# MTTB: End-to-End Framework for Oxide TFT Channel Guidelines Beyond-IGZO

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

As AMOLED displays move toward higher resolution and lower power operation, next-generation oxide semiconductors are needed to achieve high-performance, reliable backplane TFTs. However, existing ML-based screening has two limitations. It can distort design decisions by collapsing conflicts among multiple material descriptors into a single objective function. It also lacks an interpretable link between material-level predictions and device-level performance. To address these issues, we propose a four-stage hybrid framework, *MTTB* (Materials-to-TFT Bridge). (1) **AI-based property prediction**: rapidly infers material descriptors (band gap $G_x$, inverse spatial wavefunction overlaps $ISWO_C$ and $ISWO_V$, and formation energy $E_f$) from composition vectors. (2) **Pareto-based multi-objective composition search**: compresses promising candidates while preserving trade-offs among material descriptors. (3) **Candidate selection strategy**: selects final candidate compositions based on performance–stability trade-offs and target priorities. (4) **Physics-Inference Module (PIM)**: maps predicted properties to device-physics variables to efficiently infer TCAD-equivalent device metrics. Experimentally, global Pareto-knee candidates converge to Al–Mg-centered compositions under high-bandgap conditions. Composition-sensitivity analysis and PIM-based device simulations confirm the potential for balanced performance improvements over IGZO. Finally, we propose Al–Mg-based (beyond-IGZO) oxide archetypes and composition candidates that can replace IGZO.

## 1 Introduction

AMOLED (Active-Matrix Organic Light-Emitting Diode) displays offer high contrast and fast response. They are widely used as a core display technology in electronic devices, including smartphones (1). As demand increases for higher resolution and lower-power operation, requirements for the electrical performance and stability of AMOLED driving devices continue to intensify (2). Meeting these requirements requires thin-film transistors (TFTs) in the AMOLED backplane to provide both sufficient current-driving capability and long-term reliability (3).

Oxide TFTs exhibit higher carrier mobility and lower leakage current than amorphous silicon (a-Si) TFTs. Among them, indium–gallium–zinc oxide (IGZO) has been successfully commercialized and is now the most widely used representative material (4; 5). However, continued improvements in display performance require exploring multicomponent oxide materials beyond IGZO (5). The key challenge is the exponential growth of the design space as elemental combinations diversify. Conventional experiment-driven approaches alone are insufficient for identifying optimal compositions (6). Accordingly, high-speed screening frameworks that combine material-descriptor calculations (bandgap, ISWO, $E_f$) with machine learning (ML), such as the work of Kim *et al.* (2022) (14), have emerged as a key methodology for accelerating next-generation materials discovery.

Submitted to 1st 2026AI Co-Scientist Challange Korea. Do not distribute.

However, while these approaches help reduce the composition space, they have two fundamental limitations when linking predicted material properties to TFT device performance.

First, many ML-based screening methods integrate multiple material descriptors into a single objective function. This can fail to capture intrinsic physical trade-offs among bandgap, formation energy, and electronic localization (e.g., ISWO). For example, low formation energy suggests thermodynamic stability. However, if it accompanies excessive oxygen vacancies, it can reduce the bandgap and destabilize the threshold voltage (9). Collapsing multivariate trade-offs into a single scalar objective can prematurely exclude physically plausible compositions or yield unrealistic optima. Practical composition exploration therefore requires a multi-objective optimization framework that preserves balance among multiple material descriptors while refining the design space (10).

Second, many ML frameworks stop at composition or property prediction. Even when they predict device performance, they often rely on black-box models that do not explicitly reveal the underlying physical mechanisms (7). Commercial TCAD simulations can mitigate this issue, but their high computational cost limits repeated evaluation of large candidate sets (8). Some studies directly learn correlations between composition and device performance, but they do not sufficiently provide an interpretable link describing how changes in electronic structure map to device-physics variables (device parameters), such as carrier concentration or trap density (8).

The main contributions of this paper are summarized as follows.

- **MTTB framework**: We propose a four-stage hybrid framework, *MTTB* (Materials-to-TFT Bridge). It consists of (i) **AI-based material prediction**, (ii) Pareto-based multi-objective composition search, (iii) Candidate selection stratgey, and (iv) a Physics-inference module.
- **Addressing two fundamental limitations of prior ML-based approaches**: We address two limitations of prior ML-based screening studies. First, collapsing conflicting relationships among multiple material descriptors into a single objective can distort design decisions. We mitigate this using Pareto-based multi-objective optimization and decision rules. Second, many studies do not provide an explicit link between material-level predictions and device-level performance. We address this via the PIM.
- **Actionable beyond-IGZO candidates**: We apply MTTB to an Al–Mg-centered multi-component oxide composition space. We derive IGZO-replacement compositions that globally balance trade-offs among bandgap, mobility, and stability. We also provide specific next-generation (beyond-IGZO) compositions for follow-up experimental validation.

# 2 Related work

**ML-based discovery of amorphous oxide materials** van Setten *et al.* (14) proposed a framework that combines HT-DFT with support vector regression (SVR) models. It efficiently screens the composition space of amorphous multicomponent oxides. They learned composition–property relationships from computed data for primary and binary oxides. They then generalized the models to multicomponent (multinary) systems. This enabled prediction of key metrics such as bandgap and formation energy. They also prioritized promising composition regions to reduce search cost. They further demonstrated that the framework can propose new candidates. As an IGZO replacement candidate, they suggested Zn–Mg–Al–O (ZMAO). However, the approach has limitations. It is not an end-to-end workflow from material-property prediction to device-performance validation. It does not explicitly capture trade-offs among material descriptors because it relies on single-objective optimization. It also remains limited in scalability for large-batch parallel processing over large composition spaces.

# 3 Preliminary

## 3.1 Material Descriptors

To link atomic-scale electronic structure to device performance, we define four material descriptors, $\mathbf{x} = [G_x, ISWO_C, ISWO_V, E_f]$. The bandgap $G_x$ is defined as $G_x = E_{CBM} - E_{VBM}$ and is closely related to leakage current and charge-injection barriers; a larger $G_x$ is favorable for suppressing off-state leakage and improving visible-light transmittance.

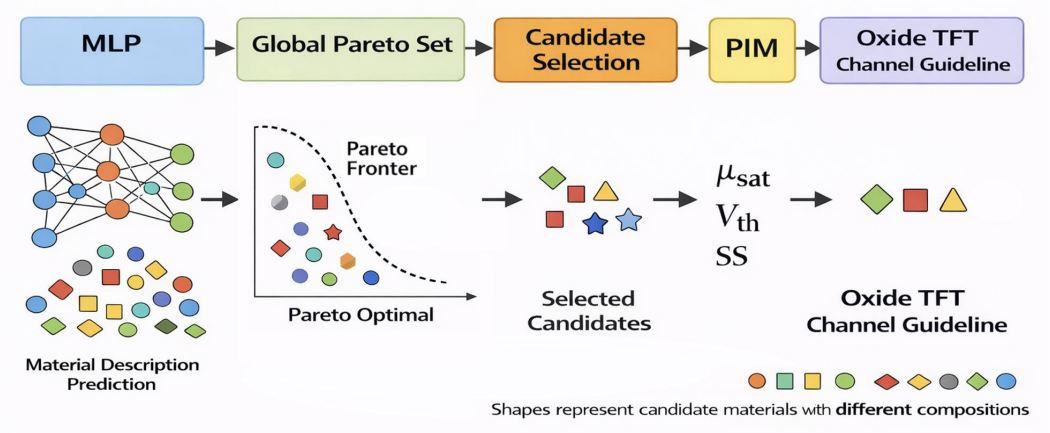



Figure 1: Overview of the MTTB framework.

Inverse Spatial Wavefunction Overlap (ISWO) is defined as the inverse of the orbital overlap integral $S$ ($ISWO \propto S^{-1}$). A smaller $ISWO_C$ indicates greater delocalization of conduction-band electron wavefunctions and thus higher electron mobility, whereas a larger $ISWO_V$ suppresses hole transport in the valence band and is favorable for n-type behavior.

The formation energy $E_f$ is defined as the total energy difference normalized by the number of oxygen atoms, $E_f = \Delta E_{tot}/N_O$, and reflects thermodynamic stability. Values lower than or comparable to IGZO indicate suppressed phase separation and reduced defect formation during processing.

To assess the de-mixing tendency of mixed amorphous oxides, we adopt the formation-energy-based stability metric $E_{\text{hull}}$ proposed by van Setten *et al.* (14). Here, $E_{\text{hull}} < 0$ indicates that the mixed state is thermodynamically more stable than separation into constituent single oxides. In this study, $E_{\text{hull}}$ is used only as a post hoc stability filter for Pareto-selected candidates.

### 3.2 TFT Performance Metrics

We summarize TFT electrical behavior using three metrics: mobility ($\mu_{sat}$), threshold voltage ($V_{th}$), and subthreshold swing ($SS$). $\mu_{sat}$ is the effective mobility extracted from the current characteristics in the saturation regime. It directly affects switching speed and on-current. $V_{th}$ is the gate voltage at which the channel forms and the drain current begins to increase substantially. It is a key parameter for low-power operation. $SS$ is the gate-voltage change required for the drain current to increase by one decade in the subthreshold regime. It is calculated as $SS = \partial V_{GS}/\partial(\log I_{DS})$.

## 4 Method

In this section, we present the MTTB framework, which integrates AI-based material property prediction, Pareto-based multi-objective screening, and physics-informed device-level inference. An overview of the framework is shown in Fig. 1.

### 4.1 AI-based Material Property Prediction

#### 4.1.1 Dataset Description

We used the amorphous oxide dataset reported in prior work (14), which comprises amorphous oxide properties computed using high-precision DFT. Amorphous structures were generated from local bonding motifs derived from unary ($X$–O) and binary ($X$–$Y$–O) crystalline structures in the Materials Project, with strictly stoichiometric compositions to exclude extrinsic n-type doping. Each unit cell contained approximately 200 atoms, and DFT calculations were performed on 10 independent amorphous configurations per composition to ensure statistical robustness. The dataset covers amorphous oxides composed of 12 metal elements (Ag, Al, Cd, Ga, In, Mg, Sb, Si, Sn, Ti, Zn, and Zr).

#### 4.1.2 MLP-based predictor

To address the limited expressiveness of the SVR model used in prior work, we adopt a two-layer multi-layer perceptron (MLP) to capture nonlinear relationships in high-dimensional composition space. The input is a normalized metal composition vector, and the model is designed for rapid inference without requiring explicit atomic-structure information.

Given the heterogeneous distributions and learning difficulties of different target properties, we trained independent MLPs for each property. The models predict four descriptors: bandgap ($E_g$), formation energy ($E_f$), $ISWO_C$, and $ISWO_V$. This design enables fast batch inference over large composition sets, significantly improving the efficiency of subsequent multi-objective screening.

### 4.2 Pareto-based Multi-objective Composition Optimization

#### 4.2.1 Rationale for Pareto Optimization in Materials Design

A weight-based single-objective function that compresses multiple properties into a single scalar can simplify the search. However, such an approach inherently fails to preserve explicit trade-offs between objectives (14). To address this, we treat each property as an independent objective and adopt a multi-objective optimization strategy to derive the Pareto-optimal set.

A Pareto-optimal set consists of solutions for which no objective can be improved without degrading at least one other. The resulting Pareto front delineates the boundary of potentially optimal candidate compositions while preserving the intrinsic trade-off structure among properties. This approach is particularly effective for materials design problems that must simultaneously consider competing requirements, such as electronic performance and stability (10; 23; 24).

#### 4.2.2 Search Space Definition and Composition Sampling

We defined the composition search space by generating all binary ($X$–$Y$–O) and ternary ($X$–$Y$–$Z$–O) oxide combinations that can be formed by selecting two or three elements from the 12-element set. Within each system, candidate compositions were generated via Dirichlet-distribution-based sampling. We used 30 compositions per ternary system and 15 compositions per binary system. To reflect practical controllability in deposition processes, we excluded trace components with an atomic fraction below 0.01% during sampling.

#### 4.2.3 Objective Function Formulation

ll objectives were formulated as minimization problems. For each candidate composition, we defined four objectives:

$$f_1 = (E_{gap} - (E_{gap,IGZO} + \Delta E_{target}))^2, \quad f_2 = ISWO_C, \quad f_3 = -ISWO_V, \quad f_4 = E_f.$$

We set $\Delta E_{target} \in \{0.0, 0.25, 0.50, 0.75, 1.00\}$. We considered $\Delta E_{target} \in \{0.0, 0.25, 0.50, 0.75, 1.00\}$. These objectives jointly capture electronic properties, carrier transport, and thermodynamic stability. The Pareto front of non-dominated solutions was then computed, enabling systematic selection of candidates that balance performance and stability beyond single-property criteria.

#### 4.2.4 Hierarchical Pareto Screening and Global Front Construction

We first computed local Pareto fronts independently for each binary and ternary composition system. System-level Pareto candidates were then aggregated across all binary systems and across all ternary systems to construct order-wise global Pareto fronts.

Next, a formation-energy-based de-mixing stability metric ($E_{hull}$) was applied to the binary and ternary Pareto candidate sets. Only compositions with no tendency toward phase separation in the mixed state ($Stability < 0$) were retained. Composition ratios were rounded to two decimal places to remove duplicates, and for identical chemical formulas, the candidate with the lowest de-mixing energy was kept. Finally, the filtered binary and ternary candidates were merged into a single set, and Pareto optimization was performed once more to obtain the final global Pareto set spanning the full composition space. From the global Pareto set (822 compositions), we formed candidate sets under three distinct criteria.

### 4.3 Candidate Selection Strategy from the Global Pareto Set

**Property-prioritized Candidate Buckets** Because industrial requirements do not reduce to a single property, we partitioned the merged global Pareto set into three candidate buckets reflecting different property priorities. The high-mobility bucket selects the top five compositions with the smallest $ISWO_C$ under target bandgap and stability constraints, whereas the high-stability bucket selects the top five compositions with the lowest formation energy $E_f$ while maintaining charge-transport properties. The balanced bucket applies a multi-criteria ranking that sequentially considers bandgap proximity, formation energy, and $ISWO_C$. This analysis serves as a supplementary tool for comparing candidate distributions under different property-driven design scenarios.

**Bandgap-conditioned Knee Point Selection** Next, for each target bandgap, we identified curvature-based knee points on the corresponding Pareto front to extract compromise candidates. Four objective functions—bandgap proximity, mobility, reliability, and formation energy—were normalized to the $[0, 1]$ range, and the sacrifice rate between adjacent candidates was computed. The knee point was defined as the candidate with the maximum sacrifice rate, characterizing locally optimal solutions for each bandgap condition.

**Global Knee-based Final Selection** Finally, in the normalized four-dimensional objective space that integrates all bandgap targets, we computed the Euclidean distance to the utopia point $(0, 0, 0, 0)$. We define the 20 compositions with the smallest global distances as the global-knee set. These candidates are considered to provide the most effective global compromise among bandgap, mobility, and stability. In this study, we propose the global-knee set as the primary target for subsequent device-physics simulations and experimental validation.

### 4.4 Physics-Inference Module (PIM)

We propose a Physics-Inference Module (PIM) that estimates device-level TCAD inputs and TFT electrical metrics from AI-predicted DFT material descriptors. Key features of the PIM include:

- **Relative anchor**: IGZO DFT predictions are used as a reference, and the target composition is interpreted through relative deviations, constraining inferred parameters to physically plausible ranges.
- **Physics-grounded linkage**: DFT descriptors are mapped to device parameters and subsequently to $I$–$V_g$ characteristics to extract $\mu_{sat}$, $V_{th}$, and $SS$.
- **Computational efficiency**: The physics-based mapping bypasses iterative TCAD simulations, enabling rapid evaluation of large candidate sets.

Specifically, based on semiconductor physics, the bandgap $G_x$ is mapped to a variable related to the intrinsic carrier concentration (16). The relative change in $ISWO_C$ is used as a proxy for changes in electron transport and mobility. Deviations in $ISWO_V$ and $E_f$ are converted into parameters that reflect changes in the distribution of defect/trap states. These parameters are reflected in the trap density and threshold-voltage variation. The estimated physical variables are then fed into an exponential model in the subthreshold regime and a gradual-channel-approximation model in strong inversion. This generates TFT transfer characteristics ($I$–$V_g$) (16; 17). Finally, we apply a parameter-extraction algorithm to the generated $I$–$V_g$ to extract $\mu_{sat}$, $V_{th}$, and $SS$ (22). Implementation details are provided in our released code.

**Scope and limitations of the PIM.** The PIM serves as a physics-informed surrogate for device-level evaluation, rather than a replacement for full TCAD simulations. Its mappings are based on simplified, relative, and IGZO-anchored relationships, and therefore do not capture detailed electrostatics, spatial defect distributions, or process-dependent variations. Accordingly, the PIM is intended for efficient comparative screening and trend analysis, while final quantitative validation requires rigorous TCAD or experimental verification.

## 5 Experiment

In this section, we evaluate the proposed framework by addressing the following research questions:
**RQ1.** Does the model accurately generalize across diverse oxide compositions?

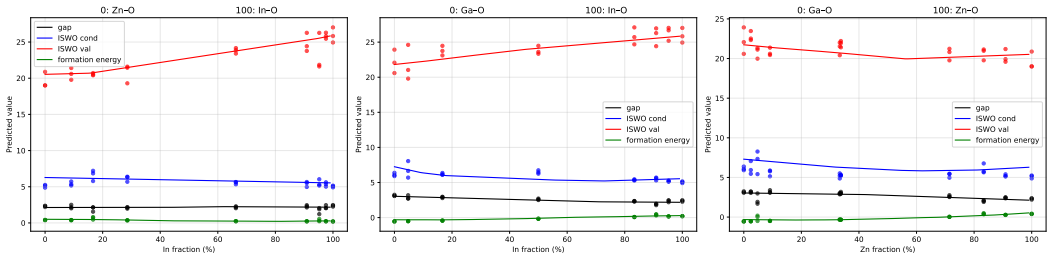



Figure 2: Validation of physical trends.Solid lines and scatter points denote MLP predictions and DFT calculations, respectively, for the In-Ga-Zn-O system.

Table 1: Comparison of property prediction performance across models.

| Target ($y$) | Gaussian | | | van Setten *et al.* (14) | | | Ours | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAE | Corr | $R^2$ | MAE | Corr | $R^2$ | MAE | Corr | $R^2$ |
| $G_x$ | 0.216 | **0.962** | -0.214 | 0.256 | 0.369 | -0.594 | **0.186** | 0.379 | **-0.037** |
| $E_f$ | 0.944 | 0.351 | -16.326 | **0.019** | **0.995** | **0.986** | 0.026 | 0.994 | 0.976 |
| $ISWO_V$ | 2.599 | -0.143 | -14.712 | 1.160 | -0.134 | -2.295 | **1.132** | **0.462** | **0.183** |
| $ISWO_C$ | 1.130 | 0.455 | **0.201** | 1.184 | **0.457** | 0.056 | **0.966** | 0.137 | -0.969 |
| Avg. | 1.222 | 0.406 | -7.763 | 0.654 | 0.422 | -0.461 | **0.577** | **0.493** | **0.038** |

**RQ2.** How do optimal oxide systems shift across varying bandgap requirements?

**RQ3.** How can Pareto-optimal candidates be categorized for device requirements?

**RQ4.** What is the ultimate material archetype for high-performance oxide semiconductors?

**RQ5.** How robust is the Al–Mg archetype against composition variations?

**RQ6.** Do AI-designed oxide TFT channels deliver gains under TCAD validation?

### 5.1 Experimental Settings

The framework was implemented in Python 3.12 (18) using PyTorch 2.9 (19) and executed on Google Colab (20) with an NVIDIA T4 GPU (21). PIM-based TCAD-equivalent analyses (Section 4.4) were performed under the following device and operating conditions. All evaluations were conducted at room temperature ($T \approx 300\,\mathrm{K}$), consistent with standard oxide TFT characterization protocols (4; 5; 3). The channel geometry was set to $W = 50\,\mu\mathrm{m}$ and $L = 10\,\mu\mathrm{m}$, with a gate insulator thickness of $t_{ox} = 20\,\mathrm{nm}$. The relative permittivity of the gate insulator was fixed at $\varepsilon_{r,ox} = 10$, and the oxide capacitance was calculated using the standard MOS expression, $C_{ox} = \varepsilon_{r,ox}\varepsilon_0/t_{ox}$ (16; 17). The implementation of the proposed MTTB framework is available at `https://anonymous.4open.science/r/MTTB-228F/`.

### 5.2 Generalization Performance of the Prediction Model

We evaluated the model's generalization performance using quantitative metrics (Table 1) and physical consistency. On the unseen Mg–Zn–Al oxide dataset, the proposed model achieved the lowest error among the compared methods (mean MAE = 0.577) and attained positive values for $ISWO_V$ and the overall mean $R^2$, where prior models failed. This demonstrates reliable prediction performance for complex multicomponent systems. In a composition sweep of the In–Ga–Zn (IGZO) system (Fig. 2), the model accurately reproduced key physical trends: $ISWO_C$ improved with increasing In content, while the bandgap ($E_g$) widened with increasing Ga content. Smooth interpolation in data-sparse regions indicates that the model stably captures the underlying physical landscape shaped by inter-element correlations, supporting its qualitative reliability for multicomponent composition exploration.

Table 2: Summary of dominant systems and predicted property ranges by target $\Delta E_g$.

| Target $\Delta E_g$ (eV) | Representative Element | $E_g$ (eV) | $ISWO_C$ | Stability |
|---|---|---|---|---|
| 0.00 | `Ag, Zr, Ti` | 0.14–0.56 | 16.5–18.4 | $\approx -3.0$ |
| 0.25 | `Mg, Ti, Al` | 2.75–2.81 | 29.8–31.2 | $\approx -2.8$ |
| 0.50 | `Zr, Cd, In, Sn` | 2.96–3.03 | 21.1–24.5 | $\approx -3.4$ |
| 0.75 | `Al, Ga, Zn, Zr` | 3.26–3.30 | 17.1–21.3 | -1.8–2.8 |
| 1.00 | `Al, Mg, Sb` | 3.43–3.68 | 8.0–9.3 | $\approx -2.1$ |

Table 3: Statistical summary of material properties by performance bucket.

| Bucket | $\Delta E_g$ (eV) | | $ISWO_C$ | | $E_f$ | | Stability | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std | Mean | Std | Mean | Std | Mean | Std |
| `balanced` | 3.02 | 0.36 | 18.80 | 8.67 | -2.41 | 1.09 | -0.39 | 0.23 |
| `high_mobility` | 2.90 | 0.44 | 6.17 | 0.73 | -0.77 | 0.51 | -0.34 | 0.24 |
| `high_stability` | 3.03 | 0.28 | 27.88 | 1.81 | -3.99 | 0.19 | -0.24 | 0.10 |

### 5.3 System Transitions Across Bandgap Regimes

We identify distinct system transitions in which dominant elemental combinations shift abruptly at specific thresholds (Table 2).

**Low-bandgap regime ($\Delta E_g \approx 0.0$ eV).** Ag–Zr-based systems dominate, maintaining high conductivity while leveraging Zr to ensure minimal amorphous-structure stability.

**Intermediate transition regime ($\Delta E_g = 0.25$–$0.75$ eV).** Ag-based systems vanish near $\Delta E_g \approx 0.25$ eV, and the optimum shifts to Mg–Ti systems, reflecting a trade-off in which mobility ($ISWO_C$) is partially sacrificed to secure a wider bandgap via lightweight-metal-based stabilization. At higher targets, a Zr-based backbone re-emerges, combined with elements such as In, Sn, and Ga to restore the mobility–stability balance and maximize design flexibility.

**High-bandgap regime ($\Delta E_g \approx 1.0$ eV).** The strategy converges to an Al–Mg backbone, with strong gap openers such as Sb, signaling entry into a property regime suitable for dielectric or barrier layers beyond semiconductor channels.

### 5.4 Device-requirement-driven Candidate Categorization

To recommend optimal compositions aligned with user design objectives within the Pareto solution set, we analyzed candidates in three performance-oriented buckets (Table 3).

**High Mobility**: Dominated by the In–Ga–Zn family, this bucket maximizes $s$-orbital overlap and achieves the lowest $ISWO_C$ ($\approx 6.17$), at the cost of reduced bandgap and higher formation energy.

**High Stability**: Strong bonding driven primarily by Zr (63%) and Ti (23%) yields the lowest formation energy ($\approx -3.99$ eV). However, the increased ionicity limits conduction-band overlap, resulting in suppressed mobility.

**Balanced**: All property metrics converge to intermediate values. This regime reflects a cooperative design strategy, with Zr contributing stability, In/Ga/Zn enhancing transport, and Al/Mg enabling bandgap tuning.

### 5.5 Emergence of an Al–Mg Archetype in Global Pareto Optima

We analyzed the 20 global optima (Global Knee) obtained by integrating all bandgap targets. As shown in Fig. 4a, unlike bucket-based candidates biased toward either mobility or stability, the Global Knee candidates cluster at the optimal compromise between these properties, occupying a region with markedly lower formation energy and superior $ISWO_C$ relative to the IGZO baseline.

Notably, all 20 global optima emerge at the maximum bandgap requirement, $\Delta E_g = 1.0$ eV (i.e., $E_g \approx$ IGZO $+ 1.0$ eV; Fig. 3), indicating that high-bandgap conditions impose strong physical constraints to maintain charge transport and structural stability. These extreme constraints act as a physical filter that drives the Pareto knee to converge to a specific composition region.

Table 4: Statistical summary of Global Knee candidates ($n = 20$).

(a) System distribution and property ranges.

| System | Count | $\Delta E_{gap}$ | $ISWO_C$ | Stability |
|---|---|---|---|---|
| (Al, Mg) | 5 | 3.80–4.27 | 9.01–10.19 | $-2.64$–$-2.25$ |
| (Al, In, Mg) | 4 | 3.59–4.03 | 8.47–9.57 | $-2.52$–$-2.24$ |
| (Al, Ga, Mg) | 3 | 3.83–4.10 | 9.10–9.76 | $-2.53$–$-2.28$ |
| (Al, Mg, Si) | 3 | 3.82–4.04 | 9.03–9.45 | $-2.39$–$-2.26$ |
| (Al, Mg, Zr) | 2 | 3.79–3.83 | 9.50–12.86 | $-2.69$–$-2.29$ |

(b) Elemental composition statistics.

| Element | Mean | Std |
|---|---|---|
| Al | 0.5285 | 0.1254 |
| Mg | 0.4420 | 0.1204 |
| In | 0.0080 | 0.0217 |
| Zr | 0.0075 | 0.0217 |
| Si | 0.0075 | 0.0185 |

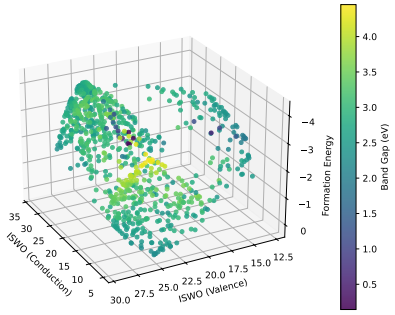



Figure 3: 3D global Pareto sets across all bandgap targets.

Consistently, the Global Knee systems share a common Al–Mg backbone (Table 4a), with Al and Mg accounting for approximately 97% of the total composition (Table 4b), suggesting convergence to a near-binary Al–Mg oxide archetype under the most stringent property requirements.

### 5.6 Composition Robustness of the Al–Mg Oxide Archetype

From the composition distributions of the global-knee optimal systems, we observe a flat region when the Al fraction exceeds 0.55, where material properties remain nearly invariant despite changes in the Mg:X ratio (Fig. 5). This indicates that the third element functions as a dopant, fine-tuning target properties without perturbing the underlying backbone.

As a result, the global optimum converges to a variable dopant configuration on a robust Al–Mg backbone. Higher Al content further suppresses sensitivity to composition variations, providing a practical materials design guideline for achieving high yield in large-area AMOLED manufacturing.

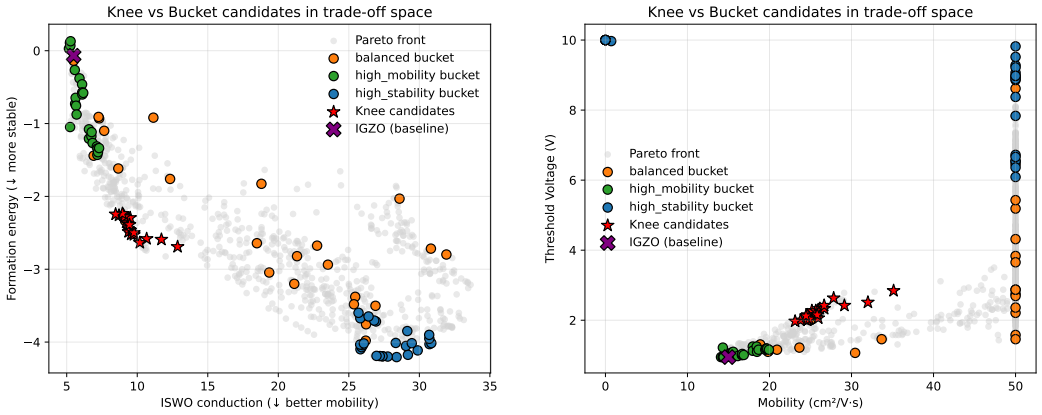
### 5.7 Device-level Validation of AI-designed Oxide TFT Channels

We performed PIM-based TCAD simulations to evaluate the practical utility of the AI-proposed compositions (Fig. 4b).

**IGZO baseline validation**: For commercial IGZO ($In_{50}Ga_{50}Zn_{50}$), we obtained $V_{th} = 0.95$ V and $\mu = 15.0$ cm$^2$/V $\cdot$ s. These values are consistent with typical experimental reports, supporting the reliability of the simulation setup.

**Device performance of Global Knee candidates**: The 20 global-optimal candidates showed strong statistical advantages over IGZO across all metrics. We obtained approximately 76% higher mobility ($\mu_{avg} \approx 26.4$) and excellent switching behavior ($SS_{avg} \approx 128.6$ mV/dec).

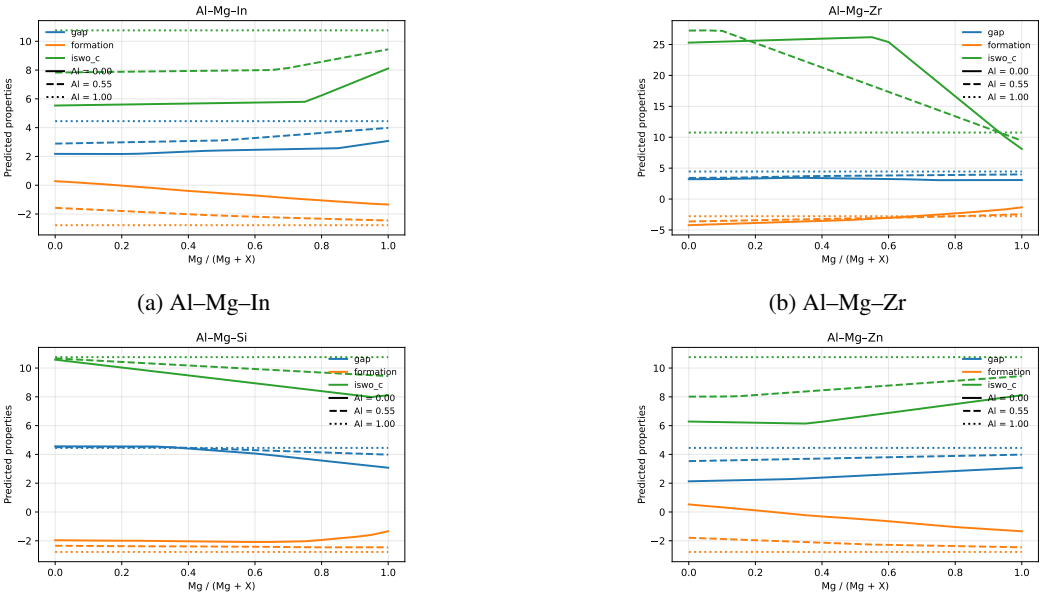
**Process reproducibility**: The threshold voltage is densely distributed within the operating window ($V_{th} = 2.26 \pm 0.23$ V), indicating high uniformity for large-area processing.




(a) $E_f$–$ISWO_C$ trade-off.

(b) TCAD performance validation.

Figure 4: Global distribution and performance of Pareto results.




(a) Al–Mg–In

(b) Al–Mg–Zr

(c) Al–Mg–Si

(d) Al–Mg–Zn

Figure 5: Fine-tuning effects of ternary elements (In, Zr, Si, Zn).

# 6 Conclusion

This study addresses the challenge of identifying oxide TFT channel materials that can replace IGZO to meet the high-performance and high-reliability demands of next-generation AMOLED backplanes. We propose the *MTTB* framework, which integrates composition-based property prediction, Pareto-based multi-objective screening, and TCAD-level device-performance inference via the Physics-Inference Module (PIM). Under stringent requirements such as a high bandgap, the design space is shown to converge to an Al–Mg binary archetype. We further demonstrate that Al–Mg–X systems exhibit low sensitivity to composition variations while enabling dopant-based fine-tuning, providing robust oxide TFT channel design guidelines for large-area AMOLED manufacturing. These systems also indicate improved device-performance potential compared with conventional IGZO.

The proposed framework supports physically consistent decision-making from material candidate discovery to device-performance validation and can be extended to incorporate process variables and device-stack-level co-design.

# References


[1] J.-H. Lee *et al.*, "Recent progress in organic light-emitting diodes for display applications," *Advanced Materials*, vol. 32, no. 34, 2002840, 2020.

[2] S. Reineke, "Complementary LED technologies," *Nature Materials*, vol. 12, pp. 474–475, 2013.

[3] E. Fortunato, P. Barquinha, and R. Martins, "Oxide semiconductor thin-film transistors: a review of recent advances," *Advanced Materials*, vol. 24, no. 22, pp. 2945–2986, 2012.

[4] K. Nomura *et al.*, "Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors," *Nature*, vol. 432, no. 7016, pp. 488–492, 2004.

[5] T. Kamiya, K. Nomura, and H. Hosono, "Present status of amorphous In–Ga–Zn–O thin-film transistors," *Science and Technology of Advanced Materials*, vol. 11, no. 4, 044305, 2010.

[6] P. Raccuglia *et al.*, "Machine-learning-assisted materials discovery using failed experiments," *Nature*, vol. 533, no. 7601, pp. 73–76, 2016.

[7] K. T. Butler *et al.*, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.

[8] Q. Ding, A. V. Kuhlmann, A. Fuhrer, and A. Schenk, "A Generalizable TCAD Framework for Silicon FinFET Spin Qubit Devices with Electrical Control," arXiv:2306.03213, 2023.

[9] D.-G. Kim *et al.*, "Negative threshold voltage shift in an a-IGZO thin film transistor under X-ray irradiation," *RSC Advances*, vol. 9, pp. 20865–20870, 2019.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[11] A. Jain *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, 011002, 2013.

[12] L. Ward *et al.*, "Matminer: An open source toolkit for materials data mining," *Computational Materials Science*, vol. 152, pp. 60–69, 2018.

[13] S. Curtarolo *et al.*, "The high-throughput highway to computational materials design," *Nature Materials*, vol. 12, pp. 191–201, 2013.

[14] M. J. van Setten *et al.*, "Complex amorphous oxides: property prediction from high throughput DFT and AI for new material search," *Materials Advances*, vol. 3, pp. 8413–8427, 2022.

[15] T. D. K"uhne *et al.*, "CP2K: An electronic structure and molecular dynamics software package – Quickstep: Efficient and accurate electronic structure calculations," *The Journal of Chemical Physics*, vol. 152, 194103, 2020.

[16] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Wiley, 2006.

[17] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd ed. Oxford University Press, 2011.

[18] Python Software Foundation, *Python Language Reference*, version 3.12, 2023.

[19] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[20] Google Research, *Google Colaboratory*, online platform documentation, accessed 2026.

[21] NVIDIA Corporation, *NVIDIA T4 Tensor Core GPU*, product brief/datasheet, 2018.

[22] A. Biswas and P. Schalberger, "PyTFT: Automated extraction of TFT parameters for simulation," software repository, University of Stuttgart, 2023.

[23] A. M. Gopakumar, P. V. Balachandran, D. Xue, J. E. Gubernatis, and T. Lookman, "Multi-objective Optimization for Materials Discovery via Adaptive Design," *Scientific Reports*, vol. 8, 3738, 2018.

[24] K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit, and B. Yoo, "Bias free multiobjective active learning for materials design and discovery," *Nature Communications*, vol. 12, 2312, 2021.

# A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission.**

# AI Co-Scientist Challenge Korea Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the four-stage MTTB framework, the motivation (limitations of single-objective ML screening), and the scope (beyond-IGZO oxide TFT channels), all of which are consistently supported by the proposed methods and experimental results.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper explicitly clarifies the scope and limitations of the Physics-Inference Module (PIM), stating that it is a physics-informed surrogate rather than a full TCAD replacement, and that it relies on simplified, IGZO-anchored mappings which do not capture detailed electrostatics or process-dependent effects.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not present formal theorems or mathematical proofs; it is primarily a framework and methodology paper combining ML, multi-objective optimization, and physics-informed modeling.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper specifies datasets, model architecture (MLP), objective formulations, Pareto procedures, PIM mappings, and experimental settings, and additionally provides an anonymized code repository to support reproducibility.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While AI Co-Scientist Challenge Korea does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides a public, anonymized repository link containing the implementation and instructions, satisfying open-access and anonymity requirements at submission time.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper details the dataset composition, prediction targets, evaluation metrics, device geometry, operating conditions, and computational environment, enabling clear interpretation of the results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

Justification: Although averages and distributions are reported, the paper does not systematically provide confidence intervals, error bars, or formal statistical significance tests for the experimental comparisons.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper specifies the software stack, execution environment (Google Colab), GPU type (NVIDIA T4), and general computational setup for the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://nips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work involves materials modeling and simulation without human subjects, personal data, or ethically sensitive applications, and preserves anonymity as required.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Although the paper does not include a standalone "Broader Impacts" section, it explicitly clarifies that the PIM is a physics-informed surrogate intended only for comparative screening, not for final decision-making. This implicitly addresses responsible use by preventing overinterpretation or misuse of surrogate predictions, while the overall framework supports positive industrial impact through more efficient and resource-aware materials exploration.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The released assets are materials design models and simulations with no clear dual-use or misuse risk.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper properly cites prior datasets, software tools, and libraries, and references their original publications.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The MTTB framework and PIM are clearly described, and accompanying code and documentation are provided in the repository.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The study does not involve human participants or crowdsourced data.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No human-subject research is conducted, so IRB approval is not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.