Inverse design of optical metasurface for CMOS imagers: a multi-objective optimization approach

Damien Maitre[1,3], Denis Rideau[1], Olivier Jeannin[1], Clémence Jamin-Mornet[2], Charly Leblanc[2], Maxime Darnon[3], Raphaël Clerc[3], Jean-Philippe Banon[3], Coumba Gaye[1], Fatima Omeis[1], Louis-Henri Fernandez-Mouron[1], Loumi Tremas[1], Mathys Le Grand[1], Adam Fuchs[1], Pascal Urard[1], James Downing[1], Bruce Rae[1]
[1] STMicroelectronics, 850 rue J. Monnet BP16, 38926 Crolles, France
[2] Minatec Campus, Leti, CEA, Grenoble, 38054, France
[3] Univ. Jean Monnet, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France
damien.maitre@st.com , denis.rideau@st.com

***Abstract*—The design of optical metasurfaces for CMOS imagers is challenging due to the influence numerous physical parameters such as pattern size, density, and layer thicknesses. In this work, an original inverse design methodology based on multi-objective and high-dimensional global optimization is proposed and applied to the design of an RGB-IR color router, achieving significant performance enhancements compared to conventional refractive microlens pixels. This approach provides insights into the design process and establishes a framework for future advances in CMOS imager technology.**

Keywords—Optical metasurface, Color Router, High Dimensional Optimization, Multi-Objective Optimization

I. Introduction

Optical metasurfaces for CMOS imagers are receiving growing interest from the electron device community [1][2], although their design faces significant challenges. Indeed, it involves the optimization of numerous physical parameters, such as the shape, size, and arrangement of nano scatterers, as well as the number and thicknesses of layers. Each of these parameters directly influences the optical response of the metasurface, including its efficiency, spectral selectivity, and angular performance. Furthermore, these parameters are often highly interdependent, leading to a complex, non-linear design space with multiple local minima. This task becomes even more challenging when considering fabrication constraints and when the metasurface operates over a broad spectral range. Therefore, advanced numerical optimization algorithms are essential to efficiently explore the high-dimensional parameter space, at the cost of computationally intensive simulations.

We present here a method for the design of optical metasurfaces. Following the approach proposed in [3] on RGB metasurface-enhanced imagers (color routers), more complex RGB and IR stacked metasurface designs are investigated and alternative optimization algorithms are examined. These developments could serve to improve the performance of architectures such as the monolithic RGBZ sensor [4], while addressing RGB decomposition and IR routing through a metasurface design compatible with fabrication constraints.

To design an RGB-IR color router, we compare several optimization strategies, including local gradient-based approaches and global optimization algorithms (Bayesian and metaheuristics), as well as single and multi-objective functions. While recent research [5][6] mostly evaluates these algorithms using single-objective functions, which can obscure multiple sub-objectives, our study addresses the challenge of multi-objective functions. We highlight the benefits of considering multiple objectives to achieve a design of an RGB-IR sensor that closely aligns with potentially complex product specifications.

Previous research has explored the optimization of RGB-IR color routers through various approaches. For example, Zhao [7] proposed a theoretically ideal 3D metastructure optimized using the adjoint method, although its fabrication presents significant challenges. Zhong [8] and Hong [9] focused on single-layer metasurfaces using the local phase approximation method, with the former targeting visible-to-IR routing and the latter addressing the separation of four wavelengths. Similarly, Peng [10] introduced a complex 3D metastructure with QR-code-like stacked layers, optimized using a multi-objective approach, though it also poses fabrication difficulties. To the best of our knowledge, most algorithmic optimization studies for RGB-IR color routers have relied on single-objective functions, except for Peng [10]. However, no prior work has provided a comprehensive comparison between single-objective and multi-objective optimization approaches. In this study, we emphasize the benefits of adopting a multi-objective optimization framework by directly comparing it with the single-objective approach. This comparison demonstrates the capability of the multi-objective method to offer a broader range of performance trade-offs, providing greater flexibility in meeting diverse design requirements.

II. RGB-IR Color Router Design

A. Color router architecture

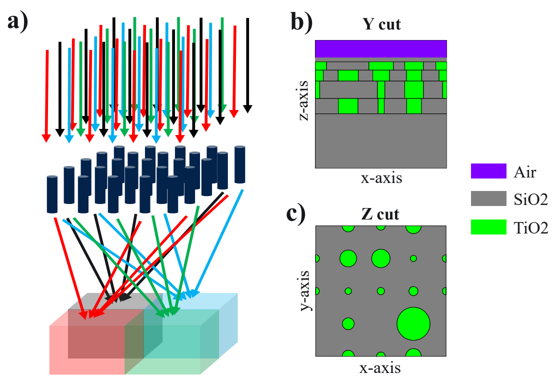

Fig. 1. Operational principle of the 3D nanostructured color router and cross-sectional views of it. a) The operational principle illustrates how a simplified metasurface directs light to different channels based on wavelength. b) The Y cut shows a layer of air (purple), an anti-reflective SiO_2 layer, 4 stacked metasurfaces, and a focal length SiO_2 layer. c) The Z cut illustrates one of the 4 metasurfaces, composed of a 4x4 array of circular TiO_2 nano-scatterers on a grid and embedded in a SiO_2 layer.

The design of the RGB-IR color router, as illustrated in Fig. 1, involves the optimization of many parameters (up to 70, including the widths and thicknesses of 4×4×4 subwavelength nano-scatterers). Simulations were conducted using an in-house 3D Rigorous Coupled-Wave Analysis (RCWA) solver, with Perfectly Matched Layer (PML) boundary conditions applied along the z-axis and periodic boundary conditions in the xy-plane. As an additional feature for minimizing the crosstalk between the four wavelengths, a typical Bayer layer of resist is added after the focal plane. However, this aspect will not be the main focus of our study, as we assume ideal filters in the analysis.

B. *Objective function*

To address the optimization challenge, the common approach consists of an objective function (to be minimized) and a dedicated optimization algorithm. In single-objective approaches, the four distinct objective functions $Loss_\lambda$ corresponding to blue ($\lambda = 450\ nm$), green ($\lambda = 550\ nm$), red ($\lambda = 650\ nm$), and infrared ($\lambda = 940\ nm$) spectral range are typically combined into a single scalar *Loss* function. The objective function used in this study, which extends the one introduced in [11] to accommodate RGB-IR operation, is detailed in equations (1,2).

$$\text{Loss} = \sum_{\lambda=R,G,B,IR} (\text{T}_\lambda \times \text{Loss}_\lambda \text{ -1})^2 \qquad (1)$$

$$\text{Loss}_\lambda = \frac{\iint_{xy-plane} [P_z(\lambda) \cdot W(\lambda)]}{\iint_{xy-plane} P_z(\lambda)} \qquad (2)$$

where T_λ is the transmittance of the color router for a given wavelength. $P_z(\lambda)$ is the component along the z-axis of the Poynting vector evaluated at the interface above the resists for the same wavelength. $W(\lambda)$ is a Gaussian weighting function that emphasizes light focusing on the center of each pixel.

C. *Optimization approaches*

Minimizing this single-objective function implicitly assumes equal importance among the four objectives. To address this limitation, a multi-objective optimization approach based on Pareto Front (PF) exploration is employed. This strategy provides a set of optimal trade-offs, allowing the designer to select the most suitable solution with respect to product specifications. By evaluating algorithms within both single and multi-objective frameworks, we identify the respective strengths and limitations of each approach, aiming to establish a robust methodology for optical metasurface design.

III. Results

A. *Gradient-based local optimization*

Gradient-based approach is one of the most used approaches when dealing with objective function minimization. Recent advancements in artificial intelligence have significantly contributed to the growing popularity of these algorithms. In particular, the development of automatic differentiation (autodiff), a technique that efficiently and accurately computes exact derivatives of functions, has played a key role. Autodiff libraries have recently been integrated into electromagnetic solvers, enabling precise computation of a system's response and its gradients. These capabilities provide clear directions for optimizing input parameters with respect to the targeted performance objectives.

For the design of our color router using gradient-based approach we perform an optimization of the metasurface using the loss function presented in eq. (1,2). We employ FMMAX [12], an autodifferentiable RCWA solver implemented in JAX, in conjunction with the well-known Adam optimizer for gradient descent. Fig. 2a illustrates the optimization process until convergence of the algorithm. The observed trend demonstrates the method's efficiency in minimizing the objective function in fewer than 200 function evaluations. However, although the objective function assigns equal weights to each wavelength, the final contributions shown in Fig. 2b vary because the algorithm optimizes the objective function without explicitly controlling their relative influences.

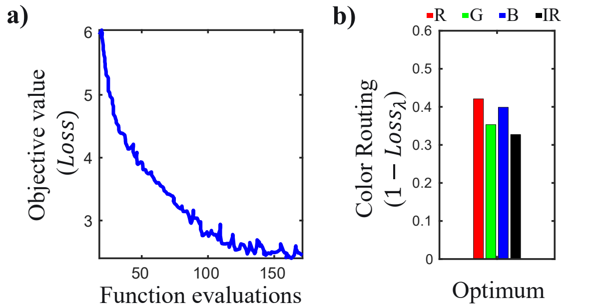

Fig. 2. Single-objective optimization process using gradient-based algorithm a) Objective value as a function of iterations during optimization process demonstrating typical function minimization. b) Color routing efficiency for each wavelengths for the optimum found by the algorithm. Color routing is quantified by the objective value for each wavelength, expressed as $(1 - Loss_\lambda)$.

This variation in performance across wavelengths is thus reflected in the resulting design characteristics. Fig. 3a presents the Poynting vector distribution for the four targeted colors, while Fig. 3b illustrates the optical efficiency across the RGB-IR spectrum. This design found by the algorithm, which effectively routes light to the targeted focal points, exhibits high routing efficiency in red and blue regions, compensating for a lower performance in the IR region.

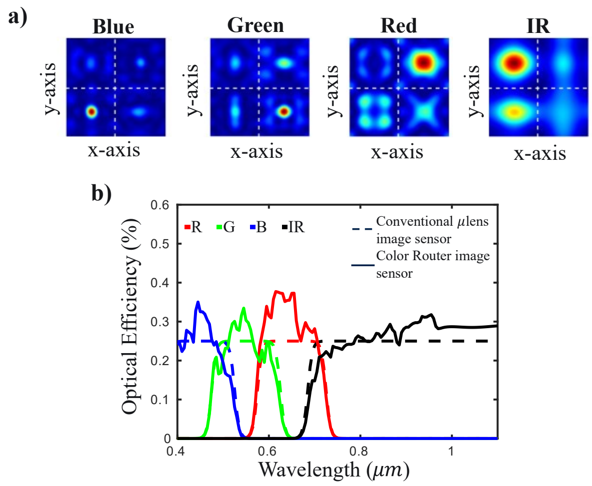

Fig. 3. Optical perfomances of the solution found by gradient-based algorithm a) Poynting vector distribution of the solution for different wavelengths (Blue, Green, Red, IR). b) Optical efficiency across various wavelengths for the solution compared to ideal performances of conventional microlens pixel (25%). Both solutions are filtered with ideal resits.

If this behavior fails to align with the desired outcome, multiple restarts of the optimization procedure may be required, along with an empirical adjustment of the weights for the four terms. However, this approach does not guarantee a comprehensive exploration of the solution space. Furthermore, it is sensitive to the choice of the initial conditions, which are determined randomly. If the initial conditions do not vary significantly, the algorithm is likely to converge to the same local minima. As a result, while this method is highly efficient for exploiting specific regions of the design space, it is less suitable for problems requiring a more global exploration of the parameter space to uncover a wider range of potential solutions.

B. Single-objective global optimization

Following a recent benchmark of high-dimensional Bayesian optimization algorithms, which included more than 20 parameters [13], and demonstrated the superior performance of TurBO [14], we adopted this approach to optimize the RGB-IR color router. Trust Region Bayesian Optimization (TurBO) is a scalable Bayesian optimization algorithm designed for high-dimensional problems. It operates by dividing the search space into localized Trust Regions (TRs), where a surrogate model, such as Gaussian Process, is used to balance exploration and exploitation. The size of each TR is dynamically adjusted based on optimization success: regions expand when improvements are found and shrink otherwise. This strategy enables efficient optimization by focusing on promising areas of the search space.

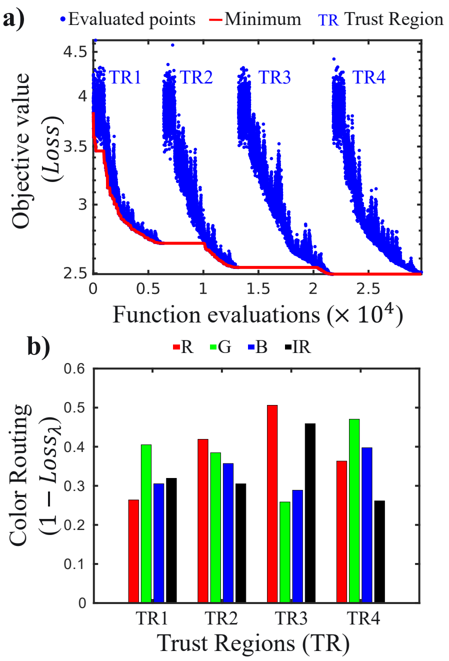

Fig. 4. Single-objective optimization process of TurBO algorithm a) Objective value as a function of iterations during optimization process demonstrating typical exploitation-exploration mechanisms of TurBO algorithm. b) Color routing efficiency for each wavelengths in the 4 Trust Regions explored by TurBO. Color routing is quantified by the objective value for each wavelength, expressed as $(1 - Loss_\lambda)$.

The results for each TR are presented in Fig. 4a, which shows clear decreases in the objective value as the number of evaluations increases. After nearly 30,000 function evaluations, the optimizer identified four distinct solutions, each located in a different TR. These four solutions are illustrated in Fig. 4b. While the solution in TR3 yields the lowest objective value and maximizes both the red and IR responses, TR2 offers a better trade-off among all four wavelengths. This analysis demonstrates that, although TurBO selects TR3 based on its minimal objective value, other TRs offer alternative solutions with different trade-offs.

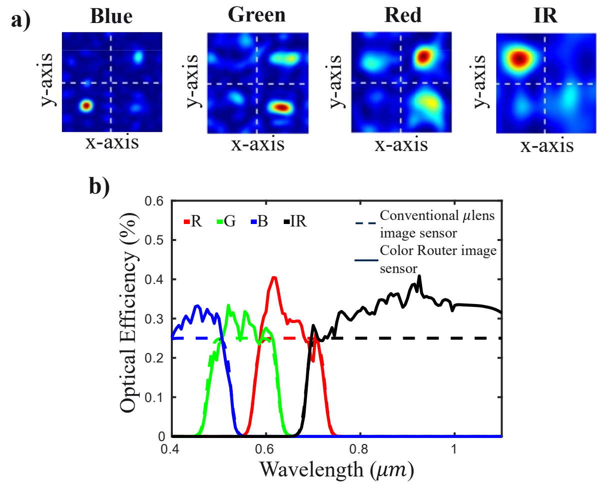

Fig. 5. Optical perfomances of the solution found by TurBO algorithm a) Poynting vector distribution of the solution found in the second trust region for different wavelengths (Blue, Green, Red, IR). b) Optical efficiency across various wavelengths for the solution found in the second trust region compared to ideal performances of conventional microlens pixel (25%). Both solutions are filtered with ideal resits.

Fig. 5a presents the solution in TR2, showing the Poynting vector distribution for the four targeted wavelengths. Fig. 5b shows the optical efficiency across the RGB-IR spectrum. As the solution found by the gradient-based approach, we observe that, for all considered wavelengths, the efficiency exceeds the one of conventional micro lens image sensors (25%). However, while TurBO offers a few solutions corresponding to different performance levels, it is important to highlight that for both methods relying on a single-objective function, the approach inherently limits the flexibility to identify a solution that best meets specific product requirements. This limitation arises because the contributions of the four RGB-IR channels are determined empirically using *ad-hoc* weights, which may not fully capture the trade-offs needed for targeted performance.

C. Multi-objective global optimization

a) General case

To overcome the limitations of single-objective optimization, we adopted here a multi-objective strategy. In this kind of approach, we intend to discover a Pareto front of solutions. This concept essentially reflects the following statement: a PF is a set of solutions that are not dominated by other solutions. This means that when we intend to minimize k conflicting objective functions with $f_i(x)$ the i-th objective function, and x being a vector of parameters in the design space X, a solution x_1 dominates another solution x_2 (denoted as $x_1 \prec x_2$) if and only if equations (3,4) are satisfied.

$$f_i(x_1) \leq f_i(x_2) \text{ for all } i \in \{1,2,\dots,k\} \quad (3)$$

$$f_i(x_1) < f_i(x_2) \text{ for at least one } i \quad (4)$$

Solutions that are not dominated by any other solution are considered Pareto-optimal. The set of all Pareto-optimal solutions forms the PF, which represents the trade-off surface in the objective space. It can be expressed as in equation (5).

$$PF = \{f(x) : x \in X, \nexists\ x' \in X \text{ s.t. } f(x') \prec f(x)\} \quad (5)$$

where $f(x)$ is the vector of objective function values. The PF then provides a set of optimal trade-offs, enabling the choice of a solution based on preferences for the conflicting objectives.

b) Color Router application

For the design of our color router, we will use a well-known multi-objective optimizer implemented in pymoo framework [15]: NSGA-III [16]. This algorithm, derived from genetic algorithm, operates by iteratively generating populations of solutions and ranking them using non-dominated sorting to identify Pareto-optimal candidates. It incorporates a reference-point-based selection mechanism, which ensures diversity by projecting solutions onto predefined reference points in the objective space, making it particularly suitable for problems with many conflicting objectives, such as color-routing trade-offs in our design.

Fig. 6a presents the PF obtained using the NSGA-III algorithm after 30,000 evaluations, matching the evaluation count of TurBO. It illustrates the variety of possible designs within the multi-objective space, each corresponding to different system performances. Examples are presented in Fig. 6b-c, corresponding to two designs favoring IR and blue transmission respectively. Fig. 6d, present a solution achieving a good balance between the four objectives. This approach significantly facilitates the exploration of the search space, enhancing the capabilities for comprehensive inverse design of optical metasurfaces.

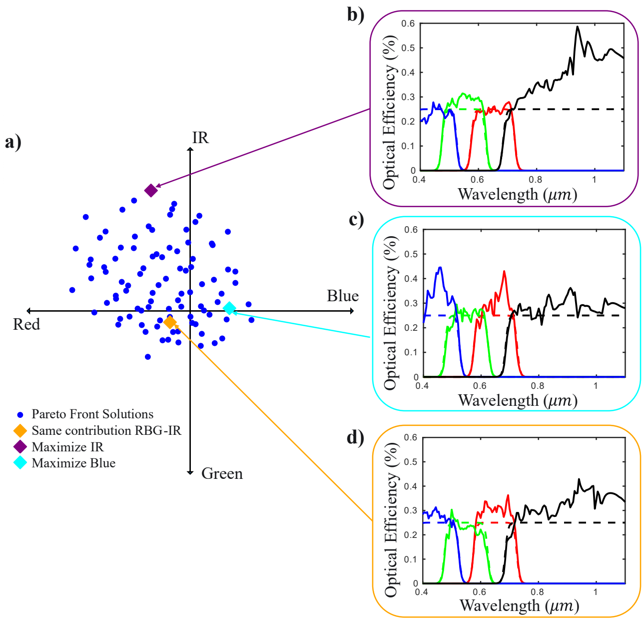

Fig. 6. Multi-objective Pareto front obtained by NSGA-III algorithm a) Star coordinate plots exhibiting the Pareto front of all the possible feasible designs leadings to different performances regarding each wavelength. b-d) Optical efficiencies of selected Pareto-optimal designs emphasizing: b) maximum IR transmission, c) maximum blue transmission, d) balanced trade-off across all four objectives.

IV. Conclusion & Discussions

In this study, we demonstrate the relevance of multi-objective global optimization for the design of RGB-IR metasurface-based CMOS sensor design. While gradient-based single-objective optimization is effective for exploiting specific regions of the design space in very few iterations, it is less suited for problems involving multiple competing objectives. Then, combining multi-objective and global optimization enables both precise design and a complete exploration of trade-offs between RGB-IR channels, which is crucial for achieving a comprehensive inverse design methodology.

The presented results focus on a controlled scenario, assuming a collimated angle of incidence and a single wavelength per color region, which streamlines the optimization process. Future work will extend this approach to more complex conditions, such as angular dispersion, off-axis illumination, and the impact of spectral crosstalk between channels. Nevertheless, the multi-objective approach proves robust, efficiently navigating the high-dimensional parameter space and offers a more comprehensive and adaptable framework compared to single-objective methods.

Acknowledgements

This project has received funding from the CHIP JU program under grant agreement No 101139941. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The authors gratefully acknowledge the contributions of the LIST team at CEA-Leti for their contribution.

References

[1] So et al., “Revisiting the Design Strategies for Metasurfaces: Fundamental Physics, Optimization, and Beyond”, *Adv. Mater., 2023*

[2] Hu et al., “A Review on Metasurface: From Principle to Smart Metadevices”, *2021*

[3] Choi et al., "Optical design of dispersive metasurface nano-prism structure for high sensitivity CMOS image sensor", *IEDM, 2023*

[4] Jamin-Mornet et al., “A RGBZ 1.4μm Image Sensor for Color and Near-Infrared Indirect Time-Of-Flight Depth Sensing on Monolithic Silicon”, *IISW, 2025*

[5] Elsawy et al., “Numerical Optimization Methods for Metasurfaces”, *Laser & Photonics Reviews 2020*

[6] Jeong et al., “A Tutorial on Inverse Design Methods for Metasurfaces”, *Curr. Opt. Photon. 2024*

[7] Zhao et al., “Perfect RGB-IR Color Routers for Sub-Wavelength Size CMOS Image Sensor Pixels”, *Adv. Photonics Res. 2021*

[8] Zhong et al., “High-Efficiency Integrated Color Routers by Simple Identical Nanostructures for Visible and Near-Infrared Wavelengths”, *Photonics, 2023*

[9] Hong et al., "A metasurface color router facilitating RGB-NIR sensing for an image sensor application", *Nanophotonics, 2024*

[10] Peng et al.,” “Inverse design of a light nanorouter for a spatially multiplexed optical filter,” *Opt. Lett., 2023*

[11] Hsu et al., "Simplest but Efficient Design of a Color Router Optimized by Genetic Algorithms", *ACS Photonics*, *2025*

[12] Schubert & Hammond, “Fourier modal method for inverse design of metasurface-enhanced micro-LEDs”, *Optic Express*, *2023*

[13] Santoni et al., “Comparison of High-Dimensional Bayesian Optimization Algorithms on BBOB”, *Association for Computing Machinery, 2024*

[14] Eriksson et al., “Scalable global optimization via local Bayesian optimization”, *Advances in Neural Information Processing Systems, 2019*

[15] Blank and Deb, “pymoo: Multi-Objective Optimization in Python”, *IEEE access, 2020*

[16] 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*