

k -space resolved surrogate models for efficient inverse design of optical metasurfaces

Damien Maitre^{1,2}, Francesco Maria Guadagnuolo¹, Jean-Philippe Banon², Maxime Darnon², Raphaël Clerc², Loumi Tremas¹, Mathys Le Grand¹, Adam Fuchs¹, Pascal Urard¹, Coumba Gaye¹, Olivier Jeannin¹, James Downing³, Denis Rideau¹

¹ STMicroelectronics, 850 rue J. Monnet BP16, 38926 Crolles, France

² Univ. Jean Monnet, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France

³ STMicroelectronics, 1 Tanfield, Edinburgh EH3 5DA, United Kingdom

damien.maitre@st.com, denis.rideau@st.com

Abstract—Inverse design of optical metasurfaces typically relies on computationally expensive full-wave solvers that constrain the search for optimal designs. Neural-network surrogates can accelerate this process but are often built on finely sampled real-space grids, leading to unnecessarily high-dimensional inputs and outputs. We instead introduce k -space resolved surrogate models that represent both permittivity and fields in a truncated Fourier basis determined from RCWA convergence. We validate the approach on an RGB color router inverse design, showing that the k -space surrogate reliably drives optimization toward high-efficiency metasurface consistent with full RCWA simulations.

Keywords—Optical metasurface, inverse design, neural operator

I. INTRODUCTION

Designing optical metasurfaces with specified functionalities is a central challenge in nanophotonics, with applications ranging from flat optics and holography to imaging and beam shaping [1,2]. These devices rely on spatially varying arrangements of subwavelength “meta-atoms” to tailor the amplitude, phase, and polarization of light. Therefore, inverse design—determining an optimal layout for a desired optical response—becomes a key task.

State-of-the-art workflows [3,4] typically rely on global or local optimization algorithms coupled with rigorous electromagnetic solvers. These methods, such as Rigorous Coupled-Wave Analysis (RCWA) [5] or Finite Difference Time Domain (FDTD) [6], are accurate but computationally expensive. By changing the objective function or the design constraints, it usually requires re-running large batches of costly simulations, thereby limiting the range of designs that can be explored.

Neural-network-based surrogate models, also known as neural operators [7], offer a way to alleviate this bottleneck by learning the mapping between the geometry of the metasurface and its optical response. These networks, trained from precomputed data, replace the full-wave solver inside the optimization loop [8,9]. A variety of model architectures have been proposed for operator learning, ranging from the use of distinct subnetworks in DeepONets [10], to cascades of convolutional kernels in CNOs [11], spectral convolutions in FNOs [12] that leverage Fourier truncation of the underlying function, and more recent transformer-based models such as Transolver [13], which benefit from efficient transformer blocks [14]. Following these two later approaches and based on the Fourier-space structure imposed by RCWA convergence, we propose a physically grounded representation for neural-network modeling of metasurfaces. We define an optimal sampling grid in reciprocal space by choosing the minimal set of spatial frequencies required for an accurate inverse design of the targeted application. This yields a compact description of both permittivity and its corresponding fields, then reducing input and output dimensionality without sacrificing accuracy. Using this physically efficient representation, we train a Vision Transformer (ViT)-based architecture [15] to predict the optical response, embedding RCWA-informed spectral content as a physically meaningful prior in the surrogate model (Fig. 1).

We test our approach on the inverse design of a “color router” device [16,17]. This metasurface spatially separates optical signals of different wavelengths by focusing them onto distinct focal points. Such a functionality is highly sensitive to both amplitude and phase control and inter-pillar coupling. By replacing RCWA solver with our k -space surrogate model within the optimization loop, we obtain high-efficiency color routing at a substantially lower computational cost. Compared

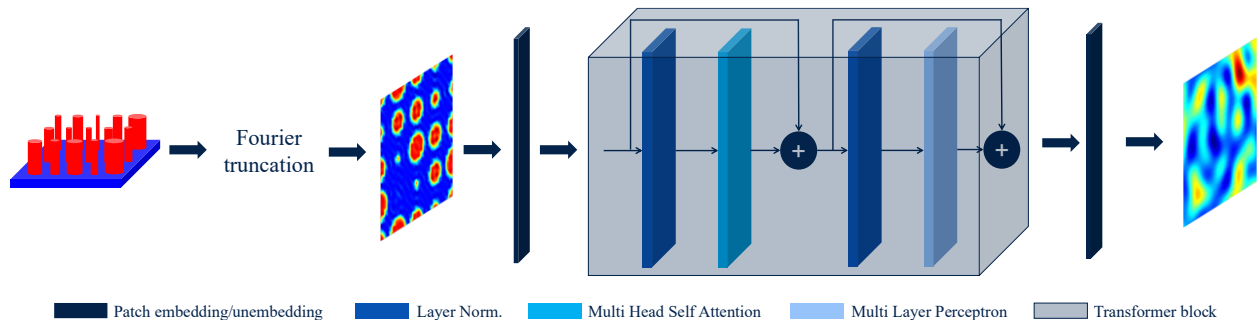


Fig. 1. Architecture for the simulation of optical metasurface leveraging Fourier space structure from RCWA and efficiency of neural networks surrogate models. The metasurface description is first truncated on a finite set of Fourier modes and fed into a Vision-Transformer-like architecture comprising patch embedding [15], transformer blocks [14], and patch unembedding, after which the output field is predicted directly on this Fourier truncation. Both input and output are not directly in k -space, but inverse transformed back to real space.

with conventional RCWA-driven optimization, our method identifies high-performance designs more efficiently while preserving accuracy, demonstrating the practical relevance of the proposed framework.

Finally, we analyze the key design choices underlying the proposed framework. In particular, we investigate how the database size affects prediction accuracy and optimization performance. This analysis confirms that maintaining high fidelity between the surrogate model and RCWA is of course essential for efficient and reliable optimization.

II. *K*-SPACE RESOLVED SURROGATE MODEL

The core idea of RCWA [5] is to represent both the permittivity distribution and the electromagnetic fields as Fourier series along the periodic directions, and to solve Maxwell's equations in this basis. For a periodic metasurface along the transverse plane, with in-plane coordinates $\mathbf{r}_{\parallel} = (x, y)$ the permittivity is written as

$$\varepsilon(\mathbf{r}_{\parallel}, z) \approx \sum_{\mathbf{G} \in \mathcal{G}_N} \varepsilon_{\mathbf{G}}(z) e^{i\mathbf{G} \cdot \mathbf{r}_{\parallel}}, \quad (1)$$

where $\mathbf{G}$ are the reciprocal lattice vectors and $\varepsilon_{\mathbf{G}}(z)$ the Fourier coefficients. This series is then truncated to a finite number N of Fourier modes $\mathcal{G}_N$. The maximum spatial frequency is set by this truncation and therefore determines the accuracy of the solution. In our surrogate model, we use this truncated reciprocal-space description directly restricted to the same set $\mathcal{G}_N$ chosen to achieve sufficient accuracy for the numerical solution. The network then learns the mapping between the truncated permittivity and the electromagnetic field.

We employ a Vision Transformer inspired architecture (Fig. 1) to model the relationship between the metasurface input representation and the corresponding optical response. The network takes as input a two-channel image, where the two channels correspond to the real and imaginary parts of the complex permittivity ε . A periodic extension of 3 pixels is first applied on each boundary, designed to respect the periodic nature of the unit cell. The extended field is then partitioned into non-overlapping 5×5 patches, and each patch is linearly projected into a 128-dimensional token embedding. The core of the ViT consists of 6 Transformer blocks [14], each using 8 self-attention heads and a feed-forward multilayer perceptron with 512 neurons. The latent representation is subsequently decoded back to the spatial domain through an unembedding layer to predict a two-channel output corresponding to the real and imaginary parts of the electromagnetic field at the original resolution. The model is trained by minimizing the mean squared error (MSE) loss using the AdamW optimizer, with an initial learning rate of 3×10^{-4} and a cosine annealing schedule. To improve generalization, a dropout rate of 0.1 is applied throughout the network. Overall, the architecture contains approximately 1.2×10^6 trainable parameters.

We fix the RCWA-informed Fourier truncation to $N = 29$ modes, based on a prior convergence study. Trained for 500 epochs on a dataset of 48,000 randomly generated samples, varying the radii of the nano-scatterers, the proposed ViT-based models achieves a MSE of 3.20×10^{-2} , 3.07×10^{-2} , and 1.40×10^{-2} , with corresponding R^2 scores of 96.7%, 96.9%, and 98.6% for the blue, green, and red wavelengths, respectively. These results demonstrate the

ability of the model to accurately predict the optical response of the metasurface.

III. COLOR ROUTER OPTIMIZATION

We apply the proposed model to the inverse design of an RGB color router consisting of a 4×4 array of TiO_2 circular nanopillars embedded in SiO_2 . Such metasurface-based devices have been introduced as a promising alternative to conventional CMOS image sensors, which typically rely on microlenses placed above the pixel array [18]. By contrast, these planar optical elements enable compact and fully integrated spectral sorting at the pixel level by directly routing the different wavelengths to distinct locations on the focal plane. In this work, the objective is to optimize the geometry of the nanopillars so that three target wavelengths, corresponding to red, green, and blue, are focused on three separate focal spots with minimal crosstalk and high routing efficiency.

For each wavelength, we use the trained ViT models within an NSGA-III [19] optimization loop. The loss function is expressed as

$$L = \sum_{\lambda=R,G,B} \iint_{xy\text{-plane}} S_z(\lambda) \odot W(\lambda), \quad (2)$$

where S_z is the Poynting vector at the target focal plane and W a weighting function. By choosing W a binarized matrix

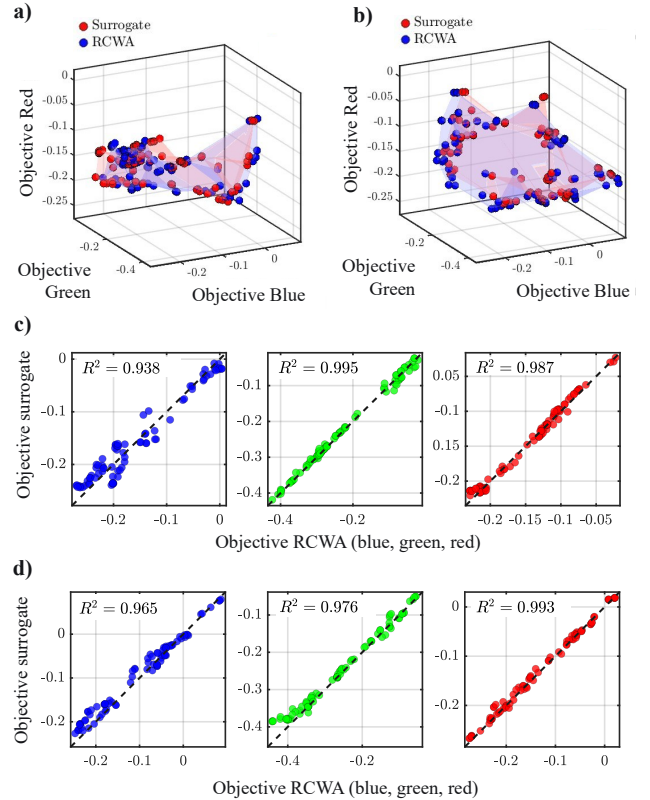


Fig. 2. Comparison of Pareto fronts (PF) obtained from surrogate-based (a) and RCWA-based (b) optimizations. In both cases, the PFs are evaluated using RCWA (blue) and the surrogate model (red), demonstrating the consistency of the surrogate in these two configurations. (c) is a flatten representation of the data in (a), and (d) from the data in (b), comparing the objective values for both configurations and thereby validating these results. Objective values are relative to the loss terms driving the optimization (eq. 2)

with ones over the targeted areas, this loss function is formulated to balance high transmission into the desired foci while maintaining a low optical crosstalk.

Figure 2 compares the optimized Pareto Fronts (PF) obtained using two optimization strategies. The first relies on the proposed surrogate model, whereas the second performs the same optimization using direct RCWA evaluations. The final PF are visible in Fig. 2a and 2b respectively. Both PF are then recomputed using the other method to check the consistency of the surrogate model on these 2 approaches. The close alignment between the PF highlights the effectiveness of the surrogate model in the two scenarios. In the first case, when the model drives the optimization, it indeed finds efficient designs that are still consistent with RCWA simulations. In the second one, compared with the RCWA-optimized PF, the surrogate remains highly accurate, demonstrating that the model reliably captures the design space. Figures 2c and 2d are in a flatten representation of these optimized PFs, comparing the objectives values computed by the full-wave solver against the predicted ones. Again, the strong fit close to $y = x$ confirms the consistency of the model and its reliability for driving an optimization.

We then pick one of the optimized designs found by the surrogate model. On this final design, as shown in Fig. 3, the surrogate predicts well-localized focal spots at the targeted positions with limited crosstalk. To validate this surrogate-optimized design, we then compute from full-wave RCWA simulations the full Poynting vector. The close match between solver results and the model predictions, in terms of focal spots, intensities, and total power delivered to the target foci, confirms that the surrogate-based optimization reliably transfers to rigorous electromagnetic simulations.

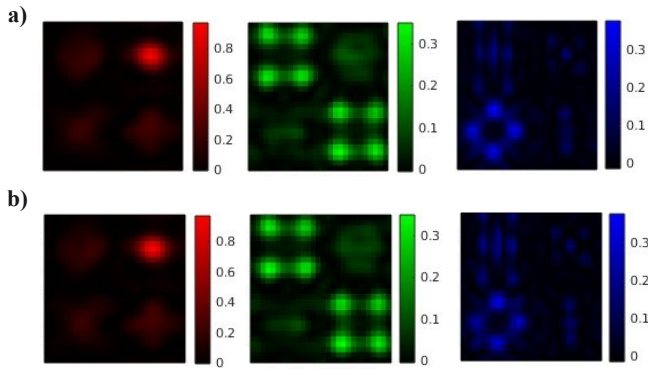


Fig. 3. z-component of the Poynting vector (RGB) of an optimal color router selected on the Pareto Front after optimization with the surrogate model. (a) prediction from the surrogate network; (b) recomputed fields using RCWA. Mean absolute errors for red, green and blue are respectively $5.7 \cdot 10^{-3}$, $5.5 \cdot 10^{-3}$ and $8.9 \cdot 10^{-3}$. The small residual differences on the effective router design indicate an excellent agreement between the model and the full-wave RCWA simulations.

Finally, we assess the ability of the surrogate model to discover designs out of the database distribution: beyond those seen during the training. As shown in Fig. 4, the Pareto front extracted from the training database is compared with the one obtained through surrogate-based optimization. The surrogate-generated solutions consistently outperform the best designs present in the database, yielding a Pareto front that entirely dominates the database Pareto front. These results demonstrate that the model is able to identify new high-

performing designs while remaining consistent with full-wave RCWA validation

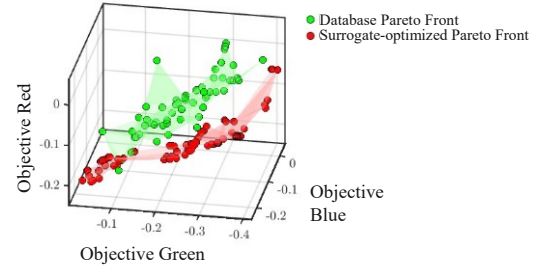


Fig. 4. Comparison of Pareto fronts (PF) obtained from the surrogate-based optimization (red) and the generated database (green). Corresponding hypervolumes are 2.11 and 1.96. The model-optimized PF dominating the database PF highlights the performance of the model to optimize and generalize out of the database distribution while still being consistent with full-wave simulation.

IV. HIGH FIDELITY MODEL

We next investigate the influence of the training-set size on the predictive performance of the surrogate model by considering nine different data splits, ranging from 10% to 90% of the database used for learning. The corresponding models are evaluated on the validation set using the R^2 score, as reported in Fig. 5. As expected, the prediction performance improves as the amount of training data increases, in agreement with standard observations in deep learning [20] and, more specifically, transformer-based architectures [21]. However, an eventual saturation of this improvement (around 70/30 split) reveals the existence of a threshold beyond which additional data no longer provide a significant gain. This result indicates that the surrogate model captured the optical response of the design space with sufficient accuracy.

The identified threshold provides a practical criterion for deciding whether a surrogate-assisted optimization is worthwhile. More specifically, the surrogate becomes advantageous when the number of full-wave simulations required by a conventional optimization exceeds the number of simulations needed to train an accurate surrogate. Since the training time of the neural network itself is negligible compared with the cost of full-wave simulations, this threshold is primarily determined by the cost of generating the training database. Conversely, if an optimization converges after only a small number of full-wave evaluations, direct simulation remains the most efficient approach. In practice, however, a surrogate is rarely trained for a single optimization

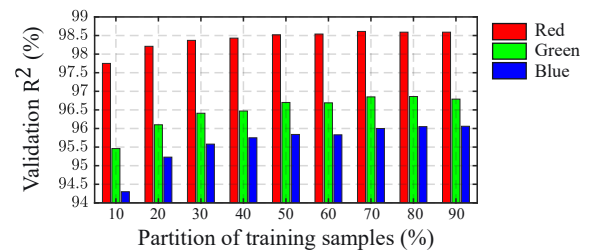


Fig. 5. Accuracy of the surrogate models for different training/validation splits. Precision is measured with R^2 score on validation set. Overall good accuracy is already reached with a few samples with a reached plateau after a certain point.

task. Once trained, the same model can be reused for multiple optimization campaigns involving different objective functions or optimization settings, allowing the initial training cost to be amortized over many design studies.

We then investigate the sufficient level of accuracy required for reliable optimization. To this end, the comparison presented in Fig. 2d is repeated using surrogate models trained with different train/test splits, resulting in different prediction accuracies. For each model, the predicted PF are compared with the corresponding RCWA evaluations on both the database PF and the RCWA-optimized PF, considering the blue wavelength as the most challenging case (Fig. 6 a,b). A second-order polynomial fit is applied in each case, with the ideal relation corresponding to the line $y = x$.

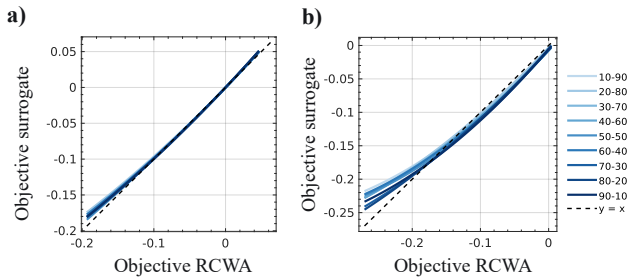


Fig. 6. Comparison of Pareto Fronts computed from RCWA solver against model prediction for different training/test repartition. Comparison is performed on the database PF (a) and the RCWA-optimized PF (b) for the blue wavelength. A second-order polynomial fit is applied for each models, ranging from lowest repartition (light blue) to highest repartition (dark blue). Close fit to $y = x$ target on database PF with lose of accuracy in better optimized points in RCWA-optimized PF highlights needs of highly accurate models to drive the optimization.

As the prediction accuracy decreases, relatively to the different models, the fitted curves progressively deviate from the ideal agreement. This behavior is particularly strong in the high-performance region of the PF (left corner). Although this tail effect is visible on both PF, it is more pronounced for the post-optimization one, highlighting the increased sensitivity of optimized designs to prediction errors. Rather than relying solely on conventional prediction metrics, this analysis provides a practical criterion for assessing whether a surrogate has reached the accuracy required to reliably guide the optimization process. It also emphasizes that a surrogate intended for inverse design must remain accurate not only over the entire design space, but especially in the vicinity of the optimal solutions.

V. CONCLUSION

In conclusion, we have shown that encoding both the metasurface permittivity and the corresponding electromagnetic fields on an RCWA-defined k -space truncation provides an efficient and physically meaningful representation for surrogate modeling. This representation, combined with a Vision Transformer architecture, enables accurate prediction of the optical response while significantly reducing the computational cost of inverse design. Using the

RGB color router as a representative example, we demonstrated that the proposed framework efficiently identifies high-performance metasurface designs, while remaining consistent with full-wave simulations. More broadly, this approach provides a practical methodology for accelerating the optimization of complex photonic devices through accurate surrogate models.

REFERENCES

- [1] Hu et al., "A Review on Metasurface: From Principle to Smart Metadevices", *Frontiers in Physics*, 2021
- [2] Zou et al., "Advanced optical imaging based on metasurfaces", *Advanced Optical Materials*, 2024
- [3] Jeong et al., "A Tutorial on Inverse Design Methods for Metasurfaces", *Curr. Opt. Photon.* 2024
- [4] Maitre et al., "Inverse design of optical metasurface for CMOS imagers: a multi-objective optimization approach", *SISPAD*, 2025
- [5] Whittaker and Culshaw, "Scattering-matrix treatment of patterned multilayer photonic structures", *Physical Review B*, 1999
- [6] Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media". *IEEE Transactions on Antennas and Propagation*, 1966
- [7] Kovachki et al. "Neural Operator: Learning Maps Between Function Spaces With Applications to PDEs", *JMLR*, 2023
- [8] Abdelraouf et al., "From maxwell's equations to artificial intelligence: The evolution of physics-guided AI in nanophotonics and electromagnetics", *Optics and Laser Technology*, 2025
- [9] Rideau et al., "Enabling Meta-surface Design with Neural Network Technologies", *META*, 2026
- [10] Lu et al., "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators", *Nature Machine Intelligence*, 2021
- [11] Raonic et al., "Convolutional Neural Operators for robust and accurate learning of PDEs", *NIPS*, 2023
- [12] Li et al., "Fourier neural operator for parametric partial differential equations", *ICLR*, 2021
- [13] Wu et al. "Transolver: A Fast Transformer Solver for PDEs on General Geometries", *PMLR*, 2024
- [14] Vaswani et al., "Attention Is All You Need", *NIPS*, 2017
- [15] Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", *ICLR*, 2021
- [16] Zhao et al., "Perfect RGB-IR Color Routers for Sub-Wavelength Size CMOS Image Sensor Pixels", *Adv. Photonics Res.* 2021
- [17] Choi et al., "Optical design of dispersive metasurface nano-prism structure for high sensitivity CMOS image sensor", *IEDM*, 2023
- [18] Yun et al., "Highly Efficient Color Separation and Focusing in the Sub-micron CMOS Image Sensor", *IEDM*, 2021
- [19] Deb et al., "An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints," *IEEE Transactions on Evolutionary Computation*, 2014
- [20] Rosenfeld et al., "A constructive prediction of the generalization error across scales", *ICLR*, 2020
- [21] Tay et al., "Scaling Laws vs Model Architectures: How does Inductive Bias Influence Scaling?", *arXiv:2207.10551v1*, 2022