

Decision-aware multi-objective discovery of Cu₂O-based photovoltaic materials integrating supply risk, stability, and band gap engineering

Author names:

Takahiro Kono¹, Souta Miyamoto^{1,2}, and Katsuaki Tanabe^{1,2,3,*}

Author address:

¹Department of Chemical Science and Engineering, Kyoto University, Nishikyo, Kyoto 615-8510, Japan

²School of Chemical Science and Technology, Kyoto University, Nishikyo, Kyoto 615-8510, Japan

³Kyoto MPI Inc., Sakyo, Kyoto 606-8501, Japan

Corresponding author:

*Email: tanabe@cheme.kyoto-u.ac.jp

Keywords: Semiconductor; Solar cell; Band gap; Machine learning; Density functional theory

Abstract

The development of photovoltaic materials that simultaneously satisfy performance requirements and practical constraints is essential for large-scale energy applications. In this study, we propose a decision-aware, multi-objective screening framework for Cu_2O -based materials targeting top-cell applications in Si-based multijunction solar cells. The framework integrates three key criteria: band gap suitability, supply risk, and thermodynamic stability. Band gap properties are predicted using machine learning models with uncertainty-aware scoring, while supply risk is quantified using the Herfindahl–Hirschman Index, and thermodynamic stability is evaluated via formation enthalpy predictions. These metrics are combined into a unified scoring function to systematically rank candidate materials. Top-ranked candidates are subsequently validated using first-principles calculations. The screening results identify Se-, P-, and S-substituted systems as high-scoring candidates. However, first-principles validation reveals that many top-ranked compounds exhibit undesirable electronic properties, such as metallic behavior or indirect band gaps. In contrast, $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ retains semiconducting character with a direct band gap and shows a reduction in band gap relative to pristine Cu_2O , while maintaining thermodynamic stability. These results demonstrate that multi-objective screening alone is insufficient to guarantee physically viable materials and highlight the importance of integrating data-driven approaches with physics-based validation. The proposed framework provides a robust approach for identifying practically viable photovoltaic materials and can be extended to other material systems under real-world constraints.

1. Introduction

The global demand for renewable energy has been rapidly increasing, leading to a significant expansion in the deployment of photovoltaic technologies. However, conventional silicon-based solar cells, which currently dominate the market, are approaching their theoretical efficiency limits [1–3]. As a result, achieving substantial performance improvements with single-junction devices has become increasingly challenging. To overcome this limitation, multi-junction solar cells have attracted considerable attention. By stacking multiple absorber layers with different band gaps, these devices enable efficient utilization of the solar spectrum over a wide energy range, thereby reducing thermalization and transmission losses and significantly improving overall conversion efficiency [4–7]. In addition to performance or efficiency, the large-scale deployment of photovoltaic technologies critically depends on material cost, abundance, and supply stability. Many high-efficiency solar cell materials contain scarce elements such as In, Ga, and Te [8–10], which suffer from limited crustal abundance and geographically concentrated production. These factors introduce significant supply risks and price volatility, posing challenges for sustainable and scalable energy systems. Therefore, identifying materials that are not only high-performing but also economically and environmentally sustainable is a central challenge for next-generation photovoltaics.

Among the candidate materials for solar cells, cuprous oxide (Cu_2O) has emerged as a particularly promising option. Cu_2O is composed of earth-abundant, low-cost, non-toxic elements and exhibits favorable optoelectronic properties [11–15]. These characteristics make Cu_2O highly promising for scalable and sustainable photovoltaic applications and an attractive candidate, for example, for the top cell in multijunction configurations with bottom cells such as Si or CIGS. However, achieving efficient carrier transport and high open-circuit voltage requires precise tuning of its electronic properties, particularly the band gap. For instance, the

intrinsic band gap of Cu_2O (~ 2.17 eV) is larger than the optimal value (~ 1.7 – 1.8 eV) required for top-cell absorbers in Si-based multijunction solar cells, necessitating band gap engineering to achieve improved device performance. Elemental substitution provides an effective strategy for tuning the electronic structure of semiconductors. In the case of Cu_2O , partial substitution at the oxygen site can modify the band structure and potentially improve carrier transport properties [16,17]. While previous studies have explored substitution strategies to optimize specific properties, most efforts have focused primarily on performance metrics such as band gap, often neglecting practical constraints such as material availability and supply stability.

The exploration of novel materials through purely experimental approaches is time-consuming and resource-intensive. In this context, machine learning has emerged as a powerful tool for accelerating materials discovery by enabling high-throughput screening of vast chemical spaces [18–22]. Nevertheless, machine learning-based screening alone does not guarantee physical validity or practical applicability, as predictions are inherently limited by training data. Therefore, integrating data-driven screening with physics-based validation is essential for reliable materials design under real-world constraints. For example, first-principles calculations based on density functional theory (DFT) provide a reliable and widely used approach for such validation, allowing detailed evaluation of electronic structures and derived properties prior to experimental verification. In this work, we propose a decision-aware, multi-objective materials discovery framework for Cu_2O -based photovoltaic materials. The framework integrates three key criteria: band gap suitability, supply risk, and thermodynamic stability, enabling systematic exploration of trade-offs among competing objectives. Candidate materials are first screened using machine learning-based predictions and quantitative scoring, followed by first-principles validation of the most promising compounds. This integrated approach aims to bridge the gap between data-driven screening and physically grounded validation, thereby enabling more reliable and application-oriented materials discovery. Through the integrated scheme, we demonstrate that Se substitution in Cu_2O is particularly

effective in achieving a balanced optimization, while also revealing that some high-scoring candidates exhibit undesirable electronic properties. These findings highlight the limitations of purely data-driven screening and underscore the importance of integrating physical validation. More broadly, this study provides a pathway toward cost-aware and scalable materials discovery, which is essential for the practical realization of next-generation photovoltaic technologies.

2. Methods

2-1. Candidate Space and Multi-Objective Screening

In this study, we focused on Cu₂O-based materials as candidate top-cell absorbers for Si-based multijunction solar cells. To tune the band gap of Cu₂O toward the optimal value for tandem applications, partial substitution at the oxygen sites was considered. In practical device configurations, Cu₂O is typically used as a *p*-type semiconductor and paired with *n*-type semiconductors such as ZnO or Ga₂O₃. Therefore, the materials explored in this study were assumed to form heterojunctions with such *n*-type counterparts. The target materials were defined by substituting oxygen atoms in Cu₂O with elements from groups 15–17, while maintaining charge neutrality. Let M denote the substituting element and *x* the substitution fraction. The compositions were defined as follows:

Group 15: Cu₂O_{1-1.5*x*}M_{*x*}

Group 16: Cu₂O_{1-*x*}M_{*x*}

Group 17: Cu₂O_{1-0.5*x*}M_{*x*}

The substitution fraction *x* was varied from 0 to 0.50 in increments of 0.05. Compounds containing elements for which no data were available in the machine learning models or evaluation metrics were excluded from the search space. To evaluate candidate materials, three criteria were considered: (i) band gap suitability, (ii) supply risk, and (iii) thermodynamic stability. The overall framework follows the multi-objective, uncertainty-aware approach proposed in Ref. [23], with modifications to the target properties and evaluation criteria. In the present study, the original objective related to catalytic performance is replaced with band gap suitability, while retaining the underlying decision-oriented scoring strategy.

Band gap values were predicted using the corrected dataset constructed in our previous

study [24], where systematic errors in DFT calculations were reduced using symbolic regression. In that approach, a physically interpretable correction function was derived to map DFT-calculated band gaps to experimental values, enabling the generation of a large, high-quality dataset. The resulting dataset significantly expands the available training data while maintaining physical consistency. Specifically, semiconductor compounds with both experimentally measured and DFT-calculated band gaps were used to derive a symbolic regression equation relating the two quantities. The obtained correction function was subsequently applied to a large collection of DFT band gaps compiled from the AFLOW and Materials Project databases, yielding corrected band gap values. The corrected values were treated as approximations of experimental band gaps and used to construct an expanded training dataset for machine learning-based band gap prediction [24]. Based on this dataset, machine learning models were trained using 129 compositional descriptors derived from elemental properties. These descriptors capture statistical, electronic, and structural characteristics of constituent elements, allowing accurate prediction of band gap values across a wide range of compounds. The band gap score A_j for compound j with a band gap $E_{g,j}$ was defined as:

$$A_j = \int_{1.6}^{1.9} \mathcal{N}(E_{g,j} | \sigma_j^2) dE_g \quad (1)$$

where N denotes a normal distribution with mean $E_{g,j}$ and standard deviation σ_j , calculated from the variance across the individual predictions generated by the 100 constituent decision trees that make up the Random Forest ensemble. This formulation represents the probability that the predicted band gap falls within the desired range. For Si-based dual-junction solar cells, the optimal top-cell band gap is approximately 1.73 eV [25,26]. Therefore, the target range was set to 1.6–1.9 eV. In practice, the variance of the Random Forest predictions was significantly larger than the target window (0.3 eV). This caused the score A_j to become uniformly small for

all candidates, making it difficult to rank the materials clearly. To resolve this and restore the ability to screen candidates, the predicted uncertainties required calibration to a realistic physical baseline. Specifically, considering that the typical distinguishable resolution of band gap values using high-accuracy electronic structure methods is approximately 0.2 eV [27], the standard deviations were globally scaled such that the average standard deviation across all candidate compounds becomes 0.2 eV. The corrected standard deviation for each compound was obtained as:

$$\sigma_{j,corr} = \frac{0.20}{\overline{\sigma_{raw}}} \sigma_{j,raw} \quad (2)$$

where $\overline{\sigma_{raw}}$ is the original average standard deviation over all compounds and $\sigma_{j,raw}$ is the original standard deviation for compound j . This procedure effectively sharpens the predicted distributions while maintaining relative uncertainty differences among compounds, thereby enabling more discriminative evaluation of band gap suitability within the target range. The reported uncertainty values correspond to the standard deviation of predictions obtained from an ensemble of independently trained Random Forest models. These values are used as a relative measure of model disagreement and candidate ranking rather than as rigorously calibrated probabilistic uncertainties.

The Herfindahl–Hirschman Index (HHI), originally developed to quantify market concentration, was employed here as a proxy for supply risk based on the geographical distribution of elemental resources. A higher HHI indicates stronger concentration and thus higher supply risk. The supply risk score C_j for compound j was defined as:

$$C_j = \frac{1}{N_t} \sum_{i=1}^{N_t} N_i \left[1 - \frac{HHI_j^i - \min(HHI)}{\max(HHI) - \min(HHI)} \right] \quad (3)$$

where N_i is the total number of elements in the compound, N_i is the number of atoms of element, and HHI_j^i is the HHI value of element i in compound j . The normalization ensures that higher scores correspond to lower supply risk. HHI data are obtained from Ref. [28], using reserve-based values.

Formation enthalpy was adopted as a metric to evaluate the thermodynamic stability of candidate compounds. The dataset was obtained from version 1.6 of the Open Quantum Materials Database (OQMD) [29]. To ensure data quality, duplicate entries corresponding to the same compound were removed by retaining only the entry with the lowest formation enthalpy. In addition, elemental entries were excluded since their formation enthalpy is defined as zero. Following the procedure described in Ref. [30], outliers beyond $\pm 5\sigma$ of the distribution (where σ is the standard deviation of the dataset) were also removed. As a result, a total of 644,008 data points were retained for model training. For the prediction of formation enthalpy, we employed Roost (Representation Learning from Stoichiometry) [31], a graph neural network-based model known for its high predictive accuracy. To determine the optimal number of training epochs, 5-fold cross-validation was performed with an 80/20 split between training and test sets. Model performance was evaluated using the mean absolute error. The final model was trained on the full dataset using the optimized hyperparameters. The trained Roost model was then applied to candidate compounds, and the predicted formation enthalpies were subsequently used to assess thermodynamic stability. The stability score S_j was defined as:

$$S_j = 1 - \frac{H_{f,j} - \min(H_f)}{\max(H_f) - \min(H_f)} \quad (4)$$

where $H_{f,j}$ is the predicted formation enthalpy of compound j , and $\max(H_f)$, $\min(H_f)$ are the maximum and minimum values within the candidate set. Lower formation enthalpy

corresponds to higher stability.

The integrated, overall score for compound j was defined as the product of the three metrics:

$$AQ_j = A_j \times C_j \times S_j \quad (5)$$

A higher AQ_j indicates a more promising candidate, balancing performance, stability, and supply considerations. It should be noted that both the band-gap prediction model and the Roost formation-enthalpy model employed in this study are composition-based and do not explicitly incorporate crystal structure information. Consequently, these models cannot distinguish polymorphs having identical compositions, nor can they capture the effects of local atomic environments, coordination geometries, defect configurations, or short-range ordering on material properties. While composition-based approaches are advantageous for rapid screening over large chemical spaces, their predictions should be regarded as approximate indicators of materials potential rather than definitive evaluations of specific crystal structures. Therefore, first-principles calculations were subsequently performed for selected high-ranking candidates to assess their structural and electronic properties more rigorously.

2-2. First-Principles Calculations

To validate the top-ranked candidates, first-principles calculations were performed based on DFT. For substitutional systems, the initial crystal structures were constructed using special quasi-random structures (SQS) to approximate random solid solutions while avoiding artificial ordering effects. SQS structures were generated using the mcsqs code in the Alloy Theoretic Automated Toolkit (ATAT) [32], based on the Cu_2O unit cell obtained from the

Materials Project database. To accurately represent target compositions under periodic boundary conditions, supercells were constructed such that the number of atoms of each element becomes an integer. For substitutions involving group 15 and 17 elements, charge balance was maintained by introducing vacancies and/or interstitial atoms where necessary. DFT calculations were performed using the Vienna Ab initio Simulation Package (VASP) [33–35], employing the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional [36]. The electronic self-consistency convergence criterion was set to 10^{-5} eV, and structural optimization was performed until atomic forces were below 0.02 eV/Å. A plane-wave cutoff energy of 520 eV was used. Brillouin zone sampling was conducted using a $2\times 2\times 2$ k -point mesh for structural relaxation and a denser $6\times 6\times 6$ mesh for static calculations. The primary quantity of interest is the band gap. Although the target band gap is 1.73 eV, it is well known that PBE systematically underestimates band gaps. For example, the DFT-calculated band gap of pristine Cu₂O (~0.5 eV) is significantly smaller than the experimental value (~2.17 eV). Therefore, instead of directly comparing absolute values, we evaluated whether the calculated band gap decreases relative to that of pristine Cu₂O, indicating a shift toward the desired range. In addition, formation enthalpy was calculated for selected candidates to assess thermodynamic feasibility. Lower formation enthalpy indicates higher likelihood of successful synthesis.

It should be noted that the PBE exchange–correlation functional is known to significantly underestimate the band gaps of transition-metal oxides due to self-interaction errors and the insufficient localization of d-electron states [37]. In particular, Cu₂O contains Cu 3d states that strongly influence the valence-band edge, and the inaccurate description of these states often results in substantial underestimation of the band gap compared with experimental measurements [38]. Previous studies have reported that while the experimental band gap of Cu₂O is approximately 2.0–2.2 eV, standard PBE calculations typically predict values around 0.5–0.8 eV. Therefore, the absolute band-gap values and electronic character predicted by PBE

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0339444

should be interpreted with caution. In the present study, first-principles calculations are primarily used to evaluate relative trends among structurally similar compounds rather than to provide quantitatively accurate electronic properties. More reliable assessment of absolute band gaps and electronic character would require higher-level approaches such as DFT+U, hybrid functionals (e.g., HSE06), or many-body perturbation methods (GW) [39,40], which are beyond the scope of the present work.

3. Results and Discussion

3-1. Evaluation of Multi-Objective Screening

The evaluation metrics defined in Section 2-1 were calculated for all candidate compounds in the search space. Regarding the underlying machine learning models, the Random Forest model for band gap prediction achieved an RMSE of 0.7538 eV. For the phase stability screening, the Roost model for formation enthalpy prediction achieved an MAE of 0.0567 eV/atom, demonstrating high predictive accuracy. The top 20 compounds ranked by the integrated score AQ_j are summarized in Table 1. Overall, compounds with anion-site substitution by P, Se, and S dominate the top-ranked candidates. Notably, Se-substituted compounds account for half of the top 20 entries, including $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$, which ranks second. This trend strongly suggests that Se substitution is particularly effective in optimizing the trade-off among band gap suitability, supply risk, and thermodynamic stability in the present material system. Among all candidates, $\text{Cu}_2\text{O}_{0.775}\text{P}_{0.150}$ exhibits the highest AQ_j , indicating that it achieves a well-balanced combination of the three evaluation criteria. This result highlights the importance of multi-objective optimization, as the top-ranked materials are not necessarily those with the best individual property but those that achieve a favorable balance across competing objectives.

Compound	A_j	C_j	S_j	AQ_j
$\text{Cu}_2\text{O}_{0.775}\text{P}_{0.150}$	0.538	0.860	0.711	0.329
$\text{Cu}_2\text{O}_{0.800}\text{Se}_{0.200}$	0.501	0.881	0.700	0.309
$\text{Cu}_2\text{O}_{0.750}\text{Se}_{0.250}$	0.521	0.878	0.672	0.307
$\text{Cu}_2\text{O}_{0.700}\text{P}_{0.200}$	0.573	0.848	0.619	0.301
$\text{Cu}_2\text{O}_{0.700}\text{Se}_{0.300}$	0.541	0.874	0.628	0.297
$\text{Cu}_2\text{O}_{0.650}\text{S}_{0.350}$	0.540	0.887	0.607	0.290

$\text{Cu}_2\text{O}_{0.850}\text{Se}_{0.150}$	0.454	0.885	0.711	0.286
$\text{Cu}_2\text{O}_{0.500}\text{S}_{0.500}$	0.543	0.883	0.587	0.281
$\text{Cu}_2\text{O}_{0.650}\text{Se}_{0.350}$	0.561	0.870	0.573	0.280
$\text{Cu}_2\text{O}_{0.850}\text{P}_{0.100}$	0.414	0.872	0.767	0.277
$\text{Cu}_2\text{O}_{0.600}\text{Se}_{0.400}$	0.600	0.867	0.532	0.276
$\text{Cu}_2\text{O}_{0.850}\text{As}_{0.100}$	0.507	0.878	0.616	0.274
$\text{Cu}_2\text{O}_{0.775}\text{As}_{0.150}$	0.558	0.868	0.564	0.274
$\text{Cu}_2\text{O}_{0.550}\text{Se}_{0.450}$	0.600	0.863	0.511	0.265
$\text{Cu}_2\text{O}_{0.550}\text{S}_{0.450}$	0.508	0.884	0.584	0.262
$\text{Cu}_2\text{O}_{0.700}\text{S}_{0.300}$	0.478	0.888	0.614	0.261
$\text{Cu}_2\text{O}_{0.850}\text{Sb}_{0.100}$	0.529	0.881	0.555	0.259
$\text{Cu}_2\text{O}_{0.900}\text{Se}_{0.100}$	0.404	0.889	0.718	0.258
$\text{Cu}_2\text{O}_{0.500}\text{Se}_{0.500}$	0.601	0.859	0.497	0.256
$\text{Cu}_2\text{O}_{0.600}\text{S}_{0.400}$	0.486	0.885	0.592	0.255

Table 1: List of the top-ranked compounds based on the integrated evaluation score AQ_j along with their corresponding band gap score (A_j), supply risk score (C_j), and thermodynamic stability score (S_j).

To further elucidate the role of each substituting element, the average values of the evaluation metrics AQ_j over all substitution ratios x were computed for each element and visualized (Fig. 1). The results confirm that Se consistently yields high AQ_j values across a wide composition range, while P and S also exhibit relatively favorable performance. In contrast, substitutions with group 17 elements and nitrogen tend to result in significantly lower scores, indicating less favorable trade-offs. To gain deeper insight into the origin of these trends, the dependence of each metric on the substitution fraction was analyzed for Se-substituted systems (Fig. 2). The band gap-related metric A_j increases monotonically with increasing Se content, whereas the thermodynamic stability and supply-related metrics exhibit different, often competing trends. These results demonstrate that the final ranking is governed by a non-trivial interplay among multiple factors, and that high-performing candidates emerge only when these competing effects are simultaneously optimized.

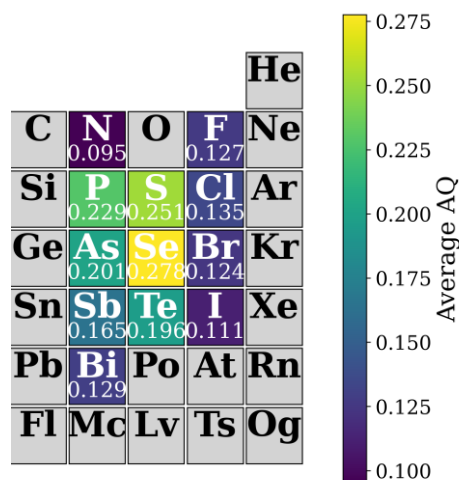


Fig. 1: Average values of the integrated evaluation score AQ_j for each substituting element considered (Groups 15–17 of the periodic table), illustrating their overall impact on material performance across all substitution ratios.

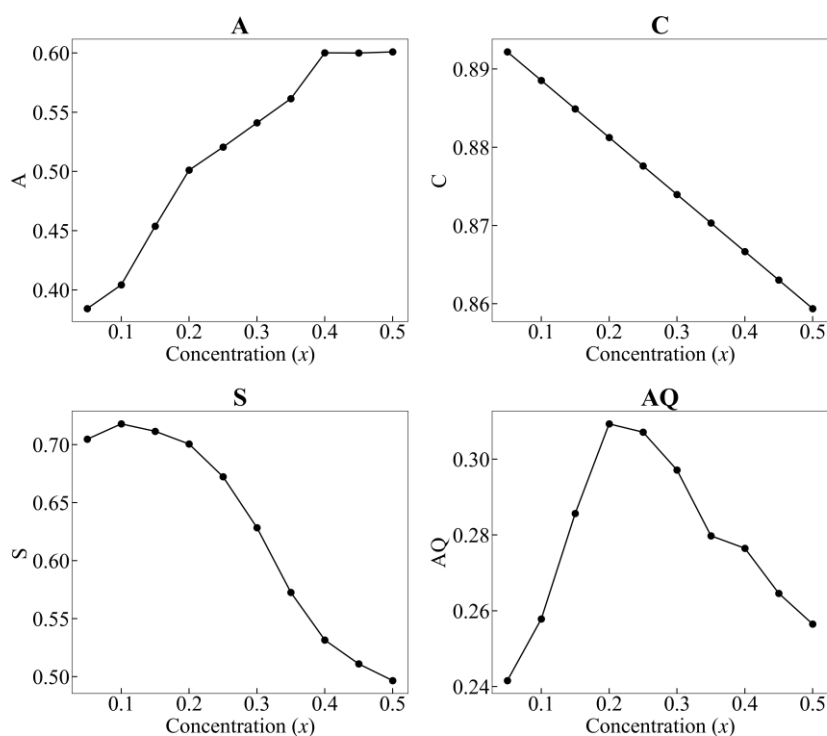


Fig. 2: Dependence of evaluation metrics (A_j , C_j , and S_j , and AQ_j) on Se substitution fraction x , illustrating trade-offs among band gap suitability, supply risk, and thermodynamic stability.

The machine learning models employed in this study rely exclusively on compositional information and do not explicitly account for crystal structure, local atomic environments, or configurational ordering effects. Consequently, discrepancies between machine learning predictions and subsequent first-principles validation are expected for some candidate materials. In addition, the predictive reliability of machine learning models generally decreases when applied to compositions that are underrepresented in the training data. Therefore, the present screening framework should be regarded as a tool for prioritizing promising candidates rather than providing definitive property predictions, highlighting the continued importance of first-principles validation.

3-2. First-Principles Validation

3-2-1. Reference DFT Calculation for Pristine Cu₂O

As a reference for comparison with the substituted compounds, the electronic structure of pristine Cu₂O was first evaluated. The density of states and band structure obtained from first-principles calculations are shown in Figs. 3 and 4, respectively. In Fig. 3, the zero of the energy axis corresponds to the Fermi level. The calculated band gap of Cu₂O was 0.5039 eV, which is in good agreement with the value reported in the Materials Project database (~0.51 eV), where calculations are primarily performed using the PBE functional. The slight discrepancy is attributed to differences in computational settings, such as *k*-point sampling. For consistency, all subsequent first-principles calculations for the top-ranked candidate compounds were performed under the same computational conditions as those used for pristine Cu₂O.

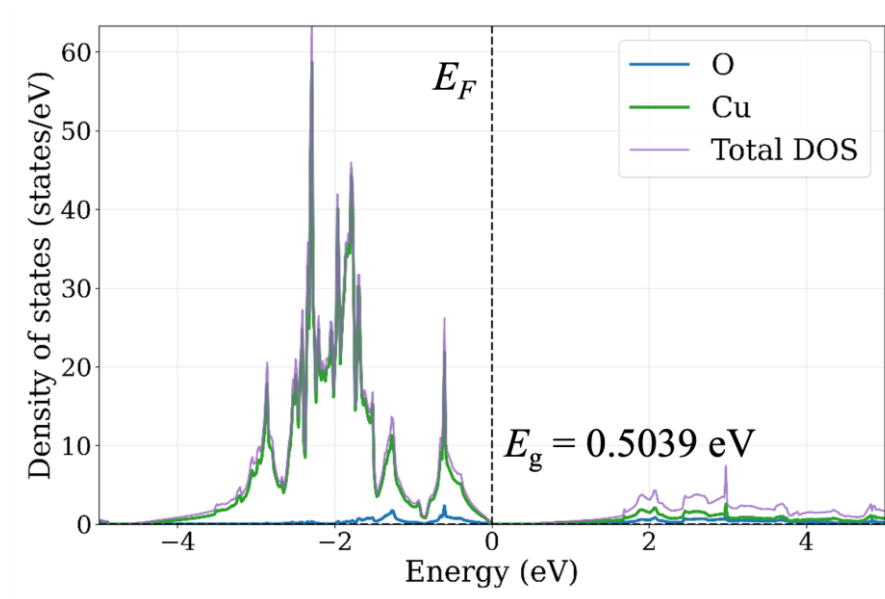


Fig. 3: Density of states of pristine Cu₂O calculated using DFT. The Fermi level is set to zero energy.

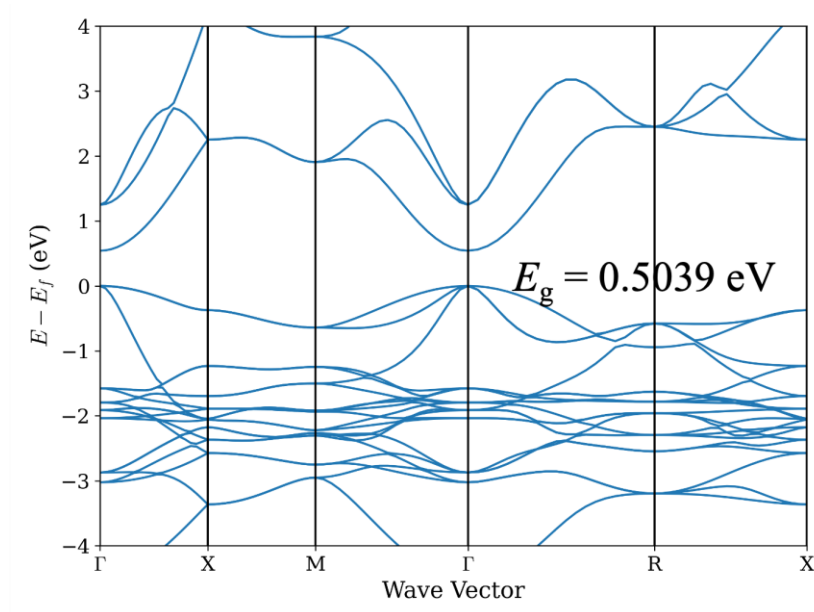


Fig. 4: Electronic band structure of pristine Cu₂O calculated using DFT.

3-2-2. Case 1: Cu₂O_{0.775}P_{0.15} (Top-ranked candidate)

The highest-ranked compound, $\text{Cu}_2\text{O}_{0.775}\text{P}_{0.15}$, was first examined. However, direct construction of this composition requires a supercell containing more than 100 atoms to satisfy the stoichiometric ratio, resulting in prohibitively high computational cost. Therefore, a slightly adjusted composition, $\text{Cu}_2\text{O}_{0.75}\text{P}_{0.167}$, was adopted, which allows construction of a 35-atom supercell while maintaining a similar substitution level. This adjusted composition was used for subsequent structural and electronic analyses. The optimized crystal structure of $\text{Cu}_2\text{O}_{0.75}\text{P}_{0.167}$ is shown in Fig. 5. To quantitatively verify the randomness of the generated SQS beyond visual inspection, we evaluated the mismatch in the pair-correlation functions using the mcsqs code. The maximum absolute difference between the SQS and an ideal random solid solution is exceptionally small for the first and second nearest-neighbor distances as shown in Table 2. This minimal mismatch mathematically demonstrates that our supercell successfully mimics an ideal random solid solution without artificial clustering. Due to the larger atomic radius of P compared to O, as well as the introduction of vacancies to maintain charge neutrality, local lattice strain is expected in the structure. During structural relaxation, atomic rearrangement occurred to minimize interatomic forces, resulting in noticeable distortion from the original Cu_2O lattice.

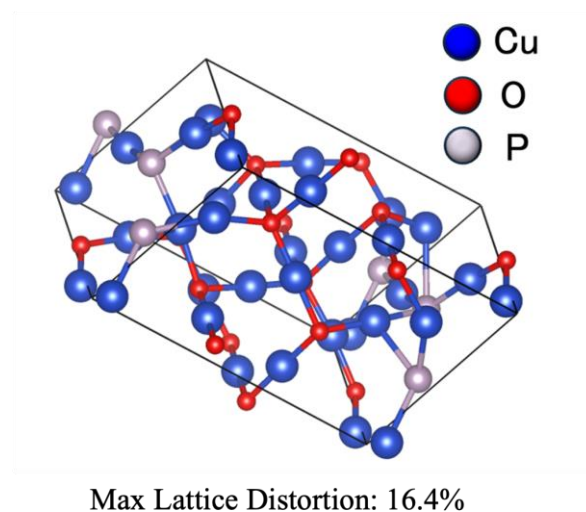


Fig. 5: Optimized crystal structure of $\text{Cu}_2\text{O}_{0.75}\text{P}_{0.167}$ generated using the SQS approach, showing the spatial distribution of substituted P atoms.

Diameter (Å)	SQS Correlation	Target Correlation	Difference
3.698	0.437500	0.390625	0.046875
3.698	0.018042	0.045106	-0.027064
3.698	0.000000	0.005208	-0.005208
3.698	0.437500	0.390625	0.046875
3.698	0.018042	0.045106	-0.027064
3.698	0.000000	0.005208	-0.005208
4.270	0.375000	0.390625	-0.015625
4.270	0.072169	0.045106	0.027063
4.270	-0.041667	0.005208	-0.046875

Table 2. Pair-correlation functions of the generated SQS supercell compared to the target ideal random solid solution.

The density of states of $\text{Cu}_2\text{O}_{0.75}\text{P}_{0.167}$ is shown in Fig. 6. A finite density of states at the Fermi level is observed, indicating that the Fermi level lies within the valence band. This behavior is characteristic of heavily doped semiconductors, where the Fermi level shifts into the band and the system exhibits degenerate or metallic-like conductivity. Such electronic characteristics are undesirable for photovoltaic applications, as they can lead to increased carrier recombination and loss of rectifying behavior. Therefore, despite its high ranking in the screening process, within the PBE framework employed in the present study, $\text{Cu}_2\text{O}_{0.775}\text{P}_{0.15}$ does not exhibit the electronic characteristics desired for top-cell absorbers in Si-based dual-junction solar cells. However, because the PBE functional is known to underestimate band gaps and may incorrectly delocalize electronic states in transition-metal oxides, the possibility that

higher-level methods such as DFT+U, HSE06, or GW could modify the predicted electronic character cannot be excluded [38–41]. Consequently, the present result should be interpreted as a qualitative assessment within the PBE framework rather than definitive evidence that the material is unsuitable for photovoltaic applications.

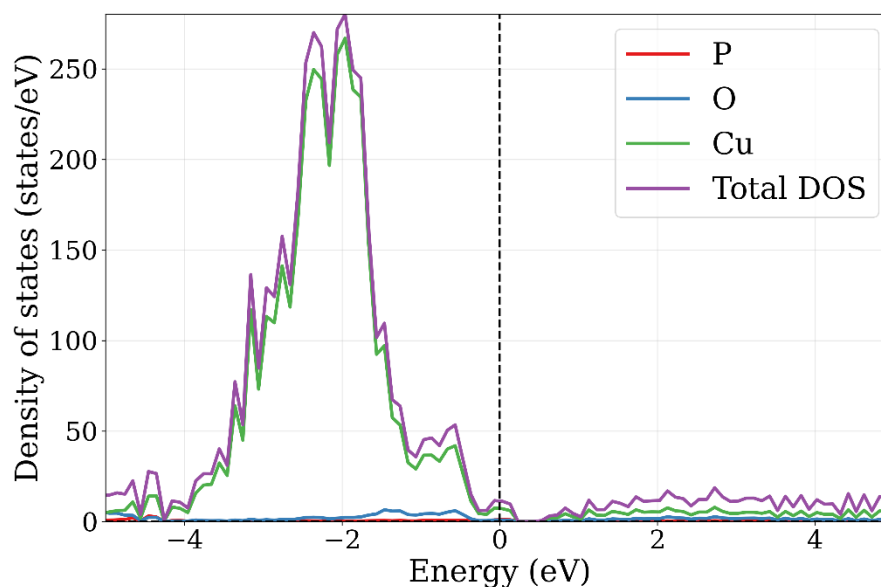


Fig. 6: Density of states of $\text{Cu}_2\text{O}_{0.75}\text{P}_{0.167}$, indicating metallic or degenerate semiconductor behavior due to high substitution levels.

3-2-3. Case 2: $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ (Second-ranked candidate)

The second-ranked compound, $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$, was further investigated in detail. The optimized crystal structure is shown in Fig. 7. As seen in the figure, the relaxed structure exhibits significant distortion from the original cubic symmetry, resulting in a parallelogram-like deformation of the lattice. This distortion is attributed to the size mismatch between Se and O atoms, where local strain induced by substitution is relieved through global lattice deformation. The density of states of $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ is shown in Fig. 8. In contrast to the P-

substituted system discussed above, the density of states at the Fermi level is zero, confirming that the compound retains semiconducting behavior with a finite band gap. The calculated band gap is 0.4874 eV, which is slightly smaller than that of pristine Cu_2O (0.5039 eV) obtained using the same computational setup. This corresponds to an approximate 3% reduction in the band gap due to Se substitution. Although it is well known that PBE-based DFT calculations systematically underestimate band gaps, the relative trends between structurally similar compounds remain meaningful. Therefore, the observed reduction in the calculated band gap suggests that the experimental band gap of $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ is likely shifted toward lower energies compared to that of pristine Cu_2O (2.17 eV [37]). The calculated DFT band gap of $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ was 0.4874 eV, slightly smaller than that of pristine Cu_2O (0.5039 eV) obtained under the same computational conditions. This small reduction suggests that Se substitution may induce a modest downward shift in the band gap within the PBE framework. However, because the difference is only 0.0165 eV and PBE substantially underestimates the absolute band gap of Cu_2O , this result should be interpreted with caution. Quantitative validation of whether Se substitution shifts the experimental band gap toward the target range for Si-based multijunction solar cells would require higher-level electronic structure calculations, such as HSE06 or GW, or experimental measurements. Therefore, $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$ should be regarded as a candidate warranting further validation rather than as a conclusively optimized top-cell absorber.

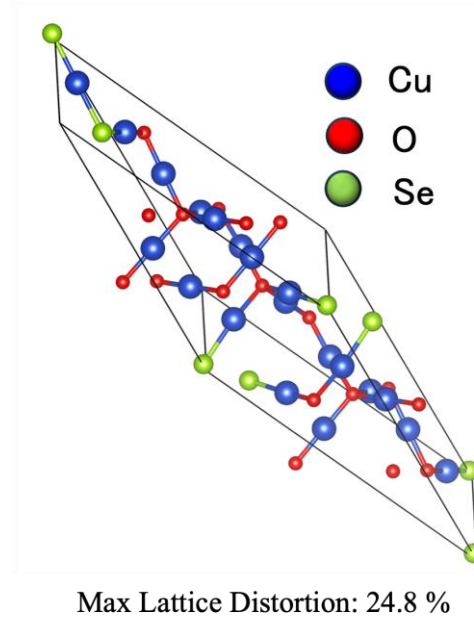


Fig. 7: Optimized crystal structure of $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$, highlighting lattice distortion induced by Se substitution.

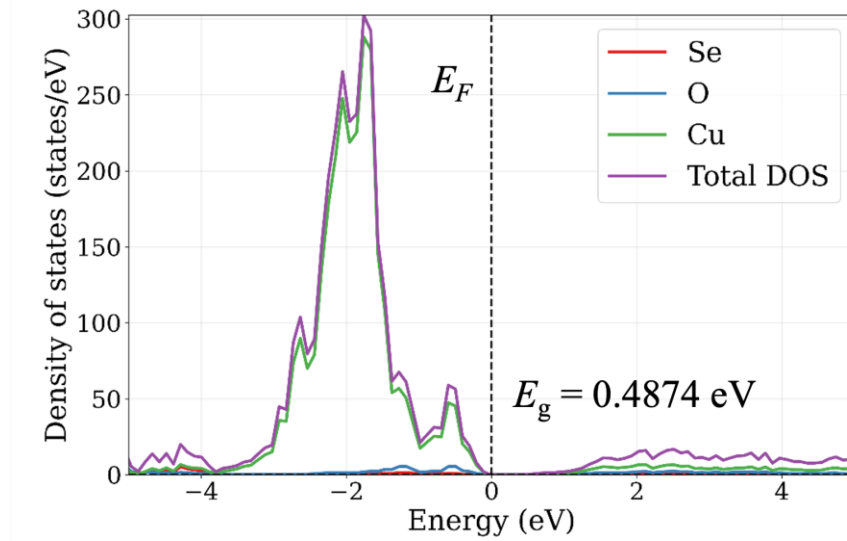


Fig. 8: Density of states of $\text{Cu}_2\text{O}_{0.80}\text{Se}_{0.20}$, showing the presence of a band gap and direct semiconducting characteristics.

It should be emphasized that the band-gap score A_j and the first-principles validation serve different purposes. A_j is based on machine learning predictions trained on corrected band-

gap values that approximate experimental band gaps, and therefore evaluates the probability that a candidate falls within the absolute target range of 1.6–1.9 eV. In contrast, the DFT calculations are not intended to provide quantitatively reliable absolute band gaps because the DFT calculation severely underestimates the band gap of Cu₂O. The DFT validation was therefore used only to assess qualitative electronic characteristics, such as whether the material remains semiconducting and whether the band gap shows a relative increase or decrease compared with pristine Cu₂O under identical computational settings. Thus, the machine learning screening and DFT validation should be regarded as complementary rather than directly equivalent evaluation steps.

The electronic band structure is shown in Fig. 9. Due to the structural distortion, strict symmetry-based high-symmetry paths would require a more complex treatment. However, for clarity and direct comparison with pristine Cu₂O, the band structure was plotted along the conventional high-symmetry path assuming an idealized lattice. The results show that both the valence band maximum and conduction band minimum are located at the Γ point, indicating a direct band gap. Direct semiconductors are advantageous for photovoltaic applications due to their high optical absorption coefficients. Furthermore, the dispersion near the valence band maximum at the Γ point is comparable to that of pristine Cu₂O (Fig. 4), suggesting that carrier transport properties are not significantly degraded by Se substitution. Nevertheless, it should be noted that the present band-structure calculations were performed for distorted SQS supercells, in which the crystal symmetry is lower than that of pristine Cu₂O. Therefore, the conventional high-symmetry k -point path adopted for visualization should be regarded as an approximate reference path. Consequently, the identification of a direct band gap should be interpreted qualitatively, and a rigorous assessment would require symmetry-adapted band unfolding or related analyses.

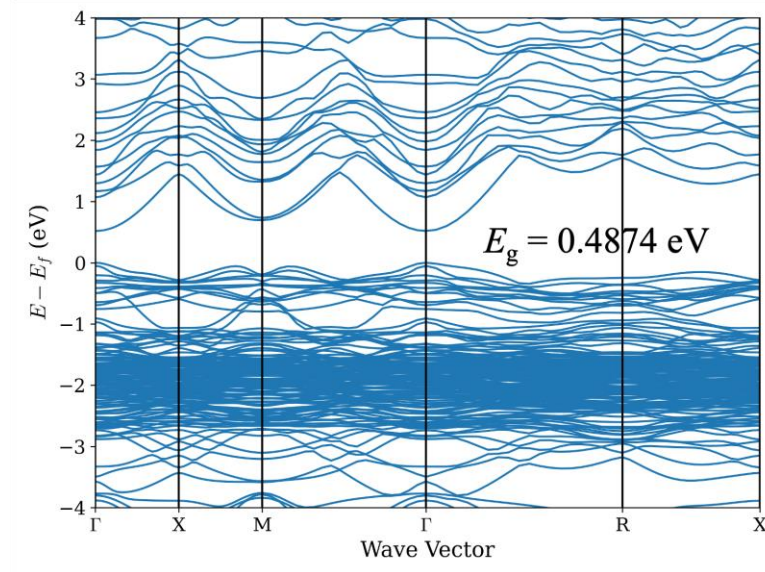


Fig. 9: Electronic band structure of Cu₂O_{0.80}Se_{0.20}, indicating a direct band gap at the Γ point.

To further assess thermodynamic stability, the formation enthalpy was calculated using:

$$\Delta H_f = \frac{E_{total} - \sum_i n_i \mu_i}{\sum_i n_i} \quad (6)$$

where E_{total} is the total energy of the supercell obtained from first-principles calculations, n_i is the number of atoms of element i , and μ_i is the reference energy per atom for each element in its stable phase. The reference energies were calculated under the same computational conditions as that for the compound search above, resulting in -3.729 , -4.930 , and -3.491 eV/atom for Cu, O, and Se, respectively. The calculated formation enthalpy of Cu₂O_{0.80}Se_{0.20} was -0.313 eV/atom, indicating favorable compound formation relative to the elemental reference states. However, a negative formation enthalpy alone does not guarantee thermodynamic stability against competing phases. According to the Cu–O–Se phase stability analysis in OQMD, the corresponding composition is predicted to decompose into Cu, Cu₂O, and Cu₃Se₂ with a decomposition enthalpy of -0.572 eV/atom [29]. Therefore, the present SQS

structure lies approximately 0.259 eV/atom above the convex hull and should be regarded as thermodynamically metastable under equilibrium conditions. Nevertheless, metastable phases may still be experimentally accessible through kinetic stabilization or non-equilibrium synthesis routes, such as low-temperature thin-film growth, rapid quenching, or strain-stabilized epitaxy [42].

3-2-4. Other Candidates

First-principles validation was also performed for all compounds ranked from third to sixth in the screening results. In addition, to assess the influence of less frequently occurring substituent elements, the highest-ranked As-substituted compound (ranked twelfth, as found in Table 1) was examined under the same computational framework. The results are summarized in Table 3. The investigated compounds can be broadly classified into three categories: (i) direct-gap semiconductors with band gaps larger than that of pristine Cu_2O , (ii) compounds exhibiting reduced band gaps but with indirect band gaps, and (iii) compounds showing metallic or degenerate electronic behavior, similar to the P-substituted system discussed earlier. None of these characteristics are suitable for the targeted application in Si-based multijunction solar cells. In particular, larger band gaps reduce spectral utilization, indirect band gaps lead to lower optical absorption efficiency, and metallic or degenerate behavior results in increased carrier recombination and loss of rectifying properties. These observations highlight that high screening scores do not necessarily guarantee desirable electronic properties, underscoring the importance of first-principles validation.

Rank	Composition	Band gap [eV]	Electronic character	Band gap trend vs Cu ₂ O
1	Cu ₂ O _{0.750} P _{0.167} [†]	—	Degenerate semiconductor	—
2	Cu ₂ O _{0.800} Se _{0.200}	0.487	Direct semiconductor	Decreased
3	Cu ₂ O _{0.750} Se _{0.250}	0.740	Direct semiconductor	Increased
4	Cu ₂ O _{0.700} P _{0.200}	—	Degenerate semiconductor	—
5	Cu ₂ O _{0.700} Se _{0.300}	0.379	Indirect semiconductor	Decreased
6	Cu ₂ O _{0.667} S _{0.333} [‡]	0.609	Direct semiconductor	Increased
12	Cu ₂ O _{0.850} As _{0.100}	—	Degenerate semiconductor	—

Table 3: Summary of first-principles validation results for candidate compounds ranked 3rd to 6th based on AQ_j , along with the highest-ranked As-substituted compound (ranked 12th).

[†] Composition adjusted from Cu₂O_{0.775}S_{0.150} to reduce computational cost.

[‡] Composition adjusted from Cu₂O_{0.650}S_{0.350} to reduce computational cost.

4. Conclusions

In this study, we explored Cu₂O-based compounds with partial substitution at oxygen sites as candidate top-cell materials for Si-based multijunction solar cells. To identify practically viable materials, we developed and applied a multi-objective, decision-aware screening framework that simultaneously considers band gap suitability, supply risk, and thermodynamic stability. Promising candidates identified through this screening were further validated using first-principles calculations. The screening results revealed that compounds with anion substitution by P, Se, and S frequently rank highly in terms of the integrated score. In particular, Se substitution was found to be especially effective in balancing the competing requirements. Among the candidates, Cu₂O_{0.80}Se_{0.20} was identified as a promising material, exhibiting a favorable combination of reduced supply risk, appropriate band gap tuning, and thermodynamic stability. The proposed framework is effective in narrowing a large candidate space and prioritizing compounds for subsequent first-principles validation. However, the present results also demonstrate that high machine-learning scores do not necessarily guarantee desirable electronic properties after DFT validation, highlighting the importance of integrating data-driven screening with physics-based verification. At the same time, first-principles validation revealed that even top-ranked candidates do not always retain the desired semiconducting properties. In particular, substitutions involving elements with significantly different atomic radii or valence electron configurations from the host lattice tend to induce metallic behavior, likely due to impurity state formation and structural distortions. This finding highlights an important limitation of purely data-driven screening approaches and underscores the necessity of incorporating physical consistency—such as structural compatibility with the host lattice—into materials design. More broadly, the present study suggests that integrating data-driven screening with physics-based validation provides a robust pathway for reliable and

application-oriented materials discovery. Future work should incorporate additional descriptors such as carrier mobility and structural compatibility, as well as higher-level electronic structure methods (e.g., HSE06, DFT+U, or GW) for more accurate validation. Future extensions of the present framework may benefit from the integration of emerging machine learning interatomic potentials (e.g., MACE and UMA) with crystal-structure prediction techniques, enabling efficient identification of low-energy polymorphs, phase stability, and dynamical properties beyond the parent Cu_2O lattice [43]. Furthermore, integration with experimental synthesis and feedback loops will be essential to establish a closed-loop materials discovery framework.

Declaration of Competing Interest

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

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This study was financially supported, in part, by the grant number 23K03942, the Japan Society for the Promotion of Science.

References

- [1] J. Zhao, A. Wang, M. A. Green, and F. Ferrazza, "19.8% efficient “honeycomb” textured multicrystalline and 24.4% monocrystalline silicon solar cells," *Appl. Phys. Lett.* **73**, 1991–1993 (1998).
<https://doi.org/10.1063/1.122345>
- [2] C. Battaglia, A. Cuevasb, and S. De Wolf, "High-efficiency crystalline silicon solar cells: Status and perspectives," *Energy Environ. Sci.* **9**, 1552–1576 (2016).
<https://doi.org/10.1039/c5ee03380b>
- [3] K. Yoshikawa, H. Kawasaki, W. Yoshida, T. Irie, K. Konishi, K. Nakano, T. Uto, D. Adachi, M. Kanematsu, H. Uzu, and K. Yamamoto, "Silicon heterojunction solar cell with interdigitated back contacts for a photoconversion efficiency over 26%," *Nat. Energy* **2**, 17032 (2017).
<https://doi.org/10.1038/nenergy.2017.32>
- [4] J. M. Olson, S. R. Kurtz, A. E. Kibbler, and P. Faine, "A 27.3% efficient Ga_{0.5}In_{0.5}P/GaAs tandem solar cell," *Appl. Phys. Lett.* **56**, 623–625 (1990).
<https://doi.org/10.1063/1.102717>
- [5] K. Tanabe, K. Watanabe, and Y. Arakawa, "III–V/Si hybrid photonic devices by direct fusion bonding," *Sci. Rep.* **2**, 349 (2012).
<https://doi.org/10.1038/srep00349>
- [6] J. F. Geisz, R. M. France, K. L. Schulte, M. A. Steiner, A. G. Norman, H. L. Guthrey, M. R. Young, T. Song, and T. Moriarty, "Six-junction III–V solar cells with 47.1% conversion efficiency under 143 Suns concentration," *Nat. Energy* **5**, 326–335 (2020).
<https://doi.org/10.1038/s41560-020-0598-5>
- [7] P. Schygulla, R. Müller, D. Lackner, O. Höhn, H. Hauser, B. Bläsi, F. Predan, J. Benick, M. Hermle, S. W. Glunz, F. Dimroth, "Two-terminal III–V//Si triple-junction solar cell with power conversion efficiency of 35.9 % at AM1.5g," *Prog. Photovolt.* **30**, 869–879 (2022).
<https://doi.org/10.1002/pip.3503>
- [8] K. Tanabe, A. Fontcuberta i Morral, H. A. Atwater, D. J. Aiken, and M. W. Wanlass, "Direct-bonded GaAs/InGaAs tandem solar cell," *Appl. Phys. Lett.* **89**, 102106 (2006).

<https://doi.org/10.1063/1.2347280>

[9] R. R. King, D. C. Law, K. M. Edmondson, C. M. Fetzer, G. S. Kinsey, H. Yoon, R. A. Sherif, and N. H. Karam, "40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells," *Appl. Phys. Lett.* **90**, 183516 (2007).

<https://doi.org/10.1063/1.2734507>

[10] J. D. Major, R. E. Treharne, L. J. Phillips, and K. Durose, "A low-cost non-toxic post-growth activation step for CdTe solar cells," *Nature* **511**, 334–337 (2014).

<https://doi.org/10.1038/nature13435>

[11] S. Ishizuka, S. Kato, Y. Okamoto, and K. Akimoto, "Control of hole carrier density of polycrystalline Cu₂O thin films by Si doping," *Appl. Phys. Lett.* **80**, 950–952 (2002).

<https://doi.org/10.1063/1.1448398>

[12] D. S. Darvish and H. A. Atwater, "Epitaxial growth of Cu₂O and ZnO/Cu₂O thin films on MgO by plasma-assisted molecular beam epitaxy," *J. Cryst. Growth* **319**, 39–43 (2011).

<https://doi.org/10.1016/j.jcrysgro.2011.01.071>

[13] Y. Nishi, T. Miyata, J. Nomoto, and T. Minami, "Influence of Cu₂O surface treatment on the photovoltaic properties of Al-doped ZnO/Cu₂O solar cells," *Thin Solid Films* **520**, 3819–3822 (2012).

<https://doi.org/10.1016/j.tsf.2011.08.032>

[14] K. Fujimoto, T. Oku, and T. Akiyama, "Fabrication and characterization of ZnO/Cu₂O solar cells prepared by electrodeposition," *Appl. Phys. Express* **6**, 086503 (2013).

<https://doi.org/10.7567/APEX.6.086503>

[15] S. T. Omelchenko, Y. Tolstova, H. A. Atwater, and N. S. Lewis, "Excitonic effects in emerging photovoltaic materials: A case study in Cu₂O," *ACS Energy Lett.* **2**, 431–437 (2017).

<https://doi.org/10.1021/acsenergylett.6b00704>

[16] V. Stevanović, A. Zakutayev, and S. Lany, "Composition dependence of the band gap and doping in Cu₂O-based alloys as predicted by an extension of the dilute-defect model," *Phys. Rev. Appl.* **2**, 044005 (2014).

<https://doi.org/10.1103/physrevapplied.2.044005>

[17] A. Zakutayev, V. Stevanovic, and S. Lany, "Non-equilibrium alloying controls

optoelectronic properties in Cu₂O thin films for photovoltaic absorber applications," Appl. Phys. Lett. **106**, 123903 (2015).

<https://doi.org/10.1063/1.4914974>

[18] G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, and T. Lookman, "Machine learning bandgaps of double perovskites," Sci. Rep. **6**, 19375 (2016).

<https://doi.org/10.1038/srep19375>

[19] P. Avery, X. Wang, C. Oses, E. Gossett, D. M. Proserpio, C. Toher, S. Curtarolo, and E. Zurek, "Predicting superhard materials via a machine learning informed evolutionary structure search," npj Comput. Mater. **5**, 89 (2019).

<https://doi.org/10.1038/s41524-019-0226-8>

[20] K. Tanabe, " Machine-learning-based composition analysis of the stability of V–Cr–Ti alloys," J. Nucl. Eng. **4**, 317–322 (2023).

<https://doi.org/10.3390/jne4020024>

[21] F. Dinic, I. Neporozhnyi, and O. Voznyy, "Machine learning models for the discovery of direct band gap materials for light emission and photovoltaics," Comput. Mater. Sci. **231**, 112580 (2024).

<https://doi.org/10.1016/j.commatsci.2023.112580>

[22] T. Masuda and K. Tanabe, "Explainable artificial intelligence for machine learning prediction of bandgap energies," J. Appl. Phys. **136**, 175703 (2024).

<https://doi.org/10.1063/5.0226151>

[23] L. Kavalsky, V. I. Hegde, B. Meredig, and V. Viswanathan, "A multiobjective closed-loop approach towards autonomous discovery of electrocatalysts for nitrogen reduction," Digit. Discov. **3**, 999–1010 (2024).

<https://doi.org/10.1039/D3DD00244F>

[24] T. Kono, S. Miyamoto, T. Masuda, and K. Tanabe, Comput. Mater. Sci. **263**, 114463 (2026).

<https://doi.org/10.1016/j.commatsci.2025.114463>

[25] S. R. Kurtz, P. Faine, and J. M. Olson, "Modeling of two-junction, series-connected tandem solar cells using top-cell thickness as an adjustable parameter," J. Appl. Phys. **68**, 1890–1895 (1990).

<https://doi.org/10.1063/1.347177>

- [26] R. Garrison and R. Kleiman, "Higher efficiency tandem solar cells through composite-cell current matching," *Opt. Express* **27**, A543–A571 (2019).
<https://doi.org/10.1364/OE.27.00A543>
- [27] J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard, "Resolution of the band gap prediction problem for materials design," *J. Phys. Chem. Lett.* **7**, 1198–1203 (2016).
<https://doi.org/10.1021/acs.jpclett.5b02870>
- [28] M. W. Gaultois, T. D. Sparks, C. K. H. Borg, R. Seshadri, W. D. Bonificio, and D. R. Clarke, "Data-driven review of thermoelectric materials: Performance and resource considerations," *Chem. Mater.* **25**, 2911–2920 (2013).
<https://doi.org/10.1021/cm400893e>
- [29] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, "The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies," *npj Comput. Mater.* **1**, 15010 (2015).
<https://doi.org/10.1038/npjcompumats.2015.10>
- [30] D. Jha, L. Ward, A. Paul, W. -K. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, "ElemNet: Deep learning the chemistry of materials from only elemental composition," *Sci. Rep.* **8**, 17593 (2018).
<https://doi.org/10.1038/s41598-018-35934-y>
- [31] R. E. A. Goodall and A. A. Lee, "Predicting materials properties without crystal structure: deep representation learning from stoichiometry," *Nat. Commun.* **11**, 6280 (2020).
<https://doi.org/10.1038/s41467-020-19964-7>
- [32] A. van de Walle, P. Tiwary, M. de Jong, D. L. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L. -Q. Chen, and Z. -K. Liu, "Efficient stochastic generation of special quasirandom structures," *Calphad* **42**, 13–18 (2013).
<https://doi.org/10.1016/j.calphad.2013.06.006>
- [33] G. Kresse and J. Hafner, "Ab initio molecular dynamics for liquid metals," *Phys. Rev. B.* **47**, 558–561 (1993).
<https://doi.org/10.1103/PhysRevB.47.558>
- [34] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy

calculations using a plane-wave basis set," Phys. Rev. B **54**, 11169–11186 (1996).
<https://doi.org/10.1103/PhysRevB.54.11169>

[35] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," Phys. Rev. B **59**, 1758–1775 (1999).
<https://doi.org/10.1103/PhysRevB.59.1758>

[36] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865–3868 (1996).
<https://doi.org/10.1103/PhysRevLett.77.3865>

[37] A. Önsten, M. Månsson, T. Claesson, T. Muro, T. Matsushita, T. Nakamura, T. Kinoshita, U.O. Karlsson, and O. Tjernberg, "Probing the valence band structure of Cu₂O using high-energy angle-resolved photoelectron spectroscopy," Phys. Rev. B **76**, 115127 (2007).
<https://doi.org/10.1103/physrevb.76.115127>

[38] D. O. Scanlon and G. W. Watson, "Understanding the *p*-type defect chemistry of CuCrO₂," J. Mater. Chem. **21**, 3655–3663 (2011).
<https://doi.org/10.1039/C0JM03852K>

[39] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, "Quasiparticle band structure based on a generalized Kohn-Sham scheme," Phys. Rev. B **76**, 115109 (2007).
<https://doi.org/10.1103/PhysRevB.76.115109>

[40] J. Robertson and S. J. Clark, "Limits to doping in oxides," Phys. Rev. B **83**, 075205 (2011).
<https://doi.org/10.1103/10.1103/PhysRevB.83.075205>

[41] M. M. Fadlallah, U. Eckern, and U. Schwingenschlögl, "Defect engineering of the electronic transport through cuprous oxide interlayers," Sci. Rep. **6**, 27049 (2016).
<https://doi.org/10.1038/srep27049>

[42] W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, and G. Ceder, "The thermodynamic scale of inorganic crystalline metastability," Sci. Adv. **2**, e1600225 (2016).
<https://doi.org/10.1126/sciadv.1600225>

[43] R. A. Gouvêa and G. Rignanese, "VibroML: An automated toolkit for high-throughput vibrational analysis and dynamic instability remediation of crystalline materials using machine-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0339444

learned potentials," arXiv:2604.27685 (2026).
<https://doi.org/10.48550/arXiv.2604.27685>