

Machine Learning Framework for the Inverse Design of Tight-Binding Parameters in Compound Semiconductors

Kaito Seike¹, and Shingo Sato²
 Graduate School of
 Science and Engineering
 Kansai University
 Osaka, Japan
 {¹k181021, ²satos}@kansai-u.ac.jp

Abstract— This paper proposes a machine learning (ML) framework for designing the tight-binding (TB) parameters in semiconductor materials. By integrating a "physics-informed loss function," we successfully predicted TB parameters that reproduce target effective masses and bandgaps with high accuracy, including materials with narrow-gap characteristics. Furthermore, sensitivity analysis revealed that strong correlations between physical properties define the fundamental precision limits of the inverse design.

Keywords— Inverse Design, Machine Learning, Tight-Binding Model, Physics-Informed Neural Networks, Semiconductor Band Structure, Parametric Coupling.

I. INTRODUCTION

The development of next-generation devices for quantum and power electronics necessitates highly precise methodologies for controlling the electronic states of materials and ensuring device reliability under extreme operating conditions. While the tight-binding (TB) method [1] is extensively utilized for large-scale device simulations owing to its computational efficiency, deriving the optimal parameter sets to accurately reproduce specific band structures remains a significant challenge [2] [3]. Traditionally, this parameterization process has demanded substantial domain expertise and extensive trial-and-error.[4][5] Furthermore, in practical device development, a critical disconnect exists between "material synthesis" and "failure analysis." A systematic cycle to rapidly feed physical degradation information—such as localized distortions in the band structure—back into new material design has not yet been established. To address these bottlenecks, this study proposes a multi-stage Autoencoder (AE) architecture incorporating physical constraints. This framework not only establishes a robust inverse design methodology to directly deduce TB parameters from desired band characteristics but also provides a novel analytical platform capable of reverse-engineering microscopic parameter shifts from macroscopic degradation phenomena.

II. METHODOLOGY

The proposed framework is based on the following three key technologies:

A. Dataset Construction

Training data were synthesized using a custom TB framework targeting compound semiconductors. The model employs 31 TB parameters to compute energy dispersions for 1,000 randomly sampled configurations. Each calculation was performed along the high-symmetry L-Γ-X-W-K path of the face-centered cubic (FCC) lattice to obtain the corresponding band structures.

B. Architecture of the Inverse Design Model

As illustrated in Fig. 1, the proposed inverse design framework employs a hierarchical multi-stage architecture to independently capture the latent representations of both the parameter and band structure spaces.

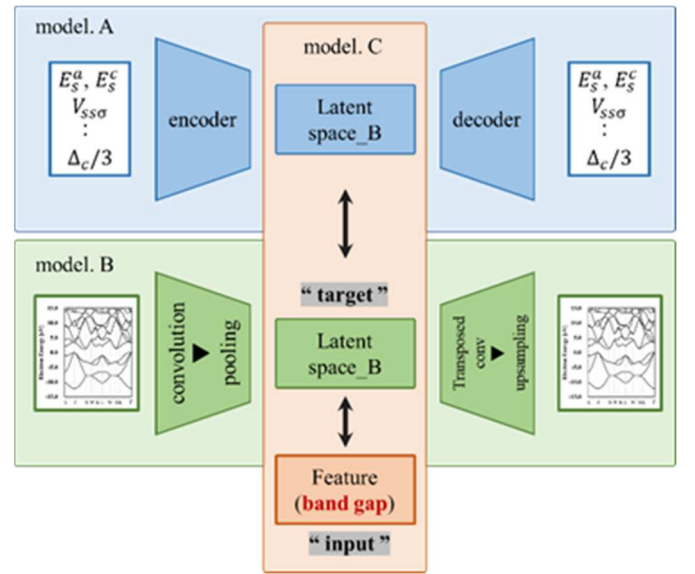


Fig. 1. Conceptual architecture of the proposed multi-stage autoencoder framework.

Model A (a fully connected autoencoder) compresses the 31-dimensional TB parameters into a 32-dimensional latent space. In parallel, Model B utilizes a convolutional autoencoder (CAE) to extract dispersion features from the 2D band structure grids (172×40), reducing them to a 64-dimensional vector. Finally, Model C bridges these domains via a deep Residual Network (ResNet) composed of three residual blocks. This architecture maps the extracted band features onto the parameter latent space, mitigating the vanishing gradient problem while capturing the highly non-linear parameter mappings. To ensure both numerical accuracy and physical consistency, we designed a custom composite loss function. The base loss combines relative, logarithmic, and absolute errors to handle physical properties across various scales. Furthermore, to eliminate unphysical solutions such as band inversion, we introduced a "Hard Sign Constraint." Additionally, curvature values which correspond to the effective mass were preprocessed using non-linear logarithmic scaling to suppress gradient explosions. Weights training and targeted sample weights were applied; specifically, we assigned 10x to 20x higher

weights to critical features like the Γ -point bandgap and X-point conduction.

C. Physics-Informed Loss Function and Weight

To mitigate the "gradient conflict" encountered by the ML optimizer and ensure physical consistency, we introduced a custom weighting strategy based on physical features. First, to prevent gradient explosion caused by extreme variations in curvature (which is inversely proportional to effective mass), a non-linear logarithmic scaling was applied as preprocessing to all curvature features

$$\tilde{C} = \text{sgn}(C) \log_{10}(|C| + 1) \quad \text{--- (1)}$$

The mapping network (Model C) was trained using the Huber loss (L_{Huber}), incorporating a dynamic sample weight w_i for each training instance i . The total loss L is defined as:

$$L = \frac{1}{N} \sum_{i=1}^N w_i \cdot L_{\text{Huber}}(y_i, \hat{y}_i) \quad \text{--- (2)}$$

To prioritize predictive accuracy in the narrow-gap regime while preserving the curvature at the extrema, the weight w_i is formulated as follows:

$$w_i = \max\left(1, 100 \exp\left(-3.2 \left|E_{g,\Gamma}^{(i)}\right|\right)\right) + \sum_j \alpha_j |C_j^{(i)}| \quad \text{--- (3)}$$

The first term applies an exponentially increasing penalty (up to 100-fold) as the Γ -point direct gap approaches zero, forcing the optimizer to fit hypersensitive narrow-gap materials. The second term serves as a linear regularization for critical curvatures ($\alpha_j \in [0.05, 0.10]$) to enforce correct band shapes, particularly aiming to reproduce the effective masses of typical wide-gap materials such as GaAs. However, as discussed later, enforcing this curvature penalty mathematically competes with the narrow-gap optimization, inducing a fundamental gradient conflict.

Before evaluating the final prediction performance, we verified the stability of the proposed training process. Fig. 2 illustrates the learning curve of the mapping network (Model C) trained with the physics-informed loss function. Despite the inherent risk of gradient explosion caused by the hypersensitivity of narrow-gap materials, the training loss smoothly converged without any divergence. This stable convergence confirms the effectiveness of the logarithmic curvature scaling (1) and the dynamic sample weighting (3) in appropriately balancing the gradient updates.

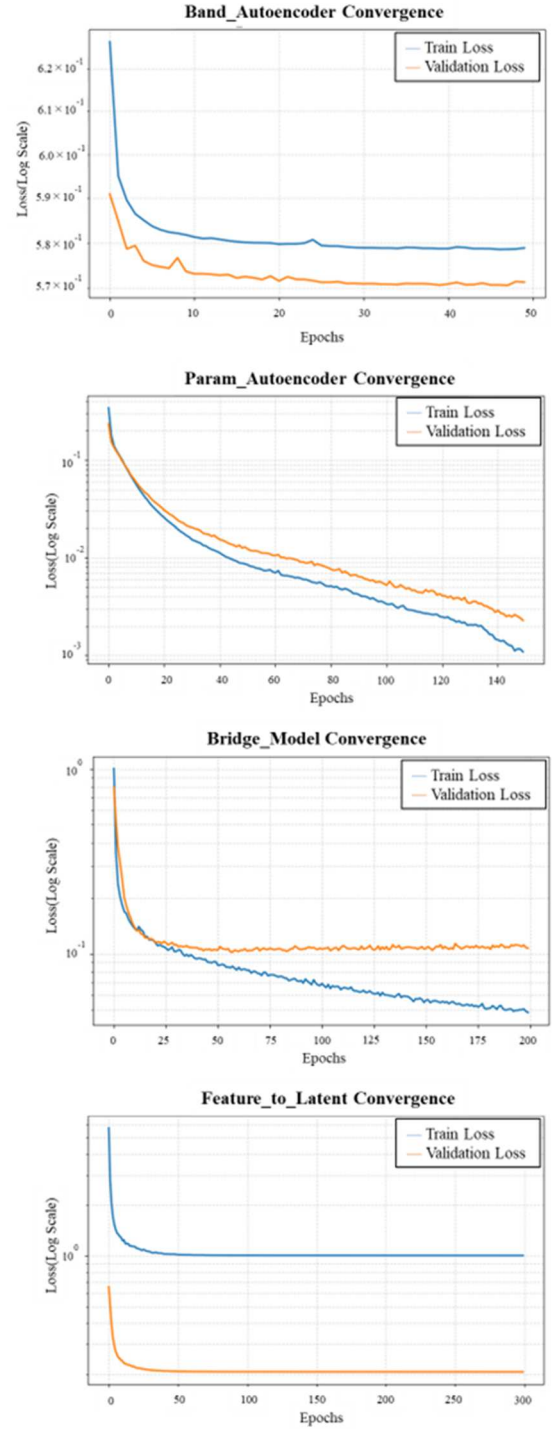


Fig. 2. Convergence profiles of the proposed ML framework.

III. RESULTS AND DISCUSSION

A. Prediction Performance

As shown in Fig. 3 and Fig. 4 the baseline multi-stage autoencoder successfully captures the overall macroscopic band profiles across the compound semiconductors. However, extensive hyperparameter tuning of the loss weight revealed that pushing the model toward zero-error optimization exposes a fundamental physical limitation of the TB framework itself.

B. Fundamental Trade-off: Macroscopic Gaps vs. Microscopic Curvatures

Manipulating the physics-informed weights (α_j) in (3) revealed a severe gradient conflict, manifesting as a "seesaw" effect between different physical properties. Specifically, when the weights were heavily biased to strictly enforce the correct microscopic curvatures (effective masses) of standard materials like GaAs, the predictive accuracy for macroscopic energy scales significantly degraded. Under this condition, the relative error for the highly sensitive narrow direct gap of InSb exploded (as shown in Fig. 3). Conversely, relaxing the curvature penalty to prioritize the narrow-gap optimization resulted in unphysical distortions in the GaAs effective masses, as demonstrated in Fig. 6. This explicit trade-off demonstrates that satisfying both macroscopic energy gaps and microscopic topologies simultaneously is fundamentally restricted within a single optimal parameter set.

C. Topological Anomaly and Relative Error Explosion in Flat Bands

Furthermore, evaluating the conduction band at the X-point in GaSb exposed a unique topological anomaly, highlighting a fundamental limitation of standard optimization and evaluation metrics. Unlike typical compound semiconductors where the conduction band at the X-point forms a standard downward-convex dispersion (positive curvature), the X-point in GaSb exhibits a locally inverted, upward-convex dispersion (negative curvature). Because the absolute magnitude of this anomalous curvature approaches zero, the energy difference between the true and predicted bands is in the order of only a few meV. While this deviation is visually indistinguishable on a macroscopic band diagram, it manifests mathematically as a massive relative curvature error exceeding 180%, as explicitly shown in Fig. 5.

Standard loss functions (e.g., mean squared error) treat this microscopic topological inversion as negligible noise, imposing almost no penalty for flipping the conduction band's curvature. However, capturing this localized topology simultaneously with hypersensitive narrow-gap characteristics is fundamentally restricted by parametric coupling. The same tight-binding parameters preserving the GaSb X-point topology are subjected to massive gradient updates to minimize the narrow-gap errors of InSb. Consequently, the optimizer sacrifices the imperceptible flat-band topology to satisfy macroscopic gap penalties.

	GaAs	InP	InAs	AlP	GaP	AlAs	AlSb	GaSb	InSb
CBM $_{\Gamma}$	9.46%	32.05%	0.81%	2.57%	4.01%	22.70%	17.29%	20.23%	20.81%
CBM $_L$	14.81%	11.65%	46.12%	27.34%	11.91%	13.37%	47.56%	16.54%	13.00%
CBM $_X$	103.39%	0.41%	10.03%	40.24%	61.65%	15.44%	55.55%	183.81%	57.37%
Gap $_{\Gamma}$	2.89%	47.08%	28.85%	2.19%	6.35%	14.05%	10.76%	9.44%	117.67%
Gap $_L$	4.75%	19.15%	13.36%	1.41%	1.86%	2.44%	3.53%	30.51%	12.24%
Gap $_X$	2.44%	2.83%	9.88%	3.79%	3.73%	0.69%	5.93%	9.16%	4.74%
VBM $_{\Gamma-L}$	35.34%	59.38%	19.45%	29.61%	20.38%	34.20%	12.65%	36.47%	64.45%
VBM $_{\Gamma-H}$	24.20%	1.33%	11.81%	15.95%	18.98%	17.41%	26.55%	24.64%	7.81%
VBM $_L$	18.13%	1.85%	6.36%	5.95%	13.77%	18.43%	42.06%	20.09%	30.47%
VBM $_X$	28.03%	19.40%	163.04%	20.60%	46.87%	1.98%	17.25%	42.13%	50.41%

Fig. 3. Relative prediction errors (%) of bandgaps and effective masses across nine compound semiconductors.

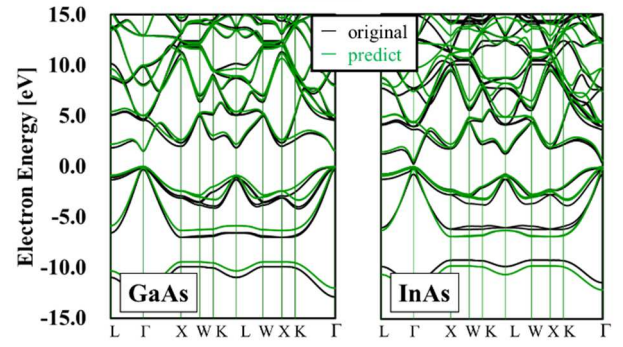


Fig. 4. Comparison between target and predicted band structures

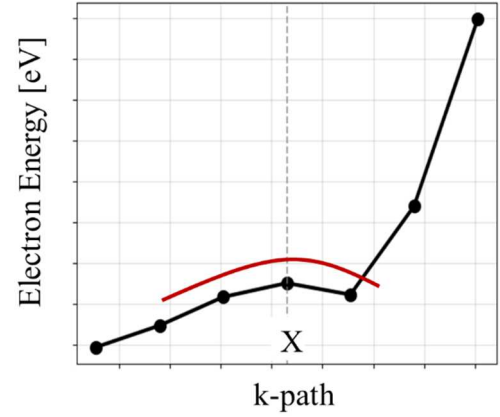


Fig. 5. Zoomed-in CBM of GaSb at the X-point (>180%).

	original	predict	error
CBM $_{\Gamma}$	133.4436448	160.4365079	20.23%
CBM $_L$	4.919936583	4.10622751	16.54%
CBM $_X$	-7.194721565	6.030191312	183.81%
Gap $_{\Gamma}$	0.821012	0.743471	9.44%
Gap $_L$	1.846411	2.40979	30.51%
Gap $_X$	4.13665	4.51553	9.16%
VBM $_{\Gamma-L}$	-52.10124968	-71.1038557	36.47%
VBM $_{\Gamma-H}$	-13.33801231	-10.05203595	24.64%
VBM $_L$	2.273087594	2.729686881	20.09%
VBM $_X$	-2.490480542	-1.441250313	42.13%

Fig. 6. Target vs. predicted curvatures for GaAs

D. Inherent Physical Constraints in Narrow-Gap Regimes

Prior to analyzing the model's behavior, we evaluated the physical relationship between bandgaps and effective masses across major symmetry points (Γ , L, X). As shown in Fig. 7, the Γ -point conduction band effective mass (CBM) decreases exponentially as the direct gap approaches zero, consistent with $k \cdot p$ perturbation theory. In contrast, the valence band mass (VBM) and indirect valleys (L, X) remain relatively stable, as they are governed by heavy-hole bands and spatial separation in k -space, respectively. This asymmetric sensitivity imposes a fundamental physical constraint on the TB framework. Because the same hopping parameters must simultaneously govern both the highly sensitive Γ -valley and the stable valleys, the ML optimizer faces a severe "gradient conflict." Consequently, while wide-gap properties are easily fitted, satisfying the stable valleys in narrow-gap materials requires minor parameter compromises that disproportionately increase the relative error of the near-zero Γ -gap.

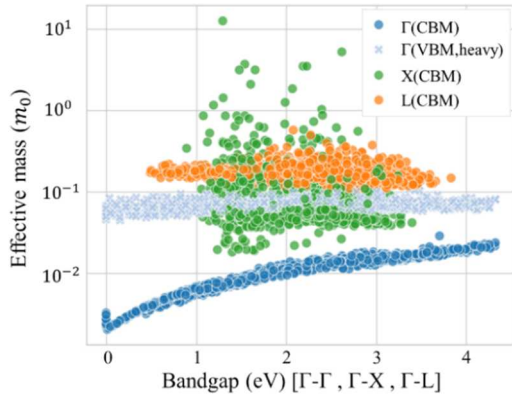


Fig. 7. Relationship between bandgaps and effective mass at Γ , L, and X points across 1,000 generated training datasets. (m_0 denotes the free electron mass)

E. Physical Insights

This parametric coupling constitutes a fundamental physical constraint inherent to the mathematical formulation of the empirical TB model. Specifically, the global nature of the localized atomic orbital basis—where a single set of hopping integrals (e.g., $V_{SS\sigma}$, $V_{Sp\sigma}$) mathematically dictates the entire Hamiltonian matrix across the entire Brillouin zone—creates an inescapable interdependence between disparate k-points. This complex interdependence is explicitly visualized in the correlation heatmap in Fig. 8. Our quantitative analysis reveals that the precision limits of the inverse design originate directly from these restrictive physical premises of the finite-basis TB representation, rather than the representational capacity or expressivity of the ML algorithm itself. Even a highly over-parameterized deep neural network cannot circumvent the physical reality that multiple, mutually independent band features must be mapped back to a highly constrained, lower-dimensional parameter space. Notably, the optimization difficulty drastically escalates for narrow-gap materials. As the target direct bandgap approaches zero, its sensitivity to infinitesimal parameter perturbations diverges exponentially. This behavior is deeply rooted in standard k-p perturbation theory, which dictates that the conduction band effective mass is inversely proportional to the energy gap due to the strong quantum mechanical repulsion between closely spaced conduction and valence bands. Within the TB framework, identical hopping parameters concurrently dictate both the absolute energy levels of band extrema and their respective local curvatures across all high-symmetry points (Γ , L, X). Consequently, an optimizer striving to reconcile diverse and physically distant targets—such as the unique flat-band topology of the GaSb X-point and the extreme narrow gap of InSb—encounters a severe "gradient conflict" during the backpropagation process. Adjusting a specific hopping parameter to perfectly fit the Γ -point gap inherently alters the dispersion at the X or L points, destroying the delicately balanced effective masses. Although band curvatures can be adequately preserved via the explicit physics-informed loss weight adjustments (3), the hypersensitivity of near-zero gaps dictates that even the most minor parameter compromises—made to satisfy the stable point constraints—induce disproportionately large relative errors

in the macroscopic bandgap. The loss landscape is essentially dominated by the steep gradients of the narrow gap, which overshadow the subtle topological gradients of the flat bands. These findings provide a rigorous physical justification for the inherent precision boundaries observed in multi-target semiconductor parameterization, demonstrating that achieving zero-error fitting across all physical properties simultaneously is fundamentally difficult when the empirical TB model is restricted to first nearest-neighbor interactions.

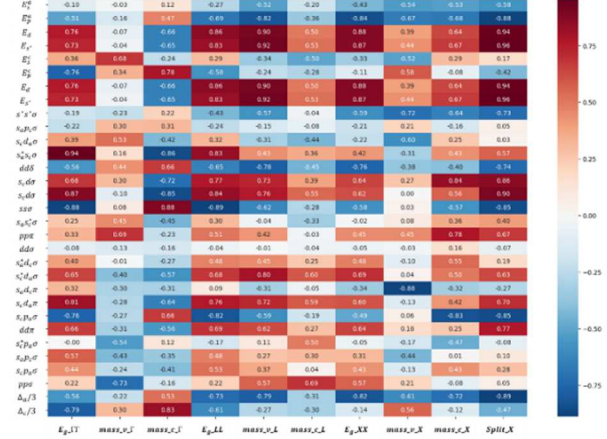


Fig. 8. Correlation heatmap between 31 tight-binding parameters and critical physical features.

IV. CONCLUSION

In conclusion, the proposed physics-informed ML framework has realized the high-speed, automated inverse design of tight-binding parameters, effectively overcoming the exhaustive trial-and-error limitations of conventional methods. By explicitly addressing the fundamental gradient conflicts inherent in parametric coupling, this methodology serves not only as a robust tool for material synthesis but also as a novel analytical platform for reliability assessment. It enables the rapid reverse-engineering of microscopic parameter shifts directly from macroscopic degradation phenomena. Future work will extend this approach to more complex configurations, including heterostructure interfaces and localized strained states, to establish a comprehensive feedback cycle for the development of next-generation quantum and power devices.

REFERENCES

- [1] Jean-Marc Jancu, et al.: Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters, Phys. Rev. B 57, 6493 (1998)
- [2] Josep Buch Palasi.: A data-driven approach for tight-binding parametrization : Universitat de Barcelona (2023)
- [3] Zifeng Wang, et al : Machine learning method for tight-binding Hamiltonian parameterization from ab-initio band structure : npj Computational Materials, 7, zp. 11 (2021)
- [4] E. Tsymalov *et al.*, "Machine learning for deep elastic strain engineering of semiconductor electronic band structure and effective mass," *npj Comput. Mater.*, vol. 7, p. 76, 2021.
- [5] T. Xie and J. C. Grossman, "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties," Phys. Rev. Lett., vol. 120, p. 145301, 2018.