OL-Transformer: A Fast and Universal Surrogate Simulator for Optical Multilayer Thin Film Structures

Taigao Ma [1] **Haozhu Wang** [2] **L. Jay Guo** [1]

Abstract

Deep learning-based methods have recently been established as fast and accurate surrogate simulators for optical multilayer thin film structures. However, existing methods only work for limited types of structures with different material arrangements, preventing their applications towards diverse and universal structures. Here, we propose the **Opto-Layer (OL) Transformer** to act as a universal surrogate simulator for enormous types of structures. Combined with the technique of *structure serialization*, our model can predict accurate reflection and transmission spectra for up to 10^{25} different multilayer structures, while still achieving a six-fold degradation in simulation time compared to physical solvers. Further investigation reveals that the general learning ability comes from the fact that our model first learns the physical embeddings and then uses the self-attention mechanism to capture the hidden relationship of light-matter interaction between each layer.

1. Introduction

Optical multilayer thin film structure (shorten as "multilayer structure") is a type of photonic structure that consists of multiple layers of different materials stacking on top of each other, with thickness typically ranging from tens of nanometers to a few micrometers. Because of the ease of fabrication, multilayer structures have been widely used in both scientific and industrial applications, including structural color (Wang et al., 2023), photovoltaic (Liu et al., 2013), display devices (Zheludev, 2007), etc. To enable these applications, researchers need to first understand the physical relationship between a multilayer structure and corresponding optical properties, e.g., transmission and reflection. Traditional simulation methods including Transfer Matrix Methods (TMM) (Byrnes, 2016) and Rigorous Coupled Wave Analysis (RCWA) (Hugonin & Lalanne, 2021) use matrix algebra to analytically or semi-analytically calculate the reflection and transmission coefficients. However, these physics-based simulation methods are usually time-consuming. In addition, a new simulation needs to be performed from scratch when facing a different structure. Thus, the development of a fast simulation method becomes fundamental for multilayer structures applications.

Recently, researchers have started to use deep learning to accelerate the simulation process by leveraging the strong generalization ability, including Multi-Layer Perceptron (MLP) (Liu et al., 2018), MLP-Mixer (Deng et al., 2021), and transformer (Chen et al., 2023). Although obtaining the training dataset through physical simulation can take some time, such an investment is a one-time payment. Once trained, these neural networks are able to capture the general mapping from the space of structures to the space of optical properties, serving as a fast and computationally efficient surrogate model to replace physical simulations.

However, many existing surrogate models can only expedite the simulation of structures with fixed material arrangements, e.g., the three-layer structure of $Ag/SiO_2/Ag$ in (Deng et al., 2021) and the six-layer structure of $MgF_2/SiO_2/Al_2O_2/TiO_2/Si/Ge$ in (Chen et al., 2023). This is because materials have dispersion, making them only accessible through categorical representations, instead of continuous variables. A universal method that can predict and simulate optical properties for diverse structures with different numbers of layers and varied materials arrangements is in great need. In this paper, we propose the Opto-Layer Transformer (OL-Transformer) as a universal surrogate model by leveraging the strong learning and generalization abilities of transformer. After training on a large dataset, our model can work as a fast and authentic surrogate solver for multilayer structures with up to 20 layers and 18 different materials, corresponding to a total of $18^{20} \sim 10^{25}$ different structures (this can be further expanded as our model is highly scalable). In addition, compared to physical

[1]University of Michigan, Michigan, USA [2]Amazon Web Services, Illinois, USA. Correspondence to: L. Jay Guo <guo@umich.edu>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023. Copyright 2023 by the author(s).

simulators, our model also achieves a six-fold degradation in simulation time and can go to a ~3800-fold speedup when using batch calculation.

2. Related Works

Deep learning-based surrogate models have been used in many scientific applications to speed up the simulation and prediction, including molecular properties prediction (Broberg et al., 2022), predicting the dynamics of physical phenomenon (Geneva & Zabaras, 2022), and meteorological predictions (Khorrami et al., 2021). They are also widely used for solving inverse problems, which seek to recover the causes given the observed results, including nanophotonic inverse design (Jiang et al., 2021), chemical material inverse design (Fu et al., 2023), etc. Our work aims to solve the simulation of optical multilayer structures, with the goal of speeding up the prediction of optical properties.

There has been some work that leverages learned knowledge to deal with different structures. For example, (Qu et al., 2019) used transfer learning to assist in the simulation of a 10-layer structure after learning on the 8-layer structure. (Kaya & Hajimirza, 2019) used transfer learning to help the optimization of a multi-layer solar cell. Meta-learning has also been applied to generalize the learning on different applications, including detector simulations and design (Zhang et al., 2020), PINN-based 1D arc simulation (Zhong et al., 2023), hydrogen storage materials simulations and design (Sun et al., 2021), etc. However, these methods usually require adaptations to new datasets, restricting their applications to general purposes. We seek to find a universal surrogate model that works for as many different structures as possible through a single training.

3. Methods

3.1. Problem Set

For a given multilayer structure with N layers (see Fig. 1a), we denote the material arrangements as $\mathbf{m} = \{m_1, m_2, \ldots, m_N\}$ and the thickness sequence as $\mathbf{t} = \{t_1, t_2, \ldots, t_N\}$. Here, m_i, t_i refers to the material and thickness at the i_{th} layer, respectively. $m_i \in \mathbf{M}$ is a discrete variable that can take several distinct values from the material database $\mathbf{M}$. Then, a multilayer structure can be described as $X = \{\mathbf{m}, \mathbf{t}\} \in \mathcal{X}$. A physical simulator $\mathcal{S} : \mathcal{X} \to \mathcal{Y}$ maps the multilayer structure X to the d-dimensional optical properties $Y = \{y_1, y_2, \ldots, y_d\} \in \mathcal{Y}$ and works as an oracle for any type of material arrangements. Existing surrogate models $\hat{\mathbf{S}}(\mathbf{t}|\theta_{\mathbf{m}}) : \mathcal{X} \to \hat{\mathcal{Y}}$ with learnable parameters $\theta_{\mathbf{m}}$ take in the structure X with different thickness $\mathbf{t}$ but with fixed material arrangement $\mathbf{m}$ and output predicted optical properties $\hat{Y} = \{\hat{y_1}, \hat{y_2}, \ldots, \hat{y_d}\}$. In this work, our surrogate model $\hat{S}(\mathbf{m}, \mathbf{t}|\theta) : \mathcal{X} \to \hat{\mathcal{Y}}$ with

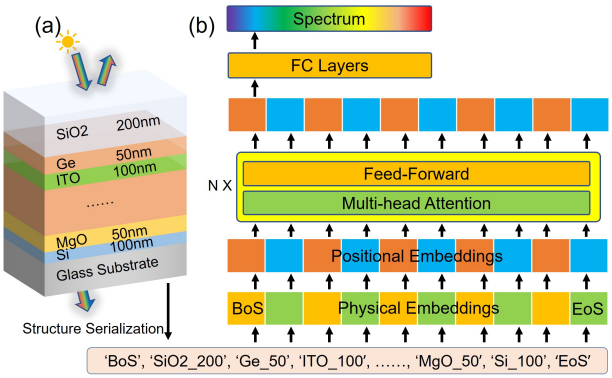

Figure 1: (a) An example of a multilayer structure as well as its structure serialization. (b) The architecture of OL-Transformer. FC Layers: Fully Connected layers.

parameters θ wants to predict the optical properties for universal structures with different $\mathbf{m}$ and different thickness $\mathbf{t}$. In this work, we consider predicting the transmission and reflection spectra from 400 nm to 1100 nm. Both spectra are discretized by 10 nm, making d=2×71=142. Other types of optical properties can also be predicted similarly.

3.2. Structure Serialization

In order to expand the model's capability towards versatile $\mathbf{m}$, we combine with the recently developed technique called structure serialization (Ma et al., 2023), where we use structure tokens to represent the material and thickness information (m_i, t_i) at each layer simultaneously, similar to how Natural Language Processing (NLP) researchers tokenize language sentences. By appending multiple tokens one-by-one, we can convert a multilayer structure into a sequence of tokens that the transformer model can deal with. We also use a special token of 'EoS' (end of sequence) to enable the learning of structures with different numbers of layers (we set the maximum to be 20 layers). There are 18 different types of materials in our material database $\mathbf{M}$, all of which are widely accessible in many nanofabrication centers. In addition, considering it is impossible to use an infinite number of tokens to describe a continuous thickness, we discretize the thickness by 10 nm and form 50 different choices from 10 nm to 500 nm. Therefore, there are 18×50+1=901 tokens in our vocabulary and the total number of structures with different material arrangements expands to $18^{20} \sim 10^{25}$. This method is scalable and can be used to include other materials and extend to more layers.

3.3. Model Architecture

Our model architecture is shown in Fig. 1b, which is a standard encoder-only transformer that takes in the sequence of tokens. Each token will go through a physical embedding and positional embedding before being passed to the

Table 1: Important parameters of our OL-Transformer

NAME	PARAMETER
NUMBER OF ENCODER BLOCK (N)	12
NUMBER OF ATTENTION HEAD (A)	16
DIM OF HIDDEN STATES (H)	1024
NUMBER OF TOKENS	901
FC LAYERS	1024-1024-142
NUMBER OF PARAMETERS	65 M

Table 2: Performance of our OL-Transformer. Batch size = 1000 for batch simulation. MSE: Mean Square Error.

ATTRIBUTE	TMM	OURS	SPEEDUP
SINGLE SIMULATION (S)	0.057	0.010	∼5.7
BATCH SIMULATION (S)	-	0.000015	∼3800
MSE	-	0.000057	-

transformer model. The physical embedding plays a similar role as the input embedding in (Vaswani et al., 2017). Similar to the class token in BERT (Devlin et al., 2018) and ViT (Dosovitskiy et al., 2020), we add a 'BoS' token (Beginning of Sequence) at the 0-th input position and treat its corresponding output from the encoder as the hidden representation for the input sentence. A fully connected layer is used to decode this output into the predicted spectra $\hat{Y}$. Our model is trained by minimizing the mean square error (MSE): $L = ||Y - \hat{Y}||^2$. Masked language modeling is not used because each layer in the multilayer structure is equally important when predicting the optical properties. The training takes about one week on a single NVIDIA 3090 GPU. Table 1 lists important parameters of our model.

3.4. Dataset Generation

Each training data contains a pair of multilayer structure and its spectra. During dataset generation, the multilayer structures are created by first randomly sampling the number of layers with an increasing ratio, and then uniformly sampling the material arrangements **m** and the thickness sequence **t**. We then obtain their reflection and transmission spectra using TMM simulation. In total, we generate 10 M pairs of structures and spectra as the training dataset and another 1 M pairs for validation, which takes ∼1200 h for a single CPU. In the future analysis and comparison, we refer TMM as the accurate physical simulator and our OL-Transformer as the surrogate model.

4. Experiments

4.1. Simulaion Speedup

In Table 2, we report the performance of simulation acceleration on the validation dataset. The physical simulation of TMM is evaluated on a single 2.4GHz CPU since there is no package available on GPU. Our model of OL-Transformer is evaluated on a single NVIDIA 3090. When making predictions for a single structure at one time, on average, our surrogate model can finish each simulation ∼5.7 times faster than the TMM. After using the GPU batch calculation (batch size = 1000), our model shows ∼3800 fold time improvement compared to TMM. Therefore, our model can be used as a faster simulator for multilayer thin film structures. This can be very helpful when a large number of simulations are needed, e.g., to inverse design or to understand the physical structure-property behaviors.

4.2. Universal Surrogate Simulator

In principle, our method can extend up to 10^{25} different structures and work as a universal surrogate simulator. However, we cannot iterate and evaluate every one of them because there are so many. Therefore, in Table 2, we report the universal prediction ability on the validation dataset and evaluate the averaged MSE on 10^6 different structures (since $10^6 \ll 10^{25}$, each structure is a distinct type with different material arrangements). For a better comparison, we randomly select six structures with different **m** and report their detailed prediction performance in Table 3. For each structure, we randomly generate 1000 structures with different **t** and report their averaged MSE. Compared to the reported work which can only predict a specific structure, ours are versatile to different structures. Notice their MSE is listed only for reference because we are predicting different optical properties and cannot be compared directly. In Fig. 2, we also give three examples to visualize the difference between target spectra from TMM (real lines) and predicted spectra by our model (dashed lines). Based on the low MSE and close-to-simulator spectra, we demonstrate that our model exhibits a strong generalization ability to predict the spectra of universal types of structures, significantly expanding the capabilities of existing surrogate models.

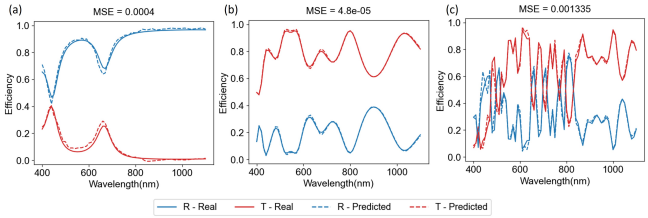

Figure 2: Three examples of predicting the transmission spectra (T) and reflection spectra (R). Title shows their MSE.

4.3. Understanding the Universal Learning Ability

To understand why our model exhibits a strong generalized learning ability, we first use t-SNE to reduce the high-dimensional physical embeddings for each token into two dimensions and visualize the results in Fig. 3. These tokens referring to the same material are marked as the same color. Clearly, we can see a transition of materials from the low

Table 3: Prediction performance on six different structures. Layers: The total number of layers in the given type of multilayer structure. Reported MSE: The reported MSE in existing work is summarized only for reference. We cannot directly compare them because the predicted optical properties are different.

DESCRIPTION OF MULTILAYER STRUCTURE	LAYERS	MSE	REPORTED MSE
$Ag/SiO_2/Ag$	3	8.2×10^{-4}	$\sim 10^{-5}$(DENG ET AL., 2021)
$MgF_2/SiO_2/Al_2O_3/TiO_2/Si/Ge$	6	4.9×10^{-4}	$\sim 10^{-6}$(CHEN ET AL., 2023)
$SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4$	6	2.6×10^{-5}	-
$TiO_2/SiO_2/Al_2O_3/Si_3N_4/ZnO/ZnS/ITO/HfO_2/Si$	9	2.5×10^{-5}	-
$ZnS/TiO_2/MgO/ZnS/Si_3N_4/ITO/SiO_2/TiO_2/Ta_2O_5/ZnO/Al_2O_3/Ag$	12	1.3×10^{-5}	-
$SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4/$ $SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4/SiO_2/Si_3N_4$	20	1.1×10^{-3}	$\sim 10^{-4}$(LIU ET AL., 2018)

refractive index region (lower left) to the high refractive index region (upper right). For each material, the increasing size of dots corresponds to the increasing thickness, e.g., 10 nm for the smallest dot and 500 nm for the largest dot. Because our model only takes in the tokens, instead of the thickness or material's properties, these observations demonstrate that our model exhibits the ability to recognize and recover the material and thickness information using hidden representations. Apart from this, intrinsic physics can also be obtained from Fig. 3. For example, these dots representing metals, e.g, Al, Ag (see zoomed view (i)), are distinguishable at small thickness but cluster together at greater thickness. This makes sense as metals with thickness greater than the penetration depth have no impact on light-matter interaction. In addition, most dielectric materials with small thickness are clustered together (see zoomed view (ii)), which reminds us of the fact that thin dielectric materials are hard to distinguish because they all have a similar impact on light propagation.

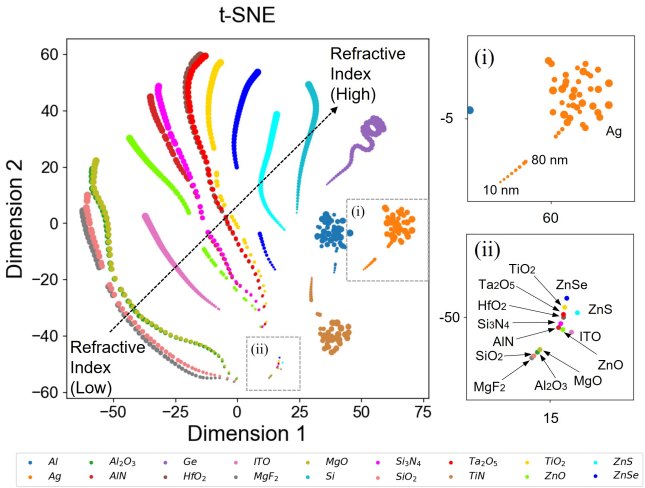

Figure 3: The 2D visualization of the hidden space using t-SNE to reduce dimension.

To further understand how our model leverages these learned embeddings for surrogate simulation, we then visualize one of the attention maps for a given multilayer structure in Fig. 4a. We also simulate and visualize the electrical field distribution from 400 nm to 1100 nm by solving the multilayer system through TMM, as shown in Fig. 4b. The attention map in Fig. 4a exhibits an obvious alternating pattern on off-diagonal elements, suggesting a high correlation in these nearby layers. Interestingly, such alternating patterns can also be observed in Fig. 4b, where the electrical field illustrates how light interacts and propagates among each layer. Intuitively, self-attention can be treated as the analogy of physical interactions inside the transformer model.

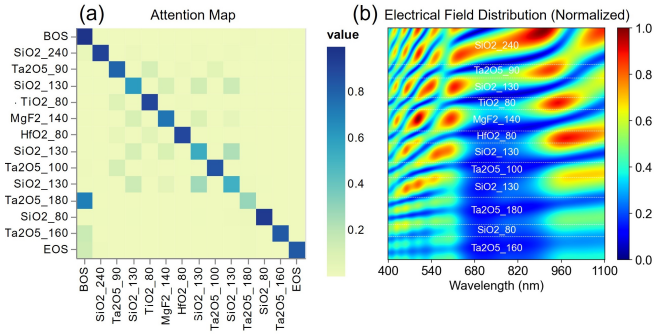

Figure 4: (a) Visualization of attention map (Block 0, Head 15). (b) The electrical field distribution.

In summary, the empirical understanding of the strong universal learning ability can be explained by 1) our model learns the unique embeddings with intrinsic physical meanings for all structure tokens and 2) our model learns to use self-attention to represent the light-matter interaction among each layer. Although there are up to 10^{25} different material arrangements, all of them can be reconstructed by 901 structure tokens. Our model learns to decompose the prediction of each type of structure into the two-step learning of physical embeddings and self-attention.

5. Conclusion

In this paper, we introduce **OL-Transformer**, a fast and universal surrogate solver for simulating the transmission and reflection spectra in optical multilayer thin film structures. Compared to existing surrogate models, our model can ex-

tend the learning capabilities from limited structures to 10^{25} different structures without adaptation used in transfer learning or meta-learning. In addition, our model still achieves a six-fold speedup compared to numerical simulation, with a potential 3800-fold when using batch calculations, facilitating the downstream applications including inverse design and understanding the structure-property behaviors. Our work expands the existing transformer applications from mainstream NLP and CV to optical simulations and demonstrates that transformer architecture (embedding and self-attention) is also an effective learner for optical physics.

6. Potential Broader Impact

In optics and photonics, understanding the structure-property relationship is vital for developing optical devices and photonic applications. By leveraging the strong generalization ability of transformer, our model can serve as a fast and universal surrogate simulator for predicting the optical properties of optical multilayer thin film structures, e.g., transmission and reflection spectra, providing an easy and straightforward to explore the light-matter interaction and inverse design for specific structures to satisfy desired optical properties. Our method can be easily scaled to predict other types of optical properties, including angled-resolved spectra and structure color, and include more materials with more complex layered structures. In addition, the way we serialize the multilayer structure as a sequence of tokens can also be directly applied to other types of photonic structures, including meta-grating structures and free-form metasurfaces, with the potential to speed up the simulation and prediction of these structures with higher complexity. On the other hand, our work also demonstrates that transformer has a strong capability of capturing intrinsic physical knowledge and using them for predicting physical behaviors, which can inspire future work to leverage transformer to solve complicated and non-trivial physical problems.

References

Broberg, J., Bånkestad, M., and Ylipää, E. Pre-training transformers for molecular property prediction using reaction prediction. *arXiv preprint arXiv:2207.02724*, 2022.

Byrnes, S. J. Multilayer optical calculations. *arXiv preprint arXiv:1603.02720*, 2016.

Chen, W., Gao, Y., Li, Y., Yan, Y., Ou, J.-Y., Ma, W., and Zhu, J. Broadband solar metamaterial absorbers empowered by transformer-based deep learning. *Advanced Science*, pp. 2206718, 2023.

Deng, Y., Dong, J., Ren, S., Khatib, O., Soltani, M., Tarokh, V., Padilla, W., and Malof, J. Benchmarking data-driven surrogate simulators for artificial electromagnetic materials. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Fu, N., Wei, L., Song, Y., Li, Q., Xin, R., Omee, S. S., Dong, R., Siriwardane, E. M. D., and Hu, J. Material transformers: deep learning language models for generative materials design. *Machine Learning: Science and Technology*, 4(1):015001, 2023.

Geneva, N. and Zabaras, N. Transformers for modeling physical systems. *Neural Networks*, 146:272–289, 2022.

Hugonin, J. P. and Lalanne, P. Reticolo software for grating analysis. *arXiv preprint arXiv:2101.00901*, 2021.

Jiang, J., Chen, M., and Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. *Nature Reviews Materials*, 6(8):679–700, 2021.

Kaya, M. and Hajimirza, S. Using a novel transfer learning method for designing thin film solar cells with enhanced quantum efficiencies. *Scientific reports*, 9(1):1–10, 2019.

Khorrami, P., Simek, O., Cheung, B., Veillette, M., Dangovski, R., Rugina, I., Soljacic, M., and Agrawal, P. Adapting deep learning models to new meteorological contexts using transfer learning. In *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4169–4177. IEEE, 2021.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Liu, D., Tan, Y., Khoram, E., and Yu, Z. Training deep neural networks for the inverse design of nanophotonic structures. *Acs Photonics*, 5(4):1365–1369, 2018.

Liu, M., Johnston, M. B., and Snaith, H. J. Efficient planar heterojunction perovskite solar cells by vapour deposition. *Nature*, 501(7467):395–398, 2013.

Ma, T., Wang, H., and Guo, L. J. Optogpt: A foundation model for inverse design in optical multilayer thin film structures. *arXiv preprint arXiv:2304.10294*, 2023.

Qu, Y., Jing, L., Shen, Y., Qiu, M., and Soljacic, M. Migrating knowledge between physical scenarios based on artificial neural networks. *ACS Photonics*, 6(5):1168–1174, 2019.

Sun, Y., DeJaco, R. F., Li, Z., Tang, D., Glante, S., Sholl, D. S., Colina, C. M., Snurr, R. Q., Thommes, M., Hartmann, M., et al. Fingerprinting diverse nanoporous materials for optimal hydrogen storage conditions using meta-learning. *Science Advances*, 7(30):eabg3983, 2021.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wang, D., Liu, Z., Wang, H., Li, M., Guo, L. J., and Zhang, C. Structural color generation: from layered thin films to optical metasurfaces. *Nanophotonics*, 12(6):1019–1081, 2023.

Zhang, J., He, Y., Li, Y.-W., Wen, C.-K., and Jin, S. Meta learning-based mimo detectors: Design, simulation, and experimental test. *IEEE Transactions on Wireless Communications*, 20(2):1122–1137, 2020.

Zheludev, N. The life and times of the led—a 100-year history. *Nature photonics*, 1(4):189–192, 2007.

Zhong, L., Wu, B., and Wang, Y. Accelerating physics-informed neural network based 1d arc simulation by meta learning. *Journal of Physics D: Applied Physics*, 56(7): 074006, 2023.