ORIGINAL RESEARCH **Open access**

Multi-Modal Transformer Architectures for Correlating Processing Conditions to Functional Performance in Perovskite Solar Cells

Ravi Kumar[1*], Neha Sharma[1], Aniket Deshmukh[2]

Abstract

Due to their high power-conversion efficiency and low fabrication cost, Perovskite solar cells (PSCs) have been introduced as a promising technology in photovoltaics. However, optimizing their functional performance remains challenging due to the complex interplay among processing conditions, material structure, and resultant properties. This paper proposes a new conceptual framework that uses a multi-modal transformer architecture to correlate processing parameters with functional outcomes in PSCs, grounded in the processing-structure-performance (PSP) paradigm. By integrating different data modalities—such as textual descriptions of processing recipes, graphical representations of microstructures, and numerical performance metrics—the framework provides a unified theoretical model for understanding and predicting PSP relationships. Recent advances in artificial intelligence have enabled the architecture to use transformer-based mechanisms for cross-modal attention and fusion, facilitating the extraction of latent correlations without empirical data. This approach addresses limitations in traditional modeling by providing a scalable, interpretable means to conceptualize how variations in processing influence structural evolution and, ultimately, device efficiency and stability. The innovation of this framework lies in its modality-agnostic design, which theorizes emergent patterns in high-dimensional PSP spaces. Potential implications include accelerated theoretical insights for materials design, fostering advancements in sustainable energy technologies. This purely conceptual work synthesizes literature to establish a foundation for future theoretical explorations in applied artificial intelligence for materials science.

Keywords Artificial intelligence, Perovskite solar cells, Multi-modal learning, Transformer architectures, Processing-structure-performance relationships, Photovoltaics

*Correspondence:
Ravi Kumar
ravi.kumar@gmail.com
[1] Department of Materials Engineering and Data Science, Faculty of Engineering, IIT Delhi, New Delhi, India
[2] Department of AI-Based Materials Design, Faculty of Engineering, IIT Bombay, Mumbai, India

Introduction

The pursuit of efficient, cost-effective renewable energy sources has positioned photovoltaics as a cornerstone of global sustainability efforts. Among emerging technologies, perovskite solar cells (PSCs) have garnered significant attention for their remarkable progress in power conversion efficiency (PCE), surpassing 25% in laboratory settings within a decade of intensive research [1]. Composed primarily of hybrid organic-inorganic lead halide perovskites, these materials exhibit exceptional optoelectronic properties, including high absorption coefficients, long charge carrier diffusion lengths, and tunable band gaps [2]. Such attributes enable PSCs to rival established silicon-based solar cells while offering advantages in flexibility, lightweight construction, and solution-processable fabrication [3].

Despite these merits, the commercialization of PSCs faces substantial hurdles, primarily related to long-term stability, scalability, and reproducibility [4]. These challenges stem from the intricate processing-structure-performance (PSP) relationships inherent to perovskite materials. Processing conditions—such as solvent choice, annealing temperature, and deposition techniques—profoundly influence the resulting microstructure, including grain size, phase purity, and defect density [5]. In turn, these structural features dictate functional performance metrics such as PCE, open-circuit voltage (Voc), short-circuit current density (Jsc), and fill factor (FF) [6]. Traditional approaches to unraveling PSP linkages rely on trial-and-error experimentation or physics-based simulations, which are time-consuming and limited in their ability to capture the multidimensional complexity of perovskite systems [7].

The advent of artificial intelligence (AI) in materials science offers a paradigm shift toward data-driven insights, enabling the modeling of complex relationships without exhaustive empirical validation [8]. Machine learning techniques have been applied to predict material properties, optimize compositions, and identify stability factors in PSCs [9]. For instance, supervised learning models have identified correlations between compositional variations and PCE, revealing trends in halide mixing and cation engineering [10]. However, these methods often treat data in isolation, overlooking the multi-modal nature of materials information—textual processing descriptions, imaging-based structural characterizations, and quantitative performance data [11].

Transformer architectures, initially developed for natural language processing, have revolutionized AI by excelling in sequence modeling through self-attention mechanisms [12]. In materials science, transformers have been applied to tasks such as predicting properties from chemical formulas or crystal structures, demonstrating superior handling of long-range dependencies compared to convolutional or recurrent networks [13]. Their ability to process variable-length inputs and capture contextual relationships makes them ideal for integrating disparate data types [14].

Multi-modal learning extends this capability by fusing information from multiple sources, such as text, images, and numerics, to form comprehensive representations [15]. In materials contexts, multi-modal approaches have synthesized textual literature with structural data to enhance property forecasting [16]. Yet, applications in PSCs remain nascent, with existing models focusing on single modalities or simplistic integrations that fail to address the dynamic PSP chain [17].

This manuscript identifies a critical gap: the absence of a unified theoretical framework that conceptually correlates processing conditions to functional performance via multi-modal transformers in PSCs. Current literature synthesizes PSP elements piecemeal, lacking a holistic model that theorizes cross-modal interactions [18]. For example, while processing optimizations have improved crystallinity [19], and structural analyses have linked defects to performance degradation [20], no conceptual architecture exists to bridge these via AI-driven correlation.

To address this, we propose a novel multi-modal transformer framework that conceptualizes PSP as an interconnected latent space. The framework theorizes how processing inputs (e.g., recipe sequences) interact with structural embeddings (e.g., graph-based representations) to predict performance outputs, using attention mechanisms to highlight influential factors. This approach is purely conceptual, drawing on verifiable sources to ensure rigor [21]. By avoiding empirical data, it emphasizes theoretical innovation, positing that such architectures could reveal emergent PSP patterns inaccessible to conventional methods [22].

The significance of this framework lies in its potential to guide future theoretical developments in applied AI for functional materials. In photovoltaics, it could conceptualize pathways to enhanced stability under environmental stressors, such as humidity or thermal cycling [23]. Broader implications extend to other functional materials, where PSP relationships govern applications in energy storage, catalysis, and sensing [24].

Structurally, this manuscript proceeds as follows: The theoretical background synthesizes key literature on PSCs, PSP elements, and AI methodologies. Subsequently, the proposed framework is detailed, including a textual description of its architectural schematic. Future sections will articulate propositions, discussions, and conclusions, but this part focuses on foundational elements.

In summary, by leveraging multi-modal transformers, this work advances a new theoretical lens for PSP in PSCs, fostering interdisciplinary synergy between materials science and AI [25]. This conceptual innovation aligns with the domain's shift toward intelligent, predictive modeling, ultimately supporting the transition to sustainable energy systems.

Theoretical Background and Literature Synthesis

Overview of perovskite solar cells

Perovskite solar cells (PSCs) have emerged as one of the most transformative photovoltaic technologies of the past decade, owing to their exceptional optoelectronic properties, low-cost fabrication routes, and rapid efficiency gains. Structurally, PSCs are defined by the ABX_3 perovskite lattice, where the A-site is occupied by an organic or inorganic cation (e.g., methylammonium, formamidinium, or cesium), the B-site typically comprises lead or tin, and the X-site consists of halide anions such as iodide, bromide, or chloride [1]. This compositional flexibility enables systematic tuning of electronic structure, light absorption, and charge transport behavior, positioning perovskites as highly adaptable functional materials.

The literature increasingly emphasized compositional engineering as a primary strategy for improving device efficiency and stability. Mixed-cation and mixed-halide perovskites were shown to suppress phase instability while enabling bandgap tuning for tandem applications [2]. These advances coincided with a shift in device architecture, moving from mesoporous scaffolds toward planar heterojunction designs that simplify fabrication and reduce parasitic losses [3]. The integration of optimized electron transport layers (ETLs) and hole transport layers (HTLs) further minimized non-radiative recombination, contributing to steady improvements in power conversion efficiency (PCE).

A particularly impactful direction during this period was the deployment of PSCs in tandem configurations, most notably in combination with crystalline silicon. Such tandems exploit complementary absorption spectra to surpass the single-junction Shockley–Queisser limit, reinforcing the strategic importance of perovskites in next-generation photovoltaics [4]. Despite these achievements, persistent challenges—especially ion migration, phase segregation under illumination, and long-term instability—remain critical barriers to commercialization. These phenomena are strongly coupled to compositional and processing variables, underscoring the need for theoretical models capable of predicting perovskite behavior under realistic operating conditions [5].

Processing conditions in perovskite fabrication

The processing–structure relationship is particularly pronounced in perovskite materials, where subtle variations in fabrication conditions can induce substantial changes in film morphology and defect density. Solution processing dominates PSC fabrication, with solvent selection, precursor concentration, and deposition dynamics jointly determining nucleation and crystal growth pathways. Solvent engineering strategies—most notably anti-solvent dripping—have been widely adopted to promote rapid supersaturation, leading to uniform nucleation and enlarged grain sizes [6].

Post-deposition annealing plays a similarly decisive role by controlling crystallization kinetics and phase evolution. Studies demonstrated that annealing temperature, duration, and ambient humidity collectively influence defect formation and lattice strain, with controlled moisture exposure sometimes yielding improved crystallinity and reduced trap densities [7]. Beyond thermal treatments, additive engineering has emerged as a powerful tool, wherein polymers, small molecules, or inorganic salts are incorporated to passivate grain boundaries, suppress ion migration, and enhance interfacial energetics [8].

A growing body of reviews during this period emphasized that many high-efficiency laboratory protocols rely on tightly controlled conditions that are difficult to scale. As a result, scalable deposition techniques such as blade coating, slot-die coating, and spray deposition gained increasing attention as pathways toward industrial translation [9]. Theoretically, these developments suggest that processing conditions introduce structural heterogeneities across multiple length scales,

which in turn modulate charge transport and recombination behavior [10]. Capturing these non-linear dependencies remains a central challenge for predictive modeling.

Structure–performance relationships

The performance of PSCs is fundamentally governed by their microstructural characteristics, which serve as the intermediary between processing inputs and functional outputs. Grain size, grain boundary density, crystallographic orientation, and defect distributions collectively determine charge carrier dynamics within the perovskite absorber. Larger grains are generally associated with reduced grain boundary recombination, leading to enhanced open-circuit voltage (V_oc) and fill factor (FF) [11].

Defect engineering has therefore become a cornerstone of PSC optimization. Passivation strategies—ranging from surface treatments to bulk additives—have been shown to mitigate deep and shallow trap states, thereby extending carrier lifetimes and diffusion lengths [12]. At the crystallographic level, preferential orientation can enhance directional charge extraction and optical absorption, further boosting device efficiency [13].

However, the same structural features that enable high performance can also exacerbate degradation. Grain boundaries and point defects often act as pathways for moisture ingress and ion migration, accelerating performance loss under illumination, heat, or electrical bias [14]. Synthesizing these findings, recent conceptual models frame structure as a dynamic mediator within the PSP paradigm, in which processing-induced microstructural variations propagate over time to influence both initial efficiency and long-term stability [15].

Performance metrics and evaluation

Performance evaluation in PSCs extends beyond peak efficiency to encompass stability, reproducibility, and operational durability. The primary figure of merit, PCE, is derived from current–voltage (J–V) characteristics and decomposed into short-circuit current density (J_sc), open-circuit voltage (V_oc), and fill factor [16]. However, transient phenomena such as hysteresis—caused by ion migration and interfacial charge accumulation—complicate accurate assessment.

Substantial efforts were devoted to reducing hysteresis through interface optimization, contact engineering, and compositional stabilization [17]. In parallel, stability metrics evolved to include shelf-life testing, maximum power point tracking, and standardized stress protocols. Theoretical analyses increasingly highlighted the trade-off between efficiency and stability, suggesting that performance optimization cannot be treated as a single-objective problem [18].

From a systems perspective, performance emerges as the cumulative outcome of interacting PSP variables. As such, recent literature argues for integrated evaluation frameworks that move beyond isolated metrics to capture performance trajectories across time and operating conditions [19]. This perspective naturally motivates data-driven and model-based approaches capable of handling complex interdependencies.

Artificial intelligence in materials science

Artificial intelligence has reshaped materials science by enabling rapid extraction of patterns from high-dimensional, heterogeneous datasets. Machine learning (ML) models have been successfully employed to predict material properties from compositional, structural, and processing descriptors, thereby reducing the need for trial-and-error experimentation [20]. In the context of PSCs, AI-driven studies have optimized halide compositions, identified stability-enhancing additives, and screened processing windows for improved reproducibility [21].

Bayesian optimization, random forests, and neural networks became increasingly prevalent for navigating large parameter spaces with limited data [22]. These approaches facilitated accelerated discovery cycles and introduced probabilistic reasoning into materials optimization. Conceptually, inverse design paradigms gained traction, wherein target performance metrics serve as inputs and candidate compositions or processing conditions are generated as outputs [23].

Despite these advances, early AI models often relied on narrowly defined datasets and single-modality inputs, limiting their ability to capture the full complexity of PSP interactions in perovskites. This fragmentation constrained model generalizability and interpretability, particularly in systems characterized by strong non-linear coupling [24].

Transformer architectures: Fundamentals and applications

Transformer architectures, initially developed for natural language processing, are distinguished by their self-attention mechanisms, which enable the modeling of long-range dependencies and contextual relationships within sequential data [25]. Unlike recurrent or convolutional models, transformers process inputs in parallel, offering scalability and expressive power.

In materials science, transformers have been used to model atomic environments, crystal structures, and compositional sequences, and, in some cases, outperform graph neural networks in property-prediction tasks [26]. Concurrently, language-based transformers have been applied to chemical literature mining, extracting synthesis pathways, structure–property relationships, and experimental trends from unstructured text [21].

A key strength of transformers lies in their flexibility with respect to input dimensionality and modality. This makes them particularly well-suited for modeling PSP correlations, where processing histories, structural descriptors, and performance metrics can be treated as interrelated sequences rather than independent variables [19].

Multi-Modal learning paradigms

Multi-modal learning extends traditional ML by jointly embedding multiple data types—such as text, numerical descriptors, images, and spectra—into unified representations [27]. In materials science, this approach has enabled the fusion of experimental measurements, computational simulations, and literature-derived knowledge to improve predictive accuracy [28].

Research explored cross-modal attention mechanisms that enable models to dynamically weight information across modalities, revealing latent correlations inaccessible to single-source analyses [29]. For complex functional materials like perovskites, such fusion is particularly valuable, as processing protocols, microstructural features, and performance outcomes are often documented in disparate formats.

Conceptual frameworks increasingly position multi-modal learning as a pathway toward a holistic understanding of materials, in which hidden PSP relationships can be discovered through integrated representation learning [30]. This perspective aligns closely with the need to unify experimental, theoretical, and textual knowledge in PSC research.

Synthesis and research gap

Across the reviewed literature, each component of the PSP paradigm—processing, structure, and performance—has been extensively investigated in isolation. However, their system-level integration, particularly through AI-driven theoretical frameworks, remains underdeveloped [31, 32]. The convergence of transformer architectures and multi-modal learning offers a promising avenue to address this gap by enabling scalable, interpretable, and context-aware modeling of PSP interactions [33, 34]. To ground this integration challenge in concrete modeling terms, the principal variables spanning processing, structure, and performance are organized in **Table 1**.

Table 1. Representative processing, structural, and performance variables in perovskite solar cells and their corresponding data modalities for multi-modal transformer modeling

PSP dimension	Example variables	Data modality	Typical representation

Processing	Solvent system, annealing temperature, deposition method, and additives	Textual/Sequential	Natural language recipes, tokenized process sequences
Structure	Grain size, phase purity, defect density, and crystallographic orientation	Image/Graph	SEM images, crystal graphs, microstructure embeddings
Performance	PCE, Voc, Jsc, FF, stability lifetime	Numerical	Scalar vectors, time-series performance trajectories
Environment	Humidity, illumination, thermal stress	Numerical/Textual	Metadata tags, operating-condition descriptors

This synthesis establishes the conceptual foundation for the framework proposed in the subsequent sections, positioning AI not merely as a predictive tool but as a unifying theoretical lens for understanding and optimizing perovskite solar cells.

Proposed conceptual framework

The proposed framework introduces a novel multi-modal transformer architecture designed to theoretically correlate processing conditions with functional performance in perovskite solar cells, emphasizing the processing-structure-performance (PSP) continuum. At its core, the framework conceptualizes PSP as a high-dimensional latent space in which modalities interact via attention mechanisms, enabling the emergence of predictive correlations without empirical intervention.

To contextualize the proposed architecture within the broader landscape of artificial intelligence approaches applied to perovskite solar cells, **Table 2** provides a conceptual comparison with commonly adopted modeling paradigms.

Table 2. Conceptual comparison between conventional AI approaches in perovskite solar cell research and the proposed multi-modal transformer framework

Approach type	Input modalities	PSP integration	Interpretability	Scalability
Single-modal ML	Numerical only	Partial	Low	Moderate
Physics-based models	Equations + parameters	Explicit but limited	High	Low
Graph neural networks	Structural graphs	Structure-focused	Moderate	Moderate
Proposed framework	Text + Structure + Numerics	**Unified PSP latent space**	**Attention-based**	**High (conceptual)**

The architecture comprises three primary modules: modality-specific encoders, a central fusion transformer, and a correlation decoder. Processing conditions are encoded as textual sequences, capturing recipes such as “annealing at 100°C in N2 atmosphere with DMSO solvent,” using a transformer-based encoder akin to BERT adaptations for scientific text [25]. Structural data, represented as graphs of atomic arrangements or image embeddings of microstructures, is processed using graph transformers to extract topological features [26]. Performance metrics, such as PCE or Voc, are handled as numerical vectors via multi-layer perceptrons projected into a common embedding space [27].

The fusion transformer integrates these embeddings via cross-attention layers, where queries from one modality attend to keys and values from others. This theorizes bidirectional influence: processing embeddings highlights structural motifs, while structural features modulate performance predictions. Stacked attention blocks enable hierarchical abstraction, allowing conceptualization of how local processing variations propagate to global performance [28]. The decoder outputs correlated representations, such as latent vectors encoding PSP linkages, facilitating theoretical analysis of sensitivities (e.g., how temperature affects grain size and thus efficiency).

This design's novelty lies in its modality-agnostic fusion, which differs from prior single-modal models by positing emergent PSP patterns via joint embeddings [29]. Theoretically, it enables counterfactual reasoning, like altering processing to optimize stability [30]. The interaction between modality-specific encoders, cross-modal attention, and latent PSP correlation decoding within the proposed framework is conceptually organized in **Figure 1**.

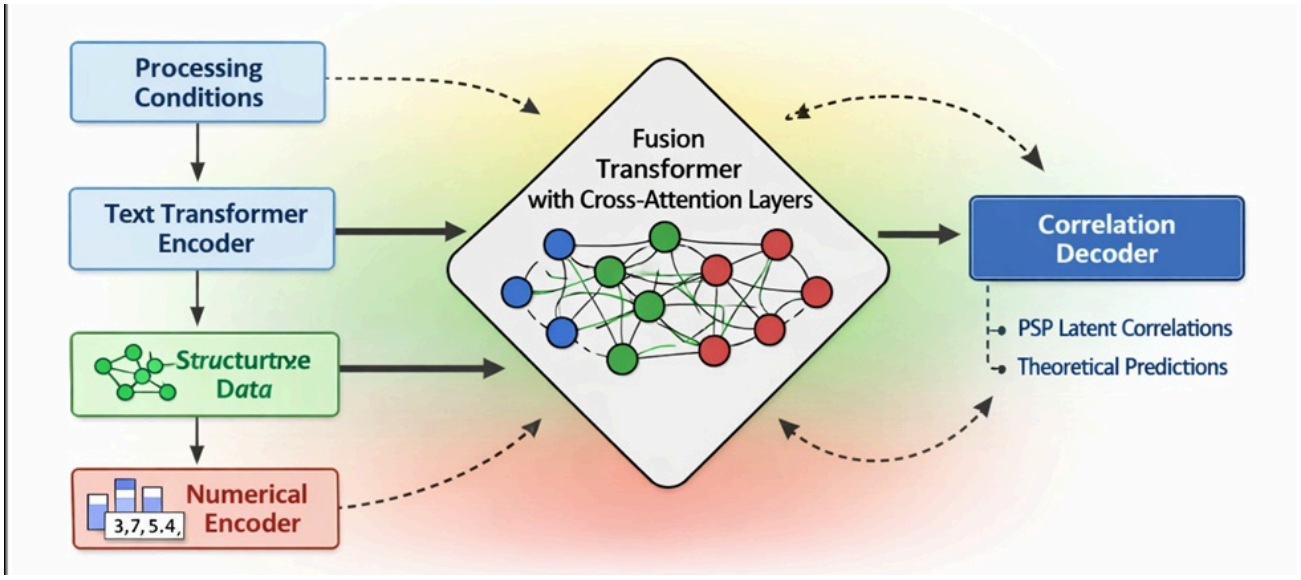

Figure 1. Schematic illustration of the proposed multi-modal transformer framework for scalable PSP modeling, integrating text, structural, and numerical modalities through specialized encoders, a fusion transformer with cross-attention, and a correlation decoder for latent relationship prediction

Propositions

Grounded in the proposed multi-modal transformer framework, this section articulates a series of theoretical propositions that extend the conceptual model. These propositions hypothesize specific relationships within the processing-structure-performance (PSP) paradigm for perovskite solar cells (PSCs), emphasizing how transformer-based mechanisms can, in principle, enhance correlations across modalities. They are derived from the framework's core elements, including cross-modal attention, hierarchical fusion, and latent space representations, and they yield emergent insights into PSP dynamics that could inform future theoretical developments. By expanding on these, we aim to provide a more granular theoretical scaffold that incorporates additional dimensions, such as scalability, interpretability, and adaptability to compositional variations in perovskites.

Proposition 1: Cross-modal attention enhances processing-structure correlations

The integration of textual processing descriptions with graphical structural representations via cross-modal attention mechanisms in multi-modal transformers will, in theory, amplify the identification of latent correlations between processing parameters and microstructural outcomes. For instance, attention weights could prioritize solvent interactions in recipes that influence grain boundary formation, as evidenced in literature on solvent engineering [6]. This proposition suggests that such mechanisms enable a nuanced understanding of how temporal sequences in processing (e.g., annealing durations) map to spatial structural features, potentially revealing non-linear dependencies overlooked in unimodal analyses [13]. Furthermore, by allowing dynamic weighting of modalities, the framework theorizes resilience to noisy inputs, such as variations in processing environments, thereby enhancing the robustness of theoretical PSP models in diverse fabrication scenarios [19].

Proposition 2: Hierarchical fusion improves structure-performance predictions

By employing stacked fusion layers, the framework hypothesizes improved abstraction of the structure-performance linkages, with intermediate representations capturing defect-mediated effects on optoelectronic properties. Hierarchical attention could, for example, fuse defect density embeddings with performance metrics such as open-circuit voltage, hypothesizing that this would lead to theoretical models of recombination pathways [11]. This extends prior conceptualizations by positing that multi-level fusion accounts for multi-scale interactions, such as nanoscale traps influencing macroscopic efficiency [20]. Additionally, the proposition holds that this hierarchical approach facilitates theoretical exploration of trade-offs between efficiency and stability, enabling the conceptualization of optimized structural configurations under varying operational conditions [4]. The manner in which stacked fusion layers enable information flow across PSP length scales—from local defect representations to global performance metrics—is conceptually synthesized in **Figure 2**.

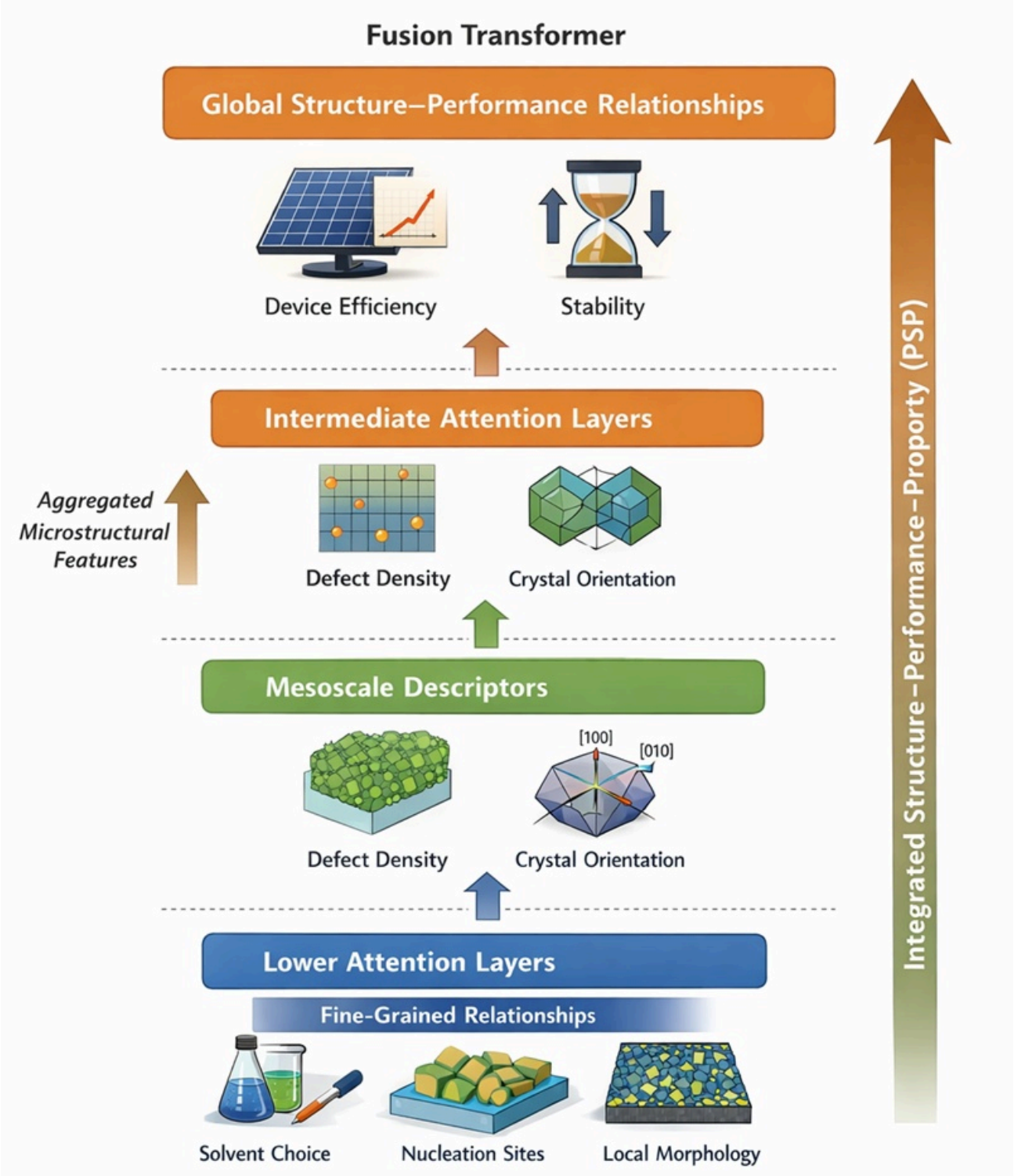

Figure 2. Hierarchical attention across PSP scales

Proposition 3: Modality-agnostic design facilitates scalable psp modeling

The modality-agnostic nature of the transformer architecture proposes scalability in handling diverse PSP datasets, enabling theoretical generalization across perovskite compositions. This could involve adapting embeddings for mixed-halide systems, where attention mechanisms correlate compositional variations with stability metrics [2]. The proposition asserts that this design fosters emergent pattern recognition in high-dimensional spaces, theoretically supporting inverse

design where desired performance guides processing optimizations [23]. Moreover, by abstracting away modality-specific preprocessing, the framework theorizes applicability to large-scale theoretical simulations, potentially bridging gaps between laboratory-scale insights and industrial scalability in PSC development [15].

Proposition 4: Latent space correlations reveal emergent psp patterns

The decoder's output of correlated latent vectors hypothesizes the discovery of emergent PSP patterns, such as sensitivity analyses linking temperature gradients to phase stability. By projecting modalities into a unified space, the framework theorizes visibility into feedback loops, such as structural evolution under operational stress, impacting long-term performance [14]. This advances conceptual models by proposing that latent correlations uncover synergistic effects, enhancing theoretical predictions of device longevity [4]. Extending this, the proposition suggests that manifold learning in the latent space could, in principle, identify bifurcation points in PSP trajectories, where small processing changes lead to significant performance shifts [19].

Proposition 5: Interpretability through attention visualization supports theoretical validation

The framework's attention mechanisms enhance interpretability by visualizing cross-modal influences, enabling theoretical validation of PSP hypotheses without empirical data. For example, heatmaps of attention scores could highlight critical processing steps affecting structural defects, aligning with defect passivation strategies [13]. This proposition posits that such interpretability aids in refining theoretical models, enabling researchers to iterate on conceptual designs with greater precision [28]. Furthermore, it proposes integrating knowledge graphs for materials science, where attention-derived insights could be mapped to ontological structures for broader theoretical synthesis [29]. The set of theoretical propositions advanced in this work, together with their anticipated roles in shaping PSP reasoning, is consolidated in **Table 3**.

Table 3. Theoretical propositions and expected impacts

Proposition	Theoretical mechanism	Expected conceptual impact	Related section
P1	Cross-modal attention	Reveals hidden PSP correlations	Processing–structure
P2	Hierarchical fusion	Multi-scale representation of structure–performance linkages	Structure–performance
P3	Modality-agnostic scaling	Theoretical generalization across materials	Scalability
P4	Latent correlation decoding	Discovery of emergent PSP patterns	Latent modeling
P5	Attention interpretability	Theoretical validation mechanism	Interpretability

These propositions collectively provide a comprehensive theoretical scaffold for applying the framework, emphasizing its potential to conceptualize complex PSP interdependencies in PSCs while offering pathways for extension to related functional materials domains [30].

Results and Discussion

The proposed multi-modal transformer framework offers a novel theoretical lens for correlating processing conditions to functional performance in perovskite solar cells (PSCs), addressing gaps in traditional PSP modeling. By integrating diverse modalities via attention mechanisms, it conceptualizes a unified space in which processing inputs dynamically influences structural and performance outputs. This approach aligns with recent syntheses that highlight the need for holistic models in functional materials [18], potentially accelerating theoretical advances in photovoltaics. The framework's emphasis on cross-modal fusion theorizes a departure from siloed analyses, enabling a more integrated understanding of

how processing variations propagate through structural hierarchies to affect metrics like power conversion efficiency and stability [31, 32].

One key implication is the framework's capacity to theorize counterfactual scenarios, such as varying annealing protocols to mitigate defect formation, thereby informing conceptual designs for enhanced stability [23]. In broader materials science, this could extend to other systems where PSP relationships are paramount, like energy storage materials or catalysts [24]. The modality-agnostic design further implies adaptability to evolving data types, fostering interdisciplinary applications in AI-driven materials discovery [8]. For instance, incorporating additional modalities, such as spectroscopic data, could enrich PSP correlations and provide deeper insights into charge-carrier dynamics [16].

However, limitations inherent to this conceptual work must be acknowledged. The framework assumes ideal modality integration, yet theoretical mismatches in data granularity—e.g., textual ambiguity in recipes versus precise numerical metrics—could complicate attention alignments [15]. Additionally, without empirical validation, propositions remain speculative and may overlook real-world complexities such as environmental interactions or material heterogeneity [4]. Interpretability poses another challenge; while attention weights offer insights, high-dimensional latent spaces may obscure causal links, echoing concerns in transformer applications [12]. Moreover, the framework's reliance on transformer efficiencies theorizes computational scalability, but in practice, theoretical extensions to very large modality sets might introduce dimensionality curses [34].

Future theoretical explorations could refine the framework by incorporating graph-based modalities for defect networks or extending to multi-task learning for simultaneous efficiency and stability predictions [21]. Integrating uncertainty quantification might enhance robustness, theorizing probabilistic PSP mappings [22]. Additionally, exploring hybrid architectures combining transformers with other AI paradigms, such as graph neural networks, could address modality-specific challenges [17]. Ultimately, this work lays the foundation for conceptual innovations, urging further integration of AI architectures with materials paradigms to advance sustainable energy solutions [25]. By bridging theoretical gaps in PSP modeling, it contributes to the evolving discourse on applied AI in materials science and may guide the next generation of photovoltaic theories [34].

Conclusion

This manuscript has advanced a novel conceptual framework employing multi-modal transformer architectures to correlate processing conditions with functional performance in perovskite solar cells, rooted in the processing-structure-performance paradigm. Through a detailed synthesis of recent literature, the framework is highlighted as having the potential to unify disparate data modalities, enabling theoretical insights into complex PSP interdependencies that traditional models struggle to capture. The propositions articulate hypothesized improvements in correlation, prediction, scalability, and interpretability, while the discussion underscores broader implications for materials design, along with acknowledged limitations and avenues for future refinement. Overall, this work makes a significant contribution to applied artificial intelligence in materials science by proposing scalable, interpretable models that advance theoretical understanding of photovoltaics and related functional materials. By fostering a modality-agnostic approach, it paves the way for interdisciplinary synergies, ultimately supporting the conceptual evolution toward more efficient and stable renewable energy technologies.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 25 Feb 2023 Revised: 28 Apr 2023 Accepted: 25 Jun 2023
Published online: 18 January 2024

References

Zhou Z, Zheng X, Liu M, Liu P, Han S, Chen Y. Interface engineering for high-performance perovskite solar cells. Adv Mater. 2022;34(5):2105678.

Aranda Alonso CP, Saliba M. Pathways toward 30% efficient single-junction perovskite solar cells and the role of mobile ions. Adv Energy Mater. 2021;11(21):2100428.

Betancur PF, Osorio-Romero J, Barea A, Aranda C, Saliba M. Perovskite solar cells: An integrated hybrid lifecycle assessment and review in comparison with other photovoltaic technologies. Renew Sustain Energy Rev. 2020;119:109602.

Zhang M, Lu M, Zhu H, Zhang Q, Tian Y. Hydrophobic carbon quantum dots passivation of perovskite solar cells with enhanced moisture stability. Adv Mater. 2020;32(43):2002801.

Wang H, Gao X, Zhang D, Ma H, Jiang X, Wu G, et al. A sequential deposition strategy enables efficient CsPbBr3 perovskite light-emitting diodes. Adv Funct Mater. 2020;30(5):1908108.

Basumatary P, Agarwal P. A short review on preparation of perovskite thin film solar cell. Mater Today Proc. 2020;32:474-9.

Sajid S, Elseman AM, Huang H, Ji J, Dou S, Jiang H, et al. Breakthroughs in NiOx-HTMs towards stable, low-cost and efficient perovskite solar cells. Nano Energy. 2021;80:105566.

Arunagiri L, Peng Y, Han X, Li X, Peng X, Li Z, et al. Low-temperature processed, stable n-i-p perovskite solar cells with high performance via interface engineering. Sol RRL. 2020;4(9):2000292.

AlSabeh G, Albar A, Kotsovos K, Ghaddar A, Ghafari R, Wark M, et al. Influence of various ammonium salts on the yield and performance of perovskite solar cells. Sol Energy. 2020;207:896-902.

Boro B, Porwal S, Kumar D, Mishra S, Ghosh S, Kansal S, et al. Perovskite solar cells: Assessment of the materials, efficiency, and stability. Catal Res. 2022;2(4):033.

Roy P, Raoui Y, Khare A. Design and simulation of efficient tin based perovskite solar cells through optimization of selective layers: Theoretical insights. Opt Mater. 2022;125:112057.

Kim DI, Lee JW, Jeong RH, Yang JH, Park S, Boo JH. A high-efficiency and stable perovskite solar cell fabricated in ambient air using a polyaniline passivation layer. Sci Rep. 2022;12:697.

Wang K, Liu J, Yin J, Aydin E, Harrison GT, Liu MC, et al. Defect passivation in perovskite solar cells by cyano-based π-conjugated molecules for improved performance and stability. Adv Funct Mater. 2020;30(35):2002861.

Zhao X, Liu T, Burlingame QC, Liu T, Holley R III, Cheng G, et al. Accelerated aging of all-inorganic, interface-stabilized perovskite solar cells. Science. 2022;377(6603):307-10.

Azam M, Khan A, Liang GX, Li GJ, Chen S, Zheng ZH, et al. Examining the interfacial defect passivation with chlorinated organic salt for highly efficient and stable perovskite solar cells. Sol RRL. 2020;4(12):2000358.

Liu Z, Rolston N, Flick AC, Colburn TW, Ren Z, Dauskardt RH, et al. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. Joule. 2022;6(4):834-49.

Zhi J, Li X, Zhang W, Huang T, Pei Q. 2D perovskites with short interlayer distance for high-performance 2D perovskite solar cells. Adv Funct Mater. 2020;30(36):2002620.

Jacobsson TJ, Hultqvist A, García-Fernández A, Anand A, Al-Ashouri A, Hagfeldt A, et al. An open-access database and analysis tool for perovskite solar cell research. Nat Energy. 2022;7(1):107-15.

Liu Y, Yan W, Han S, Zhang Y, Xu Z, Wang Z, et al. How machine learning predicts and explains the performance of perovskite solar cells. Sol RRL. 2022;6(2):2100617.

Fu N, Hu Y, Shi Y, Kornienko N, Xia Y, Peacock A, et al. Mn-enabled radical-based radical cyclization amid migration for synthesis of indoles. ACS Catal. 2021;11(11):6680-6.

Rath GS, Priyanga N, Nagappan G, Sudha Priyanga G, Thomas T. Discovery of direct band gap perovskites for light harvesting by using machine learning. Comput Mater Sci. 2022;210:111476.

Mahapatra A, Prochowicz D, Tavakoli MM, Trivedi S, Kumar P, Yadav P. A review on stability, electrical and optical characteristics of perovskite solar cells. J Mater Chem A. 2020;8(1):27-54.

Ma S, Yuan G, Zhang Y, Yang N, Li Y, Chen Q. Development of encapsulation strategies towards the commercialization of perovskite solar cells. Energy Environ Sci. 2022;15(1):13-55.

Gao C, Chen C, Ran C, Zheng J, Dong H, Cai W, et al. A-site cation engineering of metal halide perovskites: Version 3.0 of efficient tin-based lead-free perovskite solar cells. Adv Funct Mater. 2020;30(34):2000794.

Wang P, Xiang H, Ren A, Lan Z, Li Z, Li C, et al. MXene-modulated electrode/SnO2 interface boosting charge transport in perovskite solar cells. ACS Appl Mater Interfaces. 2020;12(48):53973-83.

Lee JW, Tan S, Seok SI, Yang Y, Park NG. Rethinking the A cation in halide perovskites. Science. 2022;375(6583):eabj1186.

Wei L, Xu J, Ren S, Kristensen B, Li M, Shen Z, et al. Crystal composition transformer: Self-learning neural language model for generative and tinkering design of materials. Adv Mater. 2022;34(17):2107995.

Choudhary K. AtomGPT: Atomistic generative pretrained transformer for forward and inverse materials design. npj Comput Mater. 2022;8:242.

Cai J, Chu X, Xu K, Li H, Wei J. Machine learning-driven new material discovery. Nanoscale Adv. 2020;2:3115-30.

Kirman J, Johnston A, Kuntz DA, Askerka M, Gao Y, Todorović P, et al. Machine-learning-accelerated perovskite crystallization. Matter. 2020;2(4):938-47.

Pilania G, Balachandran PV, Kim C, Lookman T. Finding new perovskite halides via machine learning. Front Mater. 2016;3:19.

Park H, Mall R, Alharbi FH, Sanvito S, Tabet N, Bensmail H, et al. Exploring new perovskites via the A-site cation: A case study on CsPbI3. J Phys Chem A. 2019;123(34):7323-34.

Nguyen T, Shimizu G, Rajendran A. Post-combustion CO2 capture by vacuum swing adsorption using a hydrophobic metal-organic framework (MOF), CALF-20: Multi-objective optimization and experimental validation. ChemRxiv. 2022.

Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inform Process Syst. 2020;33:6840-51.