World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

A review of photovoltaic interface engineering: Linking classical physics-based numerical simulation to data-driven optimization

T. Hu [1], W-Y. Yuan [2], Q. Xia [2, 3], J. Ye [4] and G.-H. (K.-W.) Zhu A. Chee [1, *]

[1] *College of IT Engineering, Kyungpook National University, Daegu 41566, Republic of Korea.*
[2] *Global Innovation Center, Zhejiang University, Ningbo, 315100, People's Republic of China.*
[3] *College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, People's Republic of China.*
[4] *University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China.*

World Journal of Advanced Research and Reviews, 2026, 30(03), 1599-1612

Publication history: Received on 18 April 2026; revised on 02 June 2026; accepted on 04 June 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.3.1514

Abstract

Photovoltaic efficiency and long-term device stability are heavily governed by interfacial dynamics, which impact charge carrier transport, energy band alignment, defect generation, and non-radiative recombination losses. Despite rapid breakthroughs in solar cell architectures and absorber materials, unfavorable band offsets and localized trap states remain persistent bottlenecks to maximizing power conversion efficiency. This study systematically examines the core physical mechanisms driving interface defects, Shockley–Read–Hall (SRH) recombination, band alignment engineering, and carrier extraction kinetics. Traditional interface optimization strategies, ranging from chemical and field-effect passivation to surface modifications and buffer layer engineering, are critically evaluated. Furthermore, we assess the role of physics-based numerical solvers (including SCAPS and TCAD) in mapping interfacial electric fields, electronic energy bands, and recombination behaviours. Particular attention is dedicated to the integration of data-driven interface optimization strategies, detailing how machine learning accelerates material screening, defect modeling, and molecular passivation design. The integration of machine learning with physics-based simulations is highlighted, and this work outlines a technical roadmap to bypass current challenges in data sparsity and model interpretability, charting a course toward physics-informed machine learning for next-generation self-optimizing, high-efficiency solar architectures.

Keywords: Interface physics; Surface states; Passivation; Heterojunctions; Carrier recombination

1. Introduction

With global energy demand rising along with the gradual depletion of non-renewable resources, the development and utilization of clean, efficient, and sustainable renewable energy resources have become a major trend in the current energy sector. At present, the mainstream renewable energy utilization technologies include solar photovoltaic power generation, wind power generation, hydropower generation, and nuclear power generation. Among the many options, solar energy has remains highly favored due to its abundance and wide distribution. In 2023, solar photovoltaic power generation accounted for 5.4% of total global electricity generation, making it the third-largest renewable generation framework behind only hydropower and wind power [1]. By 2025, as global energy demand accelerated, solar photovoltaics became the largest single source of growth, fulfilling more than 25% of the world's additional demand, while natural gas contributed 17% [2, 3].

In recent years, ongoing research focuses on maximizing the photoelectric conversion efficiency. Depending on the underlying absorber materials and architectures, solar cells are classified into silicon-based, thin-film, perovskite, III–V

---

* Corresponding author: G.-H. (K.-W.) Zhu A. Chee (guanghui-zhu@naver.com)

compound, and multi-junction tandem systems [4]. Silicon-based solar cells still dominate the current market, capturing an approximately 97-% market share due to their exceptional material stability and mature manufacturing technology [5]. III–V multi-junction solar cells exhibit the highest efficiency, exceeding 47% under concentrated illumination [6]. Nevertheless, converting solar energy efficiently remains bottlenecked by fundamental technological constraints. Prominent among these losses are optical reflections, carrier recombination, interface defects, and energy band mismatches [7-9]. Heterojunction interfaces are particularly critical in determining total device performance. Since heterojunction interfaces couple distinct material layers, they are highly prone to lattice mismatches, differences in thermal expansion coefficients, chemical composition incompatibilities, and surface defects. Structural defect states and dangling bonds are generated within the interface region, introducing localized energy levels inside the bandgap that increase non-radiative carrier recombination, especially SRH recombination, which in turn reduces carrier lifetime and device performance [9-11]. In addition, poorly aligned energy bands result in strong interface barriers that hinder electron and hole transport, compounding carrier recombination and degrading overall device output [12, 13].

So far, researchers have proposed various interface optimization methods to improve interface quality and reduce recombination losses caused by interface defects. Traditional interface optimization techniques primarily center on defect passivation, buffer layer insertion, and band alignment engineering [14, 15]. However, traditional approaches rely heavily on empirical trial-and-error laboratory experiments or isolated, single-variable physics-based numerical simulations. Classic defect passivation techniques often struggle with fine process control, resulting in incomplete passivation across the target surface [16]. Similarly, engineering precise energy band alignment faces stringent material compatibility limits, making the optimization of interface band alignment difficult to replicate [13]. Navigating these multi-variable parameter spaces through physical laboratory work generates an experimental burden characterized by high trial costs, heavy workloads, and protracted, complex optimization cycles. Consequently, deploying machine learning (ML) models has become a vital strategy to simplify and accelerate the experimental cycle [17].

While numerical simulation toolkits such as SCAPS, Silvaco, and Sentaurus are widely used to analyze semiconductor interfaces, they exhibit built-in limitations. Standard models frequently rely on oversimplified interface assumptions, lack highly precise defect-state parameters, require massive computational overhead, and struggle to adequately capture complex recombination dynamics. In addition, because these simulation outcomes depend heavily on baseline assumptions of ideal material behaviour, they often deviate noticeably from real-world experimental environments [18, 19]. As emerging solar cell architectures and interface material combinations grow increasingly complex, relying solely on standard numerical simulations is no longer a viable pathway for rapid design cycles. Compared to conventional modeling, data-driven ML offers significant advantages in efficiency, predictive capability, and data-driven optimization. ML frameworks excel at extracting complex, nonlinear relationships from large streams of experimental and simulation data, thereby allowing researchers to rapidly map exact mathematical relationships between raw material variables, localized interface properties, and macroscopic device performance metrics [20].

Furthermore, ML addresses coupled physical phenomena that traditional theoretical equations fail to describe completely or rapidly, such as the simultaneous interactions between interface charges, lattice strain, and ion migration [21]. When integrated with traditional numerical tools, ML can function as an accelerated surrogate model. This hybrid setup drastically reduces the computational cost of physics-based solvers, driving faster device prototyping and intelligent interface design for next-generation, high-efficiency photovoltaics [22].

2. Interface Physics

2.1. Interface Defects

Interface defects are one of the key factors affecting the performance and stability of solar cells. Due to issues such as lattice mismatch, differences in thermal expansion coefficients, and incomplete atomic bond formation between different materials, the interface region is prone to forming dangling bonds, vacancy defects, impurity defects, and deep-level trap states [23]. Defects introduce localized energy levels in semiconductors, thereby acting as carrier trapping and recombination centers.

The presence of defects introduces defect energy levels into the energy band structure. According to the positions of these defect levels relative to the valence band maximum (VBM) or conduction band minimum (CBM), they can be classified into deep-level defects and shallow-level defects [24, 25]. Shallow-level defects are located farther from the center of the bandgap and closer to the VBM or CBM. Therefore, trapped carriers are likely to be re-excited into the valence or conduction bands to participate in transport, making the influence of shallow-level traps on carrier recombination almost negligible. In contrast, deep-level defects are located closer to the center of the bandgap, making it difficult for trapped carriers to be re-excited into the conduction or valence bands. These trapped carriers are often lost through non-radiative SRH recombination, severely limiting device performance [26-28].

Interface Trap Density (D_{it}) is an important parameter affecting interface recombination and carrier transport behaviour in solar cells, and it is commonly used to describe the density of defect states existing at heterojunction interfaces. A high D_{it} introduces a large number of trap-assisted recombination centers, significantly enhancing interface recombination and thereby reducing carrier transport efficiency and overall device performance. A high density of interface defects not only decreases carrier lifetime and transport efficiency, but also leads to reductions in open-circuit voltage (V_{oc}) and fill factor (*FF*). In addition, excessive interface traps may hinder carrier extraction and accelerate device degradation, with particularly pronounced effects in heterojunction and perovskite solar cells [29, 30]. Therefore, effectively suppressing interface defects is crucial for achieving high-efficiency solar cells.

2.2. Shockley-Read-Hall (SRH) Recombination

SRH recombination is a typical trap-assisted non-radiative recombination mechanism. It was proposed by Shockley, Read, and Hall in 1952, and is one of the most important carrier recombination mechanisms in semiconductor physics and solar cell research. Unlike radiative recombination, no photons are emitted during SRH recombination; instead, the energy is dissipated in the form of heat through defect states, making it a non-radiative recombination process [9, 29, 31].

SRH recombination is a two-step trap-assisted process involving localized energy levels within the bandgap. First, an electron in the conduction band is captured by a trap state, and then a hole in the valence band is subsequently captured by the same trap. This sequential capture leads to electron–hole recombination without photon emission, and the excess energy is dissipated as heat [29, 32].

The SRH recombination rate can be expressed as:

$$R_{SRH} = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \qquad (1)$$

where n and p represent the electron and hole concentrations, respectively, and n_i denotes the intrinsic carrier concentration. τ_n and τ_p are the carrier lifetimes associated with electron and hole capture processes, respectively. n_1 and p_1 are related to the trap energy level within the bandgap [33].

The SRH recombination rate also varies among different types of solar cells. In crystalline silicon and heterojunction solar cells, SRH recombination is mainly influenced by interface trap states. In contrast, perovskite solar cells typically exhibit lower SRH recombination rates due to their relatively low defect density. Fan Zhang and co-workers reported that perovskite samples show low SRH recombination rates in both the bulk and at the surface, resulting in longer carrier lifetimes and higher PL efficiency [34].

Deep-level traps have a particularly significant impact on SRH recombination. Compared with shallow-level defects located near the conduction band or valence band edges, deep-level traps are typically situated near the middle of the bandgap and can more easily capture both electrons and holes, thereby significantly enhancing SRH recombination. Sunghyun Kim and co-workers demonstrated that in the semiconductor Cu_2ZnSnS_4, deep defect levels and large lattice distortion promote SRH recombination [35]. Therefore, reducing SRH recombination is a key objective in interface optimization for improving solar cell performance.

2.3. Band Alignment Engineering

Band alignment at heterojunction interfaces directly determines carrier transport, collection efficiency, and recombination dynamics. When two materials with different bandgaps, electron affinities, and work functions come into contact, discontinuities appear at the boundary. These discontinuities are quantified as the conduction band offset (ΔE_c or CBO) and the valence band offset (ΔE_v or VBO). These band offsets directly influence electron and hole transport as well as SRH recombination. For photovoltaic applications, appropriate band alignment is essential to ensure efficient carrier separation and extraction, as well as to suppress interface recombination. Depending on the relative alignment of these energy bands, semiconductor heterojunctions are classified into three configurations (Figure 1):

- Type I (Straddling Gap): The narrower-bandgap material is nested completely within the wider-bandgap material. This configuration confines both electrons and holes within the same spatial zone, promoting radiative recombination for light-emitting applications [36].
- Type II (Staggered Gap): The conduction and valence bands are staggered, driving electrons and holes to separate into adjacent layers. This spatial separation effectively suppresses non-radiative recombination, making it the ideal for photovoltaic carrier extraction [37].
- Type III (Broken Gap): The valence band maximum of one material sits above the conduction band minimum of the other, forming a tunneling junction useful for high-speed electronic devices [38-40].

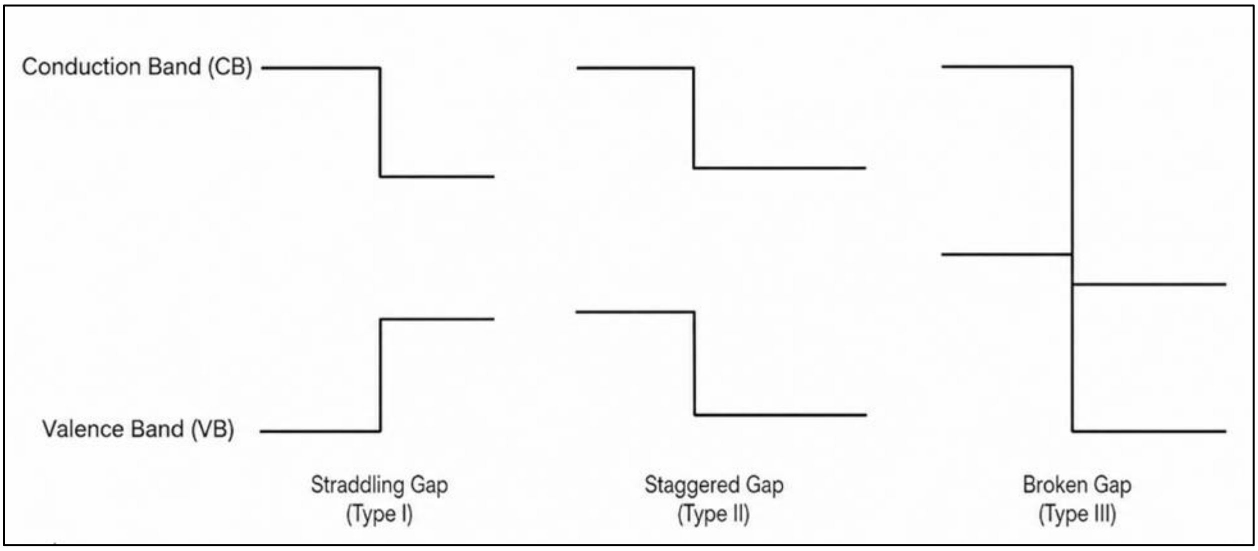

Figure 1 Classification of semiconductor heterojunctions into three types.

3. Interface optimization

3.1. Interface Passivation

To mitigate the high-density trap states and localized energetic mismatches characterized in Section 2.1, the main purpose of interface passivation is to reduce D_{it}, suppress SRH recombination, and improve carrier lifetime and device conversion efficiency [41, 42]. Interface passivation can generally be classified into two types: chemical passivation and field-effect passivation.

3.1.1. Chemical Passivation

Chemical passivation aims to saturate dangling bonds at the semiconductor interface so that they no longer participate in recombination processes, thereby reducing interface defect density and improving device performance. Hydrogen, oxygen, and dielectric materials, such as silicon dioxide (SiO_2), hydrogenated amorphous silicon nitride (SiN_x:H), and aluminium oxide (Al_2O_3), are commonly used to passivate surface states and suppress SRH recombination [43, 44]. Among various passivation materials, SiO_2 has long been regarded as one of the most effective surface passivation layers for silicon solar cells. Stefan et al. demonstrated that thermally grown or deposited SiO_2 films can significantly reduce interface defect density by saturating dangling bonds on the silicon surface, resulting in excellent chemical passivation performance [45].

However, although SiO_2 provides excellent chemical passivation, its field-effect passivation capability is relatively limited. To further suppress interface recombination, more advanced passivation materials such as Al_2O_3 have been developed. Schmidt et al. showed that ALD-grown Al_2O_3 films can significantly reduce surface recombination velocity and increase minority carrier lifetime, providing excellent surface passivation for high-efficiency crystalline silicon solar cells [46].

3.1.2. Field-Effect Passivation

Field-effect passivation relies on fixed charges in the passivation layer to form an internal electric field near the interface. This electric field repels minority carriers away from the interface, thereby suppressing interface recombination. This mechanism effectively reduces the probability of interaction between carriers and interface trap states, thus reducing SRH recombination. Field-effect passivation is particularly effective when the passivation layer contains a high density of fixed charges. For example, Al_2O_3 films are known to possess strong negative fixed charges, which can repel electrons away from the silicon surface in p-type substrates. Aber et al. demonstrated that this electrostatic repulsion significantly reduces surface recombination velocity and extends minority carrier lifetime [47].

Field-effect passivation has also been demonstrated in nanostructured photovoltaic systems. Mallorquí et al. pointed out that nanowire structures have a very high surface-to-volume ratio, making them highly sensitive to interface recombination. By introducing passivation layers that generate internal electric fields, minority carriers can be effectively separated from defect-rich regions, thereby improving carrier collection efficiency and enhancing photovoltaic performance. This highlights the crucial role of field-effect passivation in nanostructured semiconductor devices [48].

3.2. Interfacial Band Tuning Strategies

Building upon the physics of heterojunctions, practical interface optimization requires precise tuning of the CBO and VBO to maximize *Voc* and charge collection. Unfavorable energy steps generate strong transport barriers that trap charges, induce local carrier accumulation, and drastically accelerate defect-assisted SRH recombination.

In perovskite and thin-film solar architectures, optimizing these offsets relies on modifying the energy level alignment between the primary absorber and the surrounding charge transport layers. For instance, Bagade et al. demonstrated that minimizing the energetic mismatch at these interfaces directly dictates carrier extraction speed and limits voltage deficits [49]. Conversely, improper alignment causes charge stagnation. As shown by Song et al., when band offsets are excessively large, the resulting interface barriers prevent efficient carrier extraction and increase non-radiative losses [23]. Consequently, maximizing cell efficiency demands active tailoring of interfacial layers—often achieved through composition engineering, dipole insertions, or precise work-function tuning—to establish a seamless, recombination-free Type II configuration.

3.3. Buffer Layer Technology

In heterojunction solar cells and other semiconductor devices, different materials often possess mismatched lattice constants, thermal expansion coefficients, and band structures. When they are in direct contact, dislocations, cracks, and a large number of defect states are easily formed at the interface. These defects act as carrier recombination centers and promote SRH recombination, thereby reducing carrier lifetime and device performance.

To address this issue, various organic and inorganic buffer layers have been introduced to optimize interfacial energy levels and reduce defect-assisted recombination. In perovskite photovoltaics, where carrier extraction kinetics limit overall power conversion efficiency and long-term stability, researchers have employed bathocuproine (BCP) as a buffer layer to suppress interfacial carrier recombination. The introduction of the BCP layer optimizes the energy level alignment between adjacent layers and reduces carrier accumulation near the interface, thereby effectively improving the power conversion efficiency of perovskite solar cells [50]. In addition, DFT models show that inserting an appropriate buffer layer can modify charge redistribution and interfacial dipoles, thereby improving carrier transport and reducing interface recombination [51].

In organic solar cells, the insertion of suitable buffer layers can effectively optimize energy level alignment between the active layer and electrodes, thereby reducing carrier transport barriers and improving charge collection. Buffer layers prevent carrier accumulation near the interface, enhance carrier selectivity, and suppress interface recombination. Studies have also shown that buffer layers can improve interfacial stability and reduce energy losses caused by poor interfacial contact [52].

3.4. Surface Treatment Technology

During the fabrication of solar cells, material surfaces are often affected by air oxidation, chemical residues, and plasma bombardment, leading to the formation of a large number of dangling bonds, impurities, and defect centers. These defects introduce interface trap states within the bandgap, thereby promoting SRH recombination, reducing carrier lifetime, and degrading device performance. While passivation and buffer layers introduce secondary materials, surface treatment technologies modify the existing absorber termination to remove surface contaminants, native oxide layers, or process-induced plasma damage. Effectively deploying these techniques during device fabrication is of great importance for reducing the active interface trap density and improving photoelectric conversion efficiency [53].

3.4.1. Wet Chemical Treatment

Wet chemical treatment is one of the most common approaches, primarily using acidic or alkaline solutions to remove surface oxides and organic contaminants. Wet chemical cleaning is one of the most widely applied surface treatment techniques. Traditional cleaning processes include RCA cleaning, HF treatment, and ozone-based wet cleaning, which are typically used to remove surface contaminants, metal impurities, and native oxides from semiconductors. Proper wet chemical treatment can significantly improve interface passivation quality, suppress interface defect states, and reduce SRH recombination at the interface [53, 54].

3.4.2. Plasma Treatment

Plasma treatment technology can effectively remove surface contaminants, modify surface chemical properties, and tailor interface morphology without significantly damaging the material [14]. Ma and co-workers investigated plasma treatment of TiO_2 electrodes [55]. Their results showed that plasma treatment significantly improved the hydrophilicity of the TiO_2 surface, thereby enhancing perovskite crystal growth and improving interfacial contact. In addition, plasma-treated TiO_2 exhibited reduced surface defect density and improved carrier extraction efficiency. Besides air plasma treatment, argon plasma treatment can also effectively improve interface quality and enhance long-term device stability [56].

3.5. Interface Optimization Using Numerical Simulation

Numerical simulation is a commonly used method for interface optimization in solar cells. By constructing physical device models, it is possible to analyze the influence of interface parameters on device performance without extensive experimental fabrication. Widely used simulation tools include Silvaco TCAD, Sentaurus, SCAPS-1D, AFORS-HET, and wxAMPS (Table 1). These software packages model carrier transport, recombination kinetics, electric field distributions, and energy band structures by solving the semiconductor device equations self-consistently.

Table 1 Comparison of Common Simulation Softwares.

Software	Main uses	Research Direction
Silvaco TCAD	Semiconductor devices, solar cell structures, electrical and optical simulation	HJT, TOPCon, Perovskite, III-V
Synopsys Sentaurus	High-precision semiconductor device simulation	Industrial-grade research
SCAPS-1D	Thin-film solar cells	Perovskite, CIGS, CdTe
AFORS-HET	Heterojunction solar cells	a-Si:H/c-Si HJT
wxAMPS	Thin-film solar cells	CIGS, CdTe, CZTS
GPVDM	OLED/Organic/Perovskite	Organic, perovskite

In interface engineering, these simulations systematically isolate the effects of interface defect density, recombination velocity, band offsets, interfacial layer thickness, doping concentration, and work functions. For example, by varying the interface trap state density, the impact of SRH recombination on V_{OC}, short-circuit current density (J_{SC}), and FF can be analyzed. By introducing an ultrathin passivation layer or buffer layer, its effect on suppressing interface recombination and improving carrier extraction efficiency can also be evaluated.

The primary advantage of numerical modeling lies in its ability to rapidly screen key parameters and visualize physical mechanisms that are difficult to directly observe in experiments, such as simulated band diagrams, electric field distributions, recombination rate profiles, and carrier concentrations. These metrics illustrate the extent of any excessive interface barriers, severe defect-assisted recombination, or carrier accumulation at the interface. Therefore, numerical simulation can not only be used to interpret experimental results but also to predict optimal interface structures prior to experiments, thereby reducing trial-and-error costs.

However, the accuracy of numerical simulations depends on the reliability of the input material parameters and physical models. If parameters are inaccurate or the physical model is incomplete, the simulation results may deviate from real experimental conditions. Therefore, in practical research, numerical simulations are often combined with experimental characterization results, using experimental data to calibrate and refine model parameters.

4. Machine Learning-Guided Interface Optimization

4.1. Machine Learning Core Concepts

As solar cell architectures evolve into complex, high-dimensional multi-junction and heterojunction systems, evaluating their heavily coupled interfacial properties requires a departure from traditional, low-throughput pipelines. Traditional interface optimization approaches mainly rely on experimental trial-and-error or numerical simulations. These methods require repeated adjustment of parameters such as material composition, doping concentration, buffer layer thickness, passivation layer thickness, and band offsets, followed by device characterization to evaluate optimization outcomes. Based on the results, the next optimization step is then determined. Although this conventional approach is intuitive, it suffers from long experimental cycles and high costs, and it is difficult to simultaneously consider multiple interdependent interface parameters.

Compared with purely physics-based optimization, ML is increasingly regarded as a practical tool for accelerating interface design, particularly in complex multi-parameter systems. Its core idea is to use existing experimental data, simulation data, or literature data to establish a mapping relationship between interface structure, material properties, processing parameters, and device performance, thereby predicting the influence of different interface designs on performance. Compared with traditional methods, ML does not rely entirely on human experience; instead, it can automatically extract patterns from large datasets and identify key interfacial factors affecting device performance. For example, parameters such as D_{it}, CBO, VBO, interfacial layer thickness, passivation molecular structure, energy levels of charge transport layers, and defect formation energies can be used as input features for ML model compilation, mapped against output target metrics such as photovoltaic conversion efficiency, V_{oc}, J_{SC}, FF, and operational lifetime.

In recent years, ML has been applied to interface passivation material screening, buried interface engineering, defect suppression, band alignment optimization, and device performance prediction. For instance, some studies have used ML to analyze the relationship between ammonium salt molecular features and efficiency enhancement in perovskite solar cells, thereby assisting in the screening of interface passivation materials. Other studies have used ML to design buried interface engineering materials to improve charge transport, reduce recombination, and enhance device stability.

4.2. Machine Learning Models

In ML, different models are suited to different data types and task objectives. According to the training strategy, ML is generally divided into supervised learning, unsupervised learning, and reinforcement learning. Among them, supervised learning and unsupervised learning are most widely applied in solar cell interface optimization, defect prediction, and materials screening.

Supervised learning is currently the most widely used ML approach in solar cell research. Its core idea is to train models using labeled data to establish relationships between input parameters and output results. In photovoltaic interface optimization studies, supervised learning is commonly used for predicting power conversion efficiency, analyzing defect density, studying interface recombination behaviour, and screening materials. Common supervised learning models include Linear Regression (LIR), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN).

LIR is one of the most fundamental supervised learning models and is mainly used to analyze linear relationships between different parameters. In solar cell research, it is often used to study the effects of doping concentration, defect state density, and interface layer thickness on device performance. Due to its good interpretability and low computational complexity, it is commonly used for preliminary data analysis and trend prediction. However, since interfacial phenomena in solar cells are typically highly nonlinear, the predictive capability of LIR is limited in complex problems.

SVM is a supervised learning algorithm with strong generalization ability, suitable for both classification and regression tasks. SVM constructs an optimal hyperplane to effectively separate high-dimensional data, making it particularly suitable for defect classification and performance prediction under small-sample conditions. In solar cell interface engineering, SVM has been widely used to analyze defect recombination behaviour and material performance variations under different interface conditions.

RF is an ensemble learning model based on multiple decision trees. It improves prediction accuracy and reduces overfitting by aggregating the results of multiple trees. Due to its strong capability in handling nonlinear problems and its ability to evaluate feature importance, RF has become one of the most widely used ML models in solar cell research. It is commonly applied in interface defect prediction, key parameter screening, and device performance optimization.

ANNs are models that mimic the structure and function of biological neural systems and can learn complex nonlinear relationships from large datasets. In solar cell research, ANNs have been widely used for efficiency prediction, defect state analysis, and interface engineering optimization. Compared with traditional statistical methods, ANNs generally achieve higher predictive accuracy when dealing with multi-parameter coupled interfacial problems.

With the development of deep learning, Convolutional Neural Networks (CNNs) have also been increasingly applied in solar cell research. CNNs are specialized deep learning models for image recognition and feature extraction, capable of automatically extracting key features from images. In photovoltaic research, CNNs are mainly used for the automated analysis of scanning electron microscopy (SEM), transmission electron microscopy (TEM), PL images to identify grain boundary defects, interface inhomogeneities, and microstructural variations. Compared with traditional manual analysis, CNNs significantly improve defect detection efficiency and reduce human error.

In addition to supervised learning, unsupervised learning is also an important component of ML. Unlike supervised learning, unsupervised learning does not require labeled data; instead, it identifies intrinsic patterns within data for clustering, feature extraction, and pattern recognition. In solar cell interface optimization research, unsupervised learning is mainly used for high-dimensional experimental data analysis, material classification, and defect pattern recognition. Among them, Principal Component Analysis (PCA) is a commonly used dimensionality reduction technique that can extract key parameters influencing device performance and reduce data complexity, thereby improving the efficiency of subsequent model training. In addition, clustering methods are also used to study similarities among different material systems or interface states.

Overall, both supervised and unsupervised learning play important roles in solar cell interface optimization research. Supervised learning enables accurate prediction of device performance and defect behaviour based on existing experimental data, while unsupervised learning helps uncover hidden patterns from complex datasets. With the continuous expansion of experimental databases and the advancement of computational capabilities, ML is expected to further drive solar cell interface engineering toward intelligent and high-throughput optimization.

4.3. Data Collection and Model Building

The first step in ML-guided interface optimization is the construction of a reliable dataset. The data required for interface optimization mainly comes from three sources: experimental data, numerical simulation data, and literature databases. Experimental data typically include device structures, material compositions, fabrication processes, interface treatment methods, and characterization results such as *J*–*V* measurements, external quantum efficiency (EQE) measurements, PL/time-resolved photoluminescence (TRPL) measurements, and electrochemical impedance spectroscopy (EIS) measurements. Numerical simulation data, including energy band distribution, electric field strength, carrier concentrations, recombination rates, interface state density variations, and device output parameters, can be directly obtained using TCAD solvers. Literature data can be extracted from published studies to establish correlations between different interface engineering strategies and device performance.

Researchers typically use interface defect density, trap energy levels, interface recombination velocity, material band parameters, and device structural parameters as input features, while using power conversion efficiency, V_{oc}, J_{SC}, and *FF* as output targets. After training models with large amounts of experimental or simulated data, ML can rapidly predict performance trends of devices under different interface conditions, thereby guiding the design of optimization strategies.

4.4. Trap State Prediction

Traditional defect analysis mainly relies on characterization techniques such as TRPL, EIS, deep-level transient spectroscopy (DLTS), and space-charge-limited current (SCLC) measurement. These methods are widely used to investigate defect states, carrier recombination behaviour, charge transport, and trap density in photovoltaic materials and interfaces. However, because interface defects involve multiple coupled parameters, it is often difficult to rapidly and accurately determine the specific influence of different defect types on device performance using experiments alone or single-model simulations. In recent years, ML has been introduced into interface defect research, enabling the rapid prediction and optimization of defect effects by establishing a mapping relationship between “defect parameters—recombination behaviour—device performance.”

Seongtak Kim and co-workers constructed a database of 3,611 p–i–n structured perovskite solar cells with different defect conditions based on SCAPS-1D and used multiple ML models to analyze bulk and interfacial defects and their effects on device performance [57]. In their study, four algorithms—Random Forest Regression (RFR), Extreme Gradient Boosting (XGB), LIR, and Multilayer Perceptron (MLP)—were trained and used for prediction. The predicted results were then compared with SCAPS simulation outputs, and the RF model showed the highest prediction accuracy with very low loss. This demonstrates that ML-based data can be used as a reliable reference for defect state prediction and optimization design.

In addition, the study found that when bulk defect density in the perovskite absorber is low, device performance becomes more sensitive to interface defects. This implies that as the quality of bulk materials continues to improve, future efficiency improvements will increasingly rely on interface engineering. Through defect optimization strategies, the device efficiency of a PTAA/PVSK/PCBM structure was ultimately improved from 17.97% to 24.66%.

Beyond SCAPS-based ML analysis, Li et al. proposed a defect analysis method combining modulated transient photovoltage (m-TPV) with neural networks [58]. In this work, a large m-TPV curve database was established through high-throughput electrical transient simulations, and a backpropagation neural network was used to build a correlation model between m-TPV signals and defect parameters. The results showed that, compared with defect density alone, the carrier capture cross-section of defects has a more significant impact on recombination behaviour. This indicates that in interface defect analysis, it is not sufficient to consider only defect concentration; the defect energy level and carrier capture probability must also be taken into account.

Overall, the advantages of ML in interface defect and trap state prediction can be summarized in three aspects: first, it can efficiently handle complex multi-parameter coupling problems; second, it can identify key interfacial factors through interpretable models; and third, it can be integrated with SCAPS, TCAD, and transient electrical analysis for high-throughput defect prediction. Therefore, ML is gradually becoming an important tool for photovoltaic interface defect analysis and defect engineering optimization.

4.5. Passivation Material Selection

Interface passivation is a key approach for reducing interface defect density, suppressing non-radiative recombination, and improving the stability of solar cells. By introducing appropriate passivation materials, dangling bonds can be effectively saturated, vacancy defects can be compensated, trap state density can be reduced, and band alignment can be optimized. Common interface passivation materials include metal ions, organic molecules, polymers, and two-dimensional materials. However, due to the vast number of potential passivation materials, traditional experimental trial-and-error screening is often time-consuming and costly.

ML provides a new data-driven strategy for passivation material screening. Its core idea is to use existing experimental data, first-principles calculations, and materials databases to establish a relationship between “material structure—interfacial interaction—device performance,” thereby enabling rapid prediction of passivation effectiveness.

At present, ML models such as RF, XGB, SVM, and deep learning architectures have been applied to passivation material screening. By training on large amounts of experimental data, ML can automatically identify which functional groups or molecular structures are more favorable for defect passivation. For example, studies have shown that materials containing amine groups, carbonyl groups, and halogen functional groups often exhibit strong interface passivation capability, as these groups can form stable interactions with undercoordinated Pb^{2+} ions or other defect sites.

ML can not only screen existing materials but also assist in the design of novel passivation molecules. Through generative models or high-throughput prediction approaches, the most promising passivation candidates can be rapidly identified from a large chemical space, significantly reducing experimental workload. In addition, ML can be combined with first-principles density functional theory calculations (DFT) to further investigate passivation mechanisms, such as binding energies between molecules and defect sites, charge transfer behaviour, and energy level shifts.

Beyond organic molecular passivation, ML frameworks optimize multi-element co-doping and interfacial compositions. By mapping the high-dimensional relationship between doping concentrations and device performance, ML models can quickly isolate optimal doping boundaries while predicting the long-term stability and defect suppression dynamics of co-passivated architectures. This approach has demonstrated high potential across perovskite, $Cu_2ZnSn(S,Se)_4$ (CZTSSe), and silicon heterojunction (SHJ) solar cells.

Compared to traditional empirical trial-and-error methods, ML-assisted passivation material screening accelerates throughput by efficiently navigating multi-parameter design spaces. This data-driven paradigm integrates disparate experimental and high-fidelity simulation datasets, enabling the extraction of non-linear interfacial patterns that elude conventional empirical observation.

4.6. Simulation and Machine Learning Integration

With the increasing complexity of next-generation photovoltaic architectures —such as perovskite, SHJ, multi-junction tandem, and advanced thin-film solar cells— traditional interface optimization methods that rely on experimental trial-and-error are no longer sufficient to meet the demands of high-efficiency device design. The interface region in these devices involves multiple coupled parameters, including interface defect density, band offsets, localized interfacial electric fields, doping concentrations, interfacial layer thicknesses, and carrier mobilities. Since these variables are strongly interdependent, rapidly identifying an optimal interface structure is remarkably difficult using physical experiments alone or conventional, single-parameter simulation scanning.

While traditional numerical simulation methods can isolate these behaviours, they present clear computational limitations when handling multi-parameter optimization. As the parameter space expands, the number of required sequential simulations scales exponentially. This leads to high computational costs, long simulation cycles, and low optimization efficiency, especially when attempting to resolve the strong coupling effects inherent to complex heterojunctions and multi-junction devices.

To bypass these limitations and accelerate device design, a combined "numerical simulation + ML" approach has emerged as a powerful paradigm in interface engineering. The core strategy leverages the rigorous semiconductor physics models described in **Section 3.5** (such as SCAPS, TCAD, or first-principles calculations) to generate targeted baseline datasets rather than executing thousands of exhaustive, individual transport runs. By systematically varying parameters like interface trap density, recombination velocity, CBO, and trap energy levels, researchers can quickly generate comprehensive simulation data to train machine learning models.

Once trained, these ML architectures establish a highly efficient mapping relationship between "device parameters, interface behaviours, and final device performance." By serving as rapid, low-overhead proxies for computationally expensive numerical solvers, these data-driven models bypass traditional simulation bottlenecks. This integration enables the instantaneous prediction of multi-parameter shifts and allows researchers to rapidly sweep high-dimensional spaces to identify global structural optima.

5. Advantages and Challenges

ML-assisted interface optimization accelerates the evaluation of photovoltaic device performance while reducing computational overhead relative to dense DFT or drift-diffusion simulations. In addition, interpretable ML architectures, such as those utilizing Shapley Additive Explanations (SHAP) [59], quantify the relative importance of interfacial parameters, providing a crucial mathematical complement to traditional analytical transport methods.

However, widespread deployment encounters significant logistical hurdles. Foremost, the ML model accuracy remains constrained by the quality and dimensionality of training data. Experimental photovoltaic datasets are often small, often restricted to fewer than 100 localized device configurations, and inherit irreducible experimental uncertainties that propagate as prediction errors. Furthermore, the out-of-distribution generalization across different device architectures, e.g., shifting from planar to mesoporous perovskite layers) or varying material systems remains severely limited. Most empirical models are highly specialized to specific material systems like molecular donors or halide compositions, preventing direct transferability. Consequently, further studies are still required to improve the architectural universality and statistical reliability of ML-generated models, enabling their broader application.

To address these constraints, contemporary research is gearing toward Physics-Informed Machine Learning (PIML). Instead of treating neural networks as unconstrained black boxes, PIML embeds the basic semiconductor equations detailed in **Section 3.5**—specifically Poisson’s equation and the carrier continuity equations—directly into the neural network's loss functions. By penalizing solutions that violate these fundamental physical laws, PIML enforces strict physical consistency even when operating on sparse or noisy experimental training data. Ultimately, integrating these PIML frameworks with high-throughput automated synthesis loops and closed-loop artificial intelligence algorithms establishes a viable pathway toward autonomous, self-optimizing solar cell design.

6. Conclusion

This work maps the critical intersections of interface physics, numerical modeling, and data-driven optimization in modern photovoltaics. By evaluating heterojunction losses through the lens of SRH recombination, we have illustrated how localized trap states and band mismatches directly govern overall cell efficiency and operational stability. The effects of interface defects, particularly trap states, on non-radiative recombination were discussed in detail. In addition, the importance of band alignment in heterojunction structures and its influence on carrier transport and recombination dynamics were highlighted.

While conventional interface engineering approaches—namely surface passivation, band alignment tuning, and selective doping—effectively mitigate these losses, their design loop remains slow. Similarly, physics-based numerical simulations (e.g., TCAD tools) offer powerful platforms for quantitative analysis, yet they are constrained by a reliance on manual parameter iteration and prohibitive computational overhead when navigating multidimensional design spaces. Overcoming these throughput limitations necessitates a paradigm shift toward ML frameworks. Incorporating data-driven surrogate models into traditional simulation pipelines allows for rapid performance prediction and efficient exploration of complex architectural configurations. Nevertheless, critical hurdles persist. The predictive accuracy of standard ML models strongly depends on the availability of massive, high-quality datasets that are expensive and time-consuming to generate. Moreover, issues related to model interpretability and physical consistency still require further investigation. Resolving these bottlenecks will require the maturation of PIML and multi-scale modeling. Hence, future research should focus on the development of PIML methods, improvement of data efficiency, and integration of multi-scale modeling techniques to reduce computational cost and enhance model reliability. Forcing neural networks to adhere strictly to underlying thermodynamic and semiconductor transport laws will dramatically reduce baseline training data requirements while improving out-of-distribution prediction reliability. Ultimately, bridging the gap between empirical interface physics, advanced numerical software, and intelligent data optimization represents the core prerequisite for unlocking next-generation, high-efficiency multi-junction and heterostructure solar devices.

Compliance with ethical standards

Acknowledgments

The authors acknowledge institutional support and funding from the RISE Glocal 30 program (2026-Glocal30-03-001), and the Ningbo Natural Science Foundation (Grant Nos. 2025Z001, 2025Z039, 2024Z148, 2024Z234), and scientific discussions within the Global Matching Lab Support project framework contributing to this review.

Disclosure of conflict of interest

The authors declare no conflict of interest.

Funding

This work was supported by the Regional Innovation System & Education (RISE) program through the Daegu RISE Center, funded by the Ministry of Education (MOE) and Daegu Metropolitan City, Republic of Korea (2026-Glocal30-03-001), and the Ningbo Natural Science Foundation (Grant Nos. 2025Z001, 2025Z039, 2024Z148, 2024Z234).

References

[1] Agency, I.E., *Global Energy Review 2025,* IEA: Paris. https://www.iea.org/reports/global-energy-review-2025.

[2] Agency, I.E., *Global Energy Review 2026*, IEA: Paris. https://www.iea.org/reports/global-energy-review-2026.

[3] Zhang, Y., *Solar PV in the 21st century: Aligning technological growth with sustainability.* Next Energy, 2026. **10**: p. 100499.

[4] Machín, A. and F. Márquez, *Advancements in Photovoltaic Cell Materials: Silicon, Organic, and Perovskite Solar Cells.* Materials (Basel), 2024. **17**(5).

[5] ISE, F.I.f.S.E.S., *Photovoltaics Report*. 2025, Fraunhofer ISE: Freiburg, Germany.

[6] Laboratory, N.R.E. *Best Research-Cell Efficiency Chart*. 2026 [cited 2026; Available from: https://www.nrel.gov/pv/interactive-cell-efficiency.html.

[7] Zanatta, A.R., *The Shockley–Queisser limit and the conversion efficiency of silicon-based solar cells.* Results in Optics, 2022. **9**: p. 100320.

[8] Shockley, W. and H.J. Queisser, *Detailed Balance Limit of Efficiency of p-n Junction Solar Cells.* Journal of Applied Physics, 1961. **32**(3): p. 510-519.

[9] Shockley, W. and W.T. Read, *Statistics of the Recombinations of Holes and Electrons.* Physical Review, 1952. **87**(5): p. 835-842.

[10] He, F., et al., *Emitter/Absorber Interface Design Strategies for Se Solar Cells.* Chalcogenide Letters, 2025. **22**: p. 939-949.

[11] Kavanagh, S.R., et al., *Impact of metastable defect structures on carrier recombination in solar cells.* Faraday Discussions, 2022. **239**(0): p. 339-356.

[12] Anderson, R.L., *Experiments on Ge-GaAs heterojunctions.* Solid-State Electronics, 1962. **5**(5): p. 341-351.

[13] Wang, Y., J. Seunghwan, and B. Shin, *Interface engineering of antimony selenide solar cells: A review on the optimization of energy band alignments.* Journal of Physics: Energy, 2022. **4**.

[14] Njema, G.G., J.K. Kibet, and S.M. Ngari, *A review of interface engineering characteristics for high performance perovskite solar cells.* Measurement: Energy, 2024. **2**: p. 100005.

[15] Huang, L., et al., *Synergistic Interface Energy Band Alignment Optimization and Defect Passivation toward Efficient and Simple-Structured Perovskite Solar Cell.* Advanced Science, 2020. **7**: p. 1902656.

[16] Aberle, A.G., *Surface passivation of crystalline silicon solar cells: a review.* 2000. **8**.

[17] Karade, V.C., et al., *Machine Learning Drives a Path to Defect Engineering for Suppressing Nonradiative Recombination Losses in Cu2ZnSn(S,Se)4 Solar Cells.* ACS applied materials & interfaces, 2025.

[18] Khac, D.L., et al. *Influence/Effect of Deep-Level Defect of Absorber Layer and n/i Interface on the Performance of Antimony Triselenide Solar Cells by Numerical Simulation*. Sustainability, 2022. **14**, 6780 DOI: 10.3390/su14116780.

[19] Sugiura, T. and N. Nakano, *Review: Numerical simulation approaches of crystalline-Si photovoltaics.* Energy Science & Engineering, 2023. **11**(10): p. 3888-3906.

[20] Gaviria, J.F., et al., *Machine learning in photovoltaic systems: A review.* Renewable Energy, 2022. **196**: p. 298-318.

[21] Lookman, T., et al., *Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design.* npj Computational Materials, 2019. **5**(1): p. 21.

[22] Salgado-Conrado, L., C. Álvarez-Macías, and B. Reyes-Durán, *A Review of Simulation Tools for Thin-Film Solar Cells.* Materials (Basel), 2024. **17**(21).

[23] Song, Y.J., et al., *Influence of defects and band offsets on carrier transport mechanisms in amorphous silicon/crystalline silicon heterojunction solar cells.* Solar Energy Materials and Solar Cells, 2000. **64**(3): p. 225-240.

[24] Li, B., et al., *Defect Engineering toward Highly Efficient and Stable Perovskite Solar Cells.* Advanced Materials Interfaces, 2018. **5**: p. 1800326.

[25] Fu, L., et al., *Defect passivation strategies in perovskites for an enhanced photovoltaic performance.* Energy & Environmental Science, 2020. **13**(11): p. 4017-4056.

[26] Jin, H., et al., *It's a trap! On the nature of localised states and charge trapping in lead halide perovskites.* Materials Horizons, 2020. **7**(2): p. 397-410.

[27] Zhou, Y., et al., *Enhancing charge-emitting shallow traps in metal halide perovskites by >100 times by surface strain.* Joule, 2025. **9**(1): p. 101772.

[28] Wang, Z., et al., *Defects and Defect Passivation in Perovskite Solar Cells.* Molecules, 2024. **29**(9).

[29] Sherkar, T.S., et al., *Recombination in Perovskite Solar Cells: Significance of Grain Boundaries, Interface Traps, and Defect Ions.* ACS Energy Letters, 2017. **2**(5): p. 1214-1222.

[30] Leijtens, T., et al., *Carrier trapping and recombination: the role of defect physics in enhancing the open circuit voltage of metal halide perovskite solar cells.* Energy & Environmental Science, 2016. **9**(11): p. 3472-3481.

[31] Hall, R.N., *Electron-Hole Recombination in Germanium.* Physical Review, 1952. **87**(2): p. 387-387.

[32] Alnuaimi, A.A., *Effect of Interface Traps on the Performance of Heterojunction a-Si:H/c-Si Thin Film Solar Cells*. 2013, Khalifa University: Abu Dhabi.

[33] OghmaNano. *Shockley–Read–Hall (SRH) recombination theory*. Available from: https://www.oghma-nano.com/manual/theory-srh.html.

[34] Zhang, F., et al., *Comparative studies of optoelectrical properties of prominent PV materials: Halide perovskite, CdTe, and GaAs.* Materials Today, 2020. **36**: p. 18-29.

[35] Kim, S., et al., *Lone-pair effect on carrier capture in Cu2ZnSnS4 solar cells.* Journal of Materials Chemistry A, 2019. **7**(6): p. 2686-2693.

[36] Yue, Z., et al., *Bidirectional UV/violet heterojunction light-emitting diode with In0.27Al0.73N alloy film as electron transport layer.* Journal of Alloys and Compounds, 2024. **1001**: p. 175193.

[37] Kistner-Morris, J., et al., *Electric-field tunable Type-I to Type-II band alignment transition in MoSe2/WS2 heterobilayers.* Nature Communications, 2024. **15**(1): p. 4075.

[38] Liu, Z., et al., *Semiconductor heterojunctions for photocatalytic hydrogen production and Cr(VI) Reduction: A review.* Materials Research Bulletin, 2022. **147**: p. 111636.

[39] Özçelik, V.O., et al., *Band alignment of two-dimensional semiconductors for designing heterostructures with momentum space matching.* Physical Review B, 2016. **94**(3): p. 035125.

[40] Lei, C., et al., *Broken-Gap Type-III Band Alignment in WTe2/HfS2 van der Waals Heterostructure.* The Journal of Physical Chemistry C, 2019. **123**(37): p. 23089-23095.

[41] Wu, Y., et al., *Review of Interface Passivation of Perovskite Layer.* Nanomaterials (Basel), 2021. **11**(3).

[42] Shen, W., et al., *Interface passivation engineering for hybrid perovskite solar cells.* Materials Reports: Energy, 2021. **1**(4): p. 100060.

[43] Hammann, B., et al., *Review on hydrogen in silicon solar cells: From its origin to its detrimental effects.* Solar Energy Materials and Solar Cells, 2025. **282**: p. 113432.

[44] Kim, J.E., et al., *Characterization of SiNx:H thin film as a hydrogen passivation layer for silicon solar cells with passivated contacts.* Thin Solid Films, 2019. **675**: p. 109-114.

[45] Glunz, S.W. and F. Feldmann, *SiO2 surface passivation layers – a key technology for silicon solar cells.* Solar Energy Materials and Solar Cells, 2018. **185**: p. 260-269.

[46] J. Schmidt, A.M., R. Brendel, B. Hoex, M. C. M. van de Sanden, W. M. M. Kessels, *Surface passivation of high-efficiency silicon solar cells by atomic-layer-deposited* Al_2O_3*.* Progress in Photovoltaics: Research and Applications, 2008. **16**(6): p. 461–466.

[47] Aber, et al., *Field effect passivation of high efficiency silicon solar cells.* Solar Energy Materials and Solar Cells, 1993. **29**(2): p. 175-182.

[48] Mallorquí, A.D., et al., *Field-effect passivation on silicon nanowire solar cells.* Nano Research, 2015. **8**(2): p. 673-681.

[49] Bagade, S.S., et al., *Impact of band alignment at interfaces in perovskite-based solar cell devices.* Materials Today: Proceedings, 2023.

[50] He, C., et al., *Interface Engineering of BCP Buffer Layers in Planar Heterojunction Perovskite Solar Cells With NiOx Hole Transporting Layers.* Frontiers in Physics, 2018. **Volume 6 - 2018**.

[51] Huang, L., et al., *Deep insights into interface engineering by buffer layer for efficient perovskite solar cells: a first-principles study.* Science China Materials, 2020. **63**: p. 1588-1596.

[52] Li, S. and K. Bachagha, *Application of Buffer Layer to Improve the Efficiency of Organic Solar Cells.* Energy Technology, 2023. **12**.

[53] Shi, C., et al., *Surface Cleaning and Passivation Technologies for the Fabrication of High-Efficiency Silicon Heterojunction Solar Cells.* Materials (Basel), 2023. **16**(8).

[54] Deligiannis, D., et al., *Wet-chemical Treatment for Improved Surface Passivation of Textured Silicon Heterojunction Solar Cells.* Energy Procedia, 2014. **55**: p. 197-202.

[55] Ma, X., et al., *Interface Engineering of Perovskite Solar Cells with Air Plasma Treatment for Improved Performance.* ChemPhysChem, 2017. **18**(20): p. 2939-2946.

[56] Peleg, R., *New Argon plasma treatment improves the performance and stability of perovskite solar cells.* Perovskite-Info, 2025.

[57] Kim, S., et al., *Machine Learning-Assisted Defect Analysis and Optimization for P-I-N-Structured Perovskite Solar Cells.* Journal of Electronic Materials, 2023. **52**(9): p. 5861-5871.

[58] Li, Y., et al., *Accelerating defect analysis of solar cells via machine learning of the modulated transient photovoltage.* Fundamental Research, 2024. **4**(6): p. 1650-1656.

[59] Ragni, M., et al., *Machine learning for screening and predicting the best surface modifiers for a rational optimization of efficient perovskite solar cells.* APL Energy, 2024. **2**(3): p. 036102.