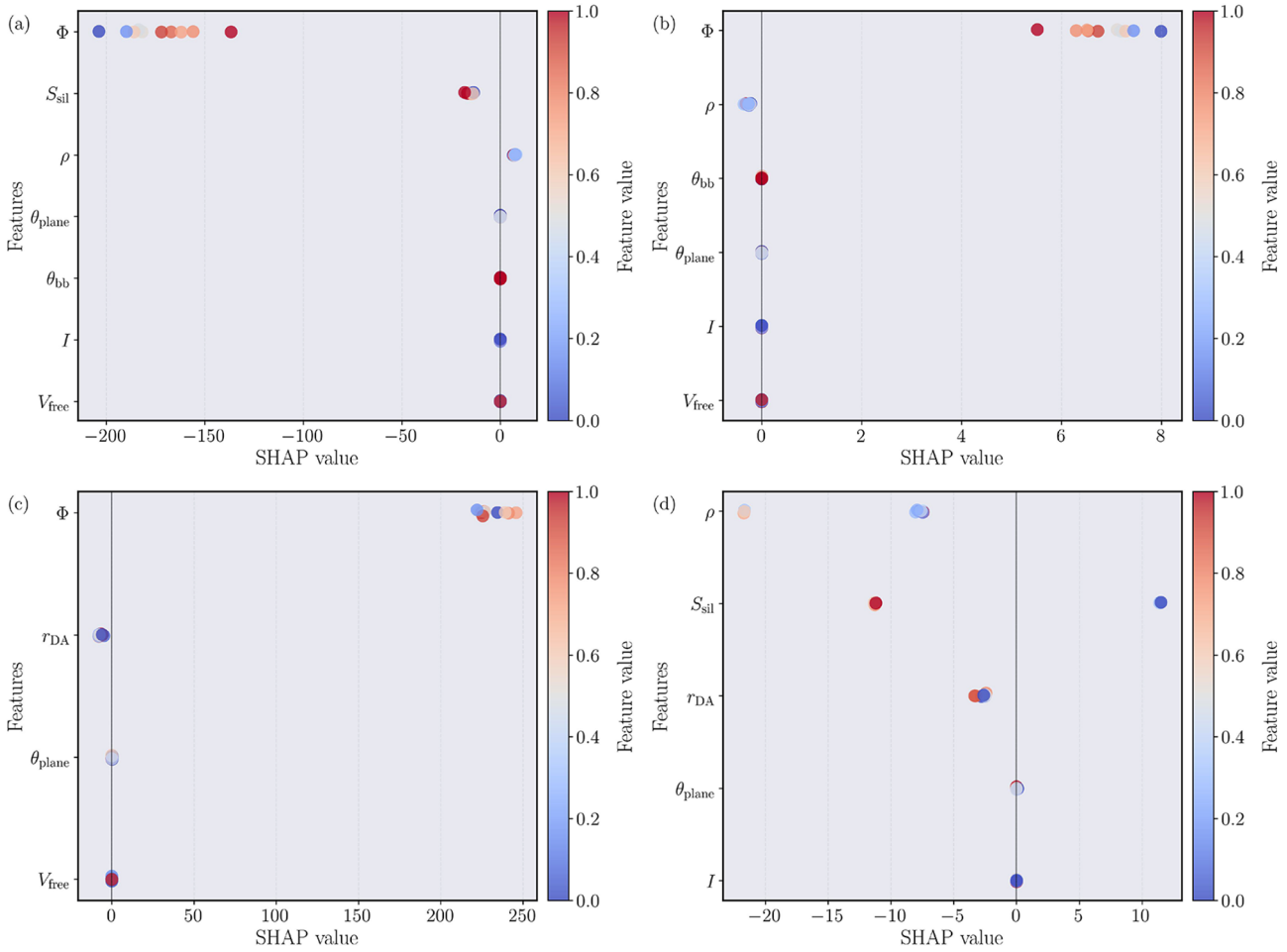


Figure 7. SHAP analysis of SISSO-derived models for all photovoltaic parameters. SHAP value distributions for (a) J_{SC}, (b) V_{OC}, (c) FF, and (d) PCE_{max}, quantifying descriptor contributions within the corresponding 3D SISSO equations.

$$T_1^{FF} = a_1 \frac{\theta_{plane}/r_{DA}}{|r_{DA} - \theta_{plane}|} \tag{13}$$

$$T_2^{FF} = a_2 \frac{r_{DA}/V_{free}}{\Phi^3} \qquad (a_0, a_1, a_2 < 0) \tag{14}$$

with RMSE (≈0.30) and Pearson ($r \approx 0.99$). All three terms act as penalties subtracted from a positive baseline (c_0). The dominant (T_2^{FF}) term encodes a geometric transport penalty, which grows with larger intermolecular spacing, smaller free volume, and low anisotropy (Φ), sharply reducing FF. Orientation mismatch (T_1^{FF}) penalizes large plane-angle deviations, while the softer T_0^{FF} enforces a spacing–orientation balance.

Device-to-device differences (Table S3) arise from the relative magnitudes of these penalties, with high-FF devices (e.g., Device 10, FF = 81.5) combine minimal Φ and r_{DA} with favorable θ_{plane} and sufficient V_{free}, whereas low-FF devices (e.g., Devices 7 and 8, FF ≈ 74−75) exhibit large r_{DA}, small plane angles, limited free volume, and high Φ, maximizing T_2^{FF} and T_1^{FF}. SHAP analysis (Figure 7c) confirms that Φ and r_{DA} dominate the importance hierarchy (eqs 11–14). Thus, FF is primarily dictated by the geometric manifold formed by spacing, orientation, and free volume, with anisotropy Φ serving as a nonlinear gating factor. Secondary descriptors such as repulsion energy, inertia, and clustering largely reflect this manifold and provide minimal additional predictive power.

3.7. SISSO-Derived Axes Governing PCE_{max}

The 3D model for PCE_{max} integrates structural, geometric, and clustering effects:

$$PCE_{max} = c_0 + T_0^{PCE_{max}} + T_1^{PCE_{max}} + T_2^{PCE_{max}} \tag{15}$$

$$T_0^{PCE_{max}} = a_0 \left| \sqrt{S_{sil}} - (\rho - S_{sil}) \right| \tag{16}$$

$$T_1^{PCE_{max}} = a_1 \frac{\ln(I)}{\rho\, r_{DA}} \tag{17}$$

$$T_2^{PCE_{max}} = a_2 \frac{\exp(r_{DA})}{\theta_{plane}^6} \qquad (a_0 < 0,\ a_1 > 0,\ a_2 << 0) \tag{18}$$

reproducing experimental PCE_{max} with RMSE ≈ 0.042 and Pearson $r \approx 0.997$. The three terms capture distinct mechanisms, with $T_0^{PCE\ max}$ is a minor clustering–packing mismatch penalty, $T_1^{PCE\ max}$ rewards mesoscale continuity (large I) under favorable density and spacing, and the dominant $T_2^{PCE\ max}$

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX−XXX

exponential term penalizes large spacing r_{DA} and poor plane-angle alignment θ_{plane}.

High-performing devices (e.g., Device 3, PCE_{max} = 18.80%) combine minimal r_{DA}, large plane angles, and moderate mesoscale continuity (Table S3), minimizing $T_2^{PCE\,max}$ while maximizing $T_1^{PCE\,max}$, with minor $T_0^{PCE\,max}$ contributions (eqs 15–18). Poor performers exhibit the opposite characteristics, with large spacing, slight tilt, or over-dense/discontinuous domains, leading to steep $T_2^{PCE\,max}$ penalties. Overall, PCE_{max} is governed by a geometric and orientational gating mechanism in which minimizing intermolecular spacing, maximizing plane angle alignment, and balancing mesoscale continuity and clustering together ensure optimal energy conversion. Correlated descriptors such as free volume, sphericity, and repulsion essentially project onto this same low-dimensional morphological manifold and add little independent predictive power once SISSO composites are considered.

3.8. Summary of Morphology–Structure Insights

While the Shockley–Queisser limit defines the thermodynamic efficiency ceiling based on band-gap and recombination physics, real device performance is critically governed by morphology, which determines how closely this limit is approached.[7,62] In organic solar cells, nanoscale packing controls charge generation, separation, transport, and recombination at donor–acceptor interfaces.[7,9] Accordingly, the morphology-derived SISSO descriptors employed here should not be viewed as replacements for electronic-structure parameters, but as physically meaningful proxies that encode how structural organization modulates the effective electronic landscape. Descriptors such as intermolecular spacing, orientation, density, free volume, and clustering directly influence electronic coupling, transport pathways, and recombination probabilities.[8] Within the chemically consistent PM6:NFA platform studied here, where intrinsic electronic variations are relatively constrained, morphology emerges as the dominant source of performance variation. In this regime, morphology-based descriptors can capture structure–property relationships with high fidelity, providing a physically interpretable framework for understanding and predicting device performance without explicitly invoking electronic parameters.[7,63]

We also note that blending can improve photovoltaic efficiency by broadening spectral coverage, as commonly explored in ternary or multicomponent systems.[64] However, the present work addresses a different question. Here, we focus on binary PM6:NFA blends, where non-fullerene acceptors already provide broad visible absorption, and where device performance is more strongly governed by nanoscale morphology than by spectral sensitization alone.[58,65] In this context, the key factors are donor–acceptor packing, orientation, free volume, and phase organization, which directly influence charge generation, transport, and recombination. The contribution of the present work lies in establishing interpretable morphology–performance relationships and design rules for binary PM6:NFA organic solar cells.

Across J_{SC}, V_{OC}, FF, and PCE_{max}, our combined statistical, latent-space, and SISSO analyses reveal a unified, physically grounded framework in which donor–acceptor morphology dictates device performance through a small set of emergent axes. For J_{SC}, dense, anisotropic domains with moderate free volume form continuous percolation pathways, enhanced by favorable orientation mismatch and minimal clustering penalties. V_{OC} is maximized in spherical, moderately dense, and low free volume morphologies, where orientation-coherence alignment mitigates energetic disorder. FF is primarily controlled by intermolecular spacing, modulated by plane angles and free volume, with anisotropy acting as a nonlinear gate. At the same time, PCE_{max} integrates these effects into a geometric–orientational gating mechanism that balances spacing, tilt, mesoscale continuity, and clustering. Notably, descriptors that appear important in isolation, such as mean distance, repulsion energy, inertia, or clustering, mainly function as proxies for these SISSO-derived composite axes.

While SISSO identifies compact analytical expressions with strong predictive performance, these expressions should not be interpreted as unique mechanistic laws. Rather, they are sparse, data-driven descriptors selected from a large candidate feature space generated using a predefined operator set.[52] Accordingly, nonlinear terms such as Φ^6, ${\theta_{plane}}^6$, and $\exp(r_{DA})$ should not be viewed as physically fundamental exponents, but as empirical transformations that improve representation of morphology–property trends within the present dataset.[66] In particular, the sixth-power terms arise naturally from repeated application of the square and cube operators in the SISSO workflow.[53] More generally, symbolic regression is non-unique, especially in the small-data regime, and multiple mathematically distinct expressions may capture similar correlations.[67] We therefore interpret the SISSO models as robust empirical descriptors rather than universal mechanistic relationships. At the same time, the recurrence of transformed descriptors such as Φ^6, $\exp(r_{DA})$, and ${\theta_{plane}}^6$ across multiple targets and descriptor dimensions indicates that they capture consistent structure–property trends within the explored morphology space. Qualitatively, the functional dependencies are physically reasonable, with distance-dependent terms consistent with the known sensitivity of charge transfer to donor–acceptor separation,[68] strong angular penalties reflecting the effect of interfacial misorientation on orbital overlap and transport,[69] and sphericity capturing geometric compactness that influences interfacial area, diffusion pathways, and recombination.[7] Therefore, the present SISSO expressions should be understood as non-unique, data-driven morphology–property descriptors whose specific exponents are not mechanistic constants, but whose repeated emergence and physically sensible trends support their interpretive value within the current chemical and structural space. Together, these insights establish a physically interpretable low-dimensional manifold linking nanoscale morphology to device physics, providing clear design rules for maximizing OSC performance and guiding future donor–NFA material selection, processing strategies, and morphology engineering.

3.9. Blind External Validation

To critically assess the out-of-sample predictive performance and transferability of the 3D SISSO models beyond the training set, we performed blind external validation on two completely unseen PM6:NFA blend systems, PM6:AC9 and PM6:Y6-BO. In both cases, the corresponding systems were entirely excluded from the initial descriptor screening, SISSO feature selection, and model parameterization workflows. The generation of amorphous morphologies, molecular dynamics simulations, and extraction of morphology descriptors were carried out using the same procedures described in Sections 2.1–2.3.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

Table 4. Blind External Validation Results for the PM6-AC9 and PM6–Y6-BO Systems Using SISSO Morphology-Based Models[a]

system	metric	J_{SC} [mA cm^{-2}]	V_{OC} [V]	FF [%]	PCE_{max} [%]
PM6-AC9	Prediction	26.34	0.929	79.83	17.45
	Uncertainty	±0.046	±0.003	±0.348	±0.155
	Experimental	26.75	0.871	79.00	18.43
	% Error	−1.54%	+6.62%	+1.06%	−5.34%
PM6–Y6-BO	Prediction	24.54	0.957	77.85	17.68
	Uncertainty	±0.046	±0.003	±0.348	±0.155
	Experimental	27.90	0.840	76.60	18.00
	% Error	−12.04%	+13.93%	+1.63%	−1.78%

[a]Uncertainties correspond to LOOCV RMSE estimates.

The resulting descriptor sets were then used as direct inputs to the trained 3D SISSO models to predict J_{SC}, V_{OC}, FF, and PCE_{max}.

As summarized in Table 4, the morphology-driven models demonstrate reliable out-of-sample accuracy. For the PM6:AC9 blend, all four photovoltaic targets are predicted with low deviations (errors $\leq$ 6.62%). For the PM6:Y6-BO blend, the model exhibits excellent quantitative calibration for the fill factor (FF error = +1.63%) and final device efficiency (PCE error = −1.78%). A larger, partially compensating divergence is observed for the individual J_{SC} (−12.04%) and V_{OC} (+13.93%) components of the PM6:Y6-BO system. This behavior reflects a well-understood physical limitation of a morphology-exclusive descriptor pool. While MD-sampled parameters like donor–acceptor distance ($\bar{r}_{DA}$) and orientation metrics (θ_{plane}, $\theta_{backbone}$) effectively dictate bulk percolation networks and interfacial packing pathways, they do not explicitly account for pure electronic or photophysical variables. Intrinsic variations in optical absorption coefficients, thin-film spectral overlap, energetic disorder, and donor–acceptor frontier orbital offsets are omitted by design. Consequently, while the dual external validation demonstrates that our framework successfully extracts robust, transferable structure–performance blueprints across the PM6:NFA family, it underscores that the integration of explicit electronic-structure parameters remains a prerequisite for universal, unconstrained multi-platform extrapolation.

4. CONCLUSIONS

We present a unified, physically interpretable framework that links nanoscale donor–acceptor morphology in polymer:NFA blends to all four key photovoltaic metrics, J_{SC}, V_{OC}, FF, and PCE_{max}, by combining atomistic MD-derived geometric, interaction, and structural descriptors with symbolic regression capabilities of SISSO++. Analysis of ten PM6:NFA systems reveals that device performance is governed by a small number of emergent low-dimensional morphological axes, namely a spacing–packing–shape axis (r_{DA}, V_{free}, ρ, Φ), an orientation–coherence axis (θ_{plane}, θ_{bb}, $I\Phi$), and a clustering–packing compatibility axis (S_{sil}). Each photovoltaic parameter corresponds to a distinct projection of this manifold, where J_{SC} is controlled by density, anisotropy, and free volume, V_{OC} by compaction and shape coherence, FF by spacing and plane orientation, and PCE_{max} by the geometric–orientational gate, modulated by structural scale and minor clustering effects.

The SISSO-derived models achieve near-perfect quantitative accuracy ($r \approx 0.97$–0.997) while remaining compact and mechanistically transparent, offering actionable insights for rational material and processing design. The present study does not explicitly include excited-state calculations, as the primary objective is to establish morphology-based principles governing charge generation and device performance. In organic solar cells, charge generation is strongly influenced by donor–acceptor separation, interfacial orientation, packing-induced delocalization, and mesoscale connectivity, rather than a simple injection mechanism.[7,70] The descriptors identified through the MD–SISSO framework capture these structural determinants, providing a physically interpretable link between morphology and photovoltaic response. While explicit excited-state calculations would offer direct insight into charge-transfer character, the ML-identified interfacial motifs presented here provide a well-defined, physically meaningful basis for future investigations.

More broadly, this work demonstrates that organic photovoltaic performance arises from a low-dimensional yet strongly nonlinear morphology–property manifold. We emphasize that the present study does not claim universal generalizability across the full organic solar cell materials space. Instead, it establishes a physically interpretable morphology–performance framework within a deliberately controlled PM6:NFA platform, where the donor, blend ratio, and simulation conditions are fixed in order to isolate morphology-driven trends. This controlled design is intentional, because simultaneous variation of donor identity, composition, and processing conditions would make it difficult to disentangle causal structure–property relationships. Indeed, several prior high-impact studies have deliberately adopted a similar strategy, examining a series of NFAs with a single donor to uncover multiscale structure–property relationships and to isolate how acceptor-side packing influences J_{SC}, V_{OC}, charge transport, and non-radiative losses.[7,9] Within this chemical space, the SISSO analysis consistently identifies three emergent morphology axes, namely (i) density, anisotropy, and free volume, (ii) orientation mismatch, and (iii) clustering and compactness, which together describe the experimental trends in J_{SC}, V_{OC}, FF, and PCE_{max}. Accordingly, the generalizability claimed here is at the descriptor and methodological levels, rather than a claim that the fitted coefficients are already transferable to all donor–acceptor systems. In this sense, the framework should be viewed as an extensible, descriptor-based strategy whose broader applicability can be tested in future studies using more chemically diverse OSC systems. The integrative approach, which bridges atomistic simulation, descriptor engineering, and interpretable symbolic regression, provides a generalizable blueprint for predictive OSC design and paves the way for inverse-design strategies that optimize molecular

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

architecture, blend composition, and processing conditions with unprecedented precision.

■ ASSOCIATED CONTENT

Data Availability Statement

The data and code to reproduce the results are available in our GitHub repository (https://github.com/Bib569/PM6-NFA-Morphology-SISSO).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsaem.6c01289.

Chemical structures, schematic plot defining computed angles, VIF analysis, SISSO-3D models, PLS-SPCA analysis, device-wise data, and detailed analysis of morphology-structure relationships (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Anirban Mondal – *Department of Chemistry, Indian Institute of Technology, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0003-3029-8840; Email: amondal@iitgn.ac.in

Authors

Bibhas Das – *Department of Chemistry, Indian Institute of Technology, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0002-9671-5275

Ram Sewak – *Department of Chemistry, Indian Institute of Technology, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0001-5700-0122

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acsaem.6c01289

Author Contributions

CRediT: **Anirban Mondal** conceived the problem; **Bibhas Das** conducted all the simulations; **Bibhas Das**, **Ram Sewak**, and **Anirban Mondal** analyzed the results and prepared the draft.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the Indian Institute of Technology Gandhinagar, India, for providing research facilities and financial support. We thank PARAM Ananta for computational resources.

■ REFERENCES

(1) Cartlidge, E. Bright outlook for solar cells. *Phys. World* **2007**, *20*, 20.

(2) Gao, J.; Yu, N.; Chen, Z.; Wei, Y.; Li, C.; Liu, T.; Gu, X.; Zhang, J.; Wei, Z.; Tang, Z.; Hao, X.; Zhang, F.; Zhang, X.; Huang, H. Over 19.2% efficiency of organic solar cells enabled by precisely tuning the charge transfer state via donor alloy strategy. *Adv. Sci.* **2022**, *9*, 2203606.

(3) Zhu, L.; Zhang, M.; Xu, J.; Li, C.; Yan, J.; Zhou, G.; Zhong, W.; Hao, T.; Song, J.; Xue, X.; Zhou, Z.; Zeng, R.; Zhu, H.; Chen, C.-C.; MacKenzie, R. C. I.; Zou, Y.; Nelson, J.; Zhang, Y.; Sun, Y.; Liu, F. Single-junction organic solar cells with over 19% efficiency enabled by a refined double-fibril network morphology. *Nat. Mater.* **2022**, *21*, 656–663.

(4) Heeger, A. J. 25th anniversary article: bulk heterojunction solar cells: understanding the mechanism of operation. *Adv. Mater.* **2014**, *26*, 10–28.

(5) Armin, A.; Li, W.; Sandberg, O. J.; Xiao, Z.; Ding, L.; Nelson, J.; Neher, D.; Vandewal, K.; Shoaee, S.; Wang, T.; Ade, H.; Heumüller, T.; Brabec, C.; Meredith, P. A history and perspective of non-fullerene electron acceptors for organic solar cells. *Adv. Energy Mater.* **2021**, *11*, 2003570.

(6) An, K.; Zhong, W.; Peng, F.; Deng, W.; Shang, Y.; Quan, H.; Qiu, H.; Wang, C.; Liu, F.; Wu, H.; Li, N.; Huang, F.; Ying, L. Mastering morphology of non-fullerene acceptors towards long-term stable organic solar cells. *Nat. Commun.* **2023**, *14*, 2688.

(7) Karki, A.; Vollbrecht, J.; Gillett, A. J.; Xiao, S. S.; Yang, Y.; Peng, Z.; Schopp, N.; Dixon, A. L.; Yoon, S.; Schrock, M.; Ade, H.; Reddy, G. N. M.; Friend, R. H.; Nguyen, T.-Q. The role of bulk and interfacial morphology in charge generation, recombination, and extraction in non-fullerene acceptor organic solar cells. *Energy Environ. Sci.* **2020**, *13*, 3679–3692.

(8) Zhu, L.; Zhang, M.; Zhong, W.; Leng, S.; Zhou, G.; Zou, Y.; Su, X.; Ding, H.; Gu, P.; Liu, F.; Zhang, Y. Progress and prospects of the morphology of non-fullerene acceptor based high-efficiency organic solar cells. *Energy Environ. Sci.* **2021**, *14*, 4341–4357.

(9) Ye, L.; Weng, K.; Xu, J.; Du, X.; Chandrabose, S.; Chen, K.; Zhou, J.; Han, G.; Tan, S.; Xie, Z.; Yi, Y.; Li, N.; Liu, F.; Hodgkiss, J. M.; Brabec, C. J.; Sun, Y. Unraveling the influence of non-fullerene acceptor molecular packing on photovoltaic performance of organic solar cells. *Nat. Commun.* **2020**, *11*, 6005.

(10) Xin, J.; Li, W.; Zhang, Y.; Liang, Q.; Song, C.; Zhao, Y.; He, Z.; Liu, J.; Ma, W. A review of nonfullerene solar cells: insight into the correlation among molecular structure, morphology, and device performance. *Battery Energy* **2023**, *2*, 20220040.

(11) Hermann, J.; Tkatchenko, A. Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals. *Phys. Rev. Lett.* **2020**, *124*, 146401.

(12) Kupgan, G.; Chen, X.-K.; Brédas, J.-L. Molecular packing of non-fullerene acceptors for organic solar cells: Distinctive local morphology in Y6 vs. ITIC derivatives. *Mater. Today Adv.* **2021**, *11*, 100154.

(13) Greenstein, B. L.; Hutchison, G. R. Organic photovoltaic efficiency predictor: data-driven models for non-fullerene acceptor organic solar cells. *J. Phys. Chem. Lett.* **2022**, *13*, 4235–4243.

(14) Suthar, R.; Abhijith, T.; Karak, S. Machine-learning-guided prediction of photovoltaic performance of non-fullerene organic solar cells using novel molecular and structural descriptors. *J. Mater. Chem. A* **2023**, *11*, 22248–22258.

(15) Das, B.; Mondal, A. Predictive Modeling and Design of Organic Solar Cells: A Data-Driven Approach for Material Innovation. *ACS Appl. Energy Mater.* **2024**, *7*, 9349–9363.

(16) Khatua, R.; Das, B.; Mondal, A. Physics-informed machine learning with data-driven equations for predicting organic solar cell performance. *ACS Appl. Mater. Interfaces* **2024**, *16*, 57467–57480.

(17) Zhao, Q.; Shan, Y.; Xiang, C.; Wang, J.; Zou, Y.; Zhang, G.; Liu, W. Predicting power conversion efficiency of binary organic solar cells based on Y6 acceptor by machine learning. *J. Energy Chem.* **2023**, *82*, 139–147.

(18) Mahmood, A.; Wang, J.-L. Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy Environ. Sci.* **2021**, *14*, 90–105.

(19) Miyake, Y.; Saeki, A. Machine learning-assisted development of organic solar cell materials: issues, analyses, and outlooks. *J. Phys. Chem. Lett.* **2021**, *12*, 12391–12401.

(20) Suthar, R.; Abhijith, T.; Sharma, P.; Karak, S. Machine learning framework for the analysis and prediction of energy loss for non-fullerene organic solar cells. *Sol. Energy* **2023**, *250*, 119–127.

(21) Scheffler, M.; Carbogno, C.; Ghiringhelli, L. M., et al. SISSO++: A C++ implementation of the sure-independence screening and sparsifying operator approach. *J. Open Source Softw.* **2022**, *7*, 3960.

(22) Sewak, R.; Khatua, R.; Mondal, A. Non-covalent interactions in conjugated polymer blends: Insights into the stability of PVC/PM6 and CPE/PM6 systems. *J. Chem. Phys.* **2024**, *161*, 214902.
(23) Gartner, T. E. III; Jayaraman, A. Modeling and simulations of polymers: a roadmap. *Macromolecules* **2019**, *52*, 755–786.
(24) Nielsen, A.; Holder, A. *Gauss View 5.0. User's Reference*; Gaussian Inc.: Pittsburgh, PA, **2008**.
(25) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
(26) Lu, T. *Sobtop. Version [dev3. 1]*, **2024**. Available online: https://sobereva.com/soft/Sobtop, (accessed: November 2024).
(27) Sewak, R.; Singh, K.; Mondal, A. Decoding Morphological Control in Isomeric Non-Fullerene Acceptor–Polymer Blends for Organic Solar Cells. *ACS Appl. Energy Mater.* **2025**, *8*, 11456–11467.
(28) Bai, Y.; Hong, L.; Dou, Y.; Zhu, S.; Tang, H.; Liu, D.; Cao, Y.; Chen, J.; Chen, S.; Shao, L., et al. C-Shape or S-Shape? The Molecular Geometry Control of Fused-Ring Nonfullerene Acceptors for Lower Energy Loss in Organic Solar Cells. *ACS Energy Lett.* **2024**, *9*, 1786–1795.
(29) Manz, T. A.; Limas, N. G. Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.* **2016**, *6*, 47771–47801.
(30) Manz, T. *Chargemol program for performing DDEC analysis, version 2.2 beta*. **2013**.
(31) Wildman, J.; Repiscak, P.; Paterson, M. J.; Galbraith, I. General force-field parametrization scheme for molecular dynamics simulations of conjugated materials in solution. *J. Chem. Theory Comput.* **2016**, *12*, 3813–3824.
(32) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.
(33) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(34) Fu, J.; Fong, P. W.; Liu, H.; Huang, C.-S.; Lu, X.; Lu, S.; Abdelsamie, M.; Kodalle, T.; Sutter-Fella, C. M.; Yang, Y.; Li, G. 19.31% binary organic solar cell and low non-radiative recombination enabled by non-monotonic intermediate state transition. *Nat. Commun.* **2023**, *14*, 1760.
(35) Wen, Z.; Wang, T.; Chen, Z.; Jiang, T.; Feng, L.; Feng, X.; Qin, C.; Hao, X. Influence of donor: acceptor ratio on charge transfer dynamics in non-fullerene organic bulk heterojunctions. *Chin. Chem. Lett.* **2021**, *32*, 529–534.
(36) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(37) Berendsen, H. J.; Postma, J.v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(38) Van Gunsteren, W. F.; Berendsen, H. J. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1988**, *1*, 173–185.
(39) Kotadiya, N. B.; Lu, H.; Mondal, A.; Ie, Y.; Andrienko, D.; Blom, P. W.; Wetzelaer, G.-J. A. Universal strategy for Ohmic hole injection into organic semiconductors with high ionization energies. *Nat. Mater.* **2018**, *17*, 329–334.
(40) Sanyam; Mondal, A. Multiscale modeling of charge transfer in hole-transporting materials: Linking molecular morphology to charge mobility. *J. Chem. Phys.* **2025**, *162*, 184702.
(41) Kryachko, E. S.; Ludena, E. V. Density functional theory: Foundations reviewed. *Phys. Rep.* **2014**, *544*, 123–239.
(42) Van Gunsteren, W. F.; Oostenbrink, C. Methods for classical-mechanical molecular simulation in chemistry: Achievements, limitations, perspectives. *J. Chem. Inf. Model.* **2024**, *64*, 6281–6304.
(43) Xia, X.; Lau, T.-K.; Guo, X.; Li, Y.; Qin, M.; Liu, K.; Chen, Z.; Zhan, X.; Xiao, Y.; Chan, P. F.; Liu, H.; Xu, L.; Cai, G.; Li, N.; Zhu, H.; Li, G.; Zhu, Y.; Zhu, T.; Zhan, X.; Wang, X.-L.; Lu, X. Uncovering the out-of-plane nanomorphology of organic photovoltaic bulk heterojunction by GTSAXS. *Nat. Commun.* **2021**, *12*, 6226.
(44) Fratini, S.; Nikolka, M.; Salleo, A.; Schweicher, G.; Sirringhaus, H. Charge transport in high-mobility conjugated polymers and molecular semiconductors. *Nat. Mater.* **2020**, *19*, 491–502.
(45) Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592.
(46) Su, P.; Li, H. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.* **2009**, *131*, 014102.
(47) Lu, T.; Chen, Q. van der Waals potential: an important complement to molecular electrostatic potential in studying intermolecular interactions. *J. Mol. Model.* **2020**, *26*, 315.
(48) Wadell, H. Volume, shape, and roundness of quartz particles. *J. Geol.* **1935**, *43*, 250–280.
(49) Kim, D.-S.; Kim, J.-K.; Won, C.-I.; Kim, C.-M.; Park, J. Y.; Bhak, J. Sphericity of a protein via the β-complex. *J. Mol. Graph. Modell.* **2010**, *28*, 636–649.
(50) Kim, J. H. Multicollinearity and misleading statistical results. *Korean J. Anesthesiol.* **2019**, *72*, 558–569.
(51) Marcoulides, K. M.; Raykov, T. Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educ. Psychol. Meas.* **2019**, *79*, 874–882.
(52) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802.
(53) Purcell, T. A.; Scheffler, M.; Ghiringhelli, L. M. Recent advances in the SISSO method and their implementation in the SISSO++ code. *J. Chem. Phys.* **2023**, *159*, 114110.
(54) Shimakawa, H.; Kumada, A.; Sato, M. Extrapolative prediction of small-data molecular property using quantum mechanics-assisted machine learning. *npj Comput. Mater.* **2024**, *10*, 11.
(55) Coropceanu, V.; Chen, X.-K.; Wang, T.; Zheng, Z.; Brédas, J.-L. Charge-transfer electronic states in organic solar cells. *Nat. Rev. Mater.* **2019**, *4*, 689–707.
(56) Dong, Y.; Nikolis, V. C.; Talnack, F.; Chin, Y.-C.; Benduhn, J.; Londi, G.; Kublitski, J.; Zheng, X.; Mannsfeld, S. C.; Spoltore, D.; Muccioli, L.; Li, J.; Blase, X.; Beljonne, D.; Kim, J.-S.; Bakulin, A. A.; D'Avino, G.; Durrant, J. R.; Vandewal, K. Orientation dependent molecular electrostatics drives efficient charge generation in homojunction organic solar cells. *Nat. Commun.* **2020**, *11*, 4617.
(57) Khan, S.-U.-Z.; Bertrandie, J.; Gui, M.; Sharma, A.; Alsufyani, W.; Gorenflot, J. F.; Laquai, F.; Baran, D.; Rand, B. P. Quantifying the effect of energetic disorder on organic solar cell energy loss. *Joule* **2022**, *6*, 2821–2834.
(58) Wang, J.; Xie, Y.; Chen, K.; Wu, H.; Hodgkiss, J. M.; Zhan, X. Physical insights into non-fullerene organic photovoltaics. *Nat. Rev. Phys.* **2024**, *6*, 365–381.
(59) Zhao, L.; Ji, H.; Li, S.; Miao, X.; Shi, Q.; Zhu, C.; Wang, W.; Zhao, S.; Huang, D.; Dong, X. Dual additive strategy to regulate the phase separation of bulk heterojunction layer for efficiency enhancement in non-fullerene organic solar cells. *Org. Electron.* **2022**, *105*, 106495.
(60) Moore, G. J.; Günther, F.; Yallum, K. M.; Causa', M.; Jungbluth, A.; Réhault, J.; Riede, M.; Ortmann, F.; Banerji, N. Direct visualization of the charge transfer state dynamics in dilute-donor organic photovoltaic blends. *Nat. Commun.* **2024**, *15*, 9851.
(61) Lee, H.; Lee, D.; Sin, D. H.; Kim, S. W.; Jeong, M. S.; Cho, K. Effect of donor–acceptor molecular orientation on charge photogeneration in organic solar cells. *NPG Asia Mater.* **2018**, *10*, 469–481.
(62) Shockley, W.; Queisser, H. *Renewable energy*; Routledge, **2018**, *2*; 35–54.
(63) Seifrid, M.; Lo, S.; Choi, D. G.; Tom, G.; Le, M. L.; Li, K.; Sankar, R.; Vuong, H.-T.; Wakidi, H.; Yi, A.; Zhu, Z.; Schopp, N.; Peng, A.; Luginbuhl, B. R.; Nguyen, T.-Q.; Aspuru-Guzik, A. Beyond molecular structure: critically assessing machine learning for designing organic photovoltaic materials and devices. *J. Mater. Chem. A* **2024**, *12*, 14540–14558.
(64) Chang, L.; Sheng, M.; Duan, L.; Uddin, A. Ternary organic solar cells based on non-fullerene acceptors: A review. *Org. Electron.* **2021**, *90*, 106063.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX

(65) Yan, C.; Barlow, S.; Wang, Z.; Yan, H.; Jen, A. K.-Y.; Marder, S. R.; Zhan, X. Non-fullerene acceptors for organic solar cells. *Nat. Rev. Mater.* **2018**, *3*, 18003.
(66) Makke, N.; Chawla, S. Interpretable scientific discovery with symbolic regression: a review. *Artif. Intell. Rev.* **2024**, *57*, 2.
(67) de França, F. O. Alleviating overfitting in transformation-interaction-rational symbolic regression with multi-objective optimization. *Genet. Program. Evol. Mach.* **2023**, *24*, 13.
(68) Kuss-Petermann, M.; Wenger, O. S. Unusual distance dependences of electron transfer rates. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18657–18664.
(69) Shavez, M.; Mahapatra, S. Influence of molecular orientation on photovoltaic performance in double donor with fullerene and non-fullerene acceptor-based heterojunctions. *Phys. Chem. Chem. Phys.* **2025**, *27*, 13403–13414.
(70) Pratik, S. M.; Kupgan, G.; Brédas, J.-L.; Coropceanu, V. Analysis of the charge generation and recombination processes in the PM6: Y6 organic solar cell. *Energy Environ. Sci.* **2025**, *18*, 841–852.

https://doi.org/10.1021/acsaem.6c01289
ACS Appl. Energy Mater. XXXX, XXX, XXX–XXX