Challenging the Limits of SRG Waveguides: A Human-AI Collaborative Design Concept

Sébastien de Cunsel, Kazue Shimizu**

***Sony Semiconductor Solutions, Kanagawa, Japan**

Abstract

Surface-relief grating (SRG) waveguides are widely used for Augmented Reality (AR) head mounted displays mainly due to their ideal form-factor and transparency. As nanophotonic fabrication methods are progressing, the design-space of periodic nanostructures is constantly expanding. Sony Semiconductor Solutions (SSS) is pushing the boundaries of SRG waveguide design using its unique AI-based simulation software solution combined with its hyper-parametrization concept.

Author Keywords

Waveguide; surface-relief gratings; SRG; head mounted display; Augmented Reality; AR; Mixed Reality; MR; design method; design tools; optimization; segmentation; complexity; Artificial Intelligence; AI.

1. Introduction

Surface-relief grating (SRG) waveguides adoption as a display solution for Augmented Reality (AR) devices is constantly growing. This technology offers very interesting characteristics particularly appreciated for AR-enabled smart-glasses. SRG waveguides can be as thin as regular eyeglasses and can offer very good transparency, allowing its user to clearly appreciate the surroundings while obtaining augmented information. Nevertheless, their efficiency tends to be lower than other wearable display technologies. Also, the use of Diffractive Optical Elements (DOE) combined with the necessity to preserve the projected image uniformity represents one of many challenges SRG waveguide must face.

Because of these difficult challenges, many solutions have been developed, all with their unique set of pros and cons. A variety of SRG waveguide layout exists, from the number of gratings area required to efficiently display the projected image, to the positioning and shapes of those same area. Some solutions involve the use of gratings on a single surface, others on both surfaces. The DOE used within those grating area are also very different from one concept to another. Those DOE can sometimes be using a rectangular, slanted, or blazed profile in two dimensions, sometimes, free-from DOE in three dimensions are used. Mixed solutions can also be found.

One specific solution seems to make consensus to improve the efficiency of SRG waveguides. To face the natural power decay resulted from the pupil replications within the waveguide, many designs show segmented grating area, in particular turn gratings and output couplers (Figure 1). Height modulation is a widely adopted answer to the power decay problem, compensating the reduced waveguide internal reflection by a gradually increasing diffraction efficiency. However, the use of complex DOE shapes and segmentation represents another challenge, especially from a simulation software perspective. To optimize their waveguide, engineers must balance the complexity involved in their design to achieve reasonable computational time.

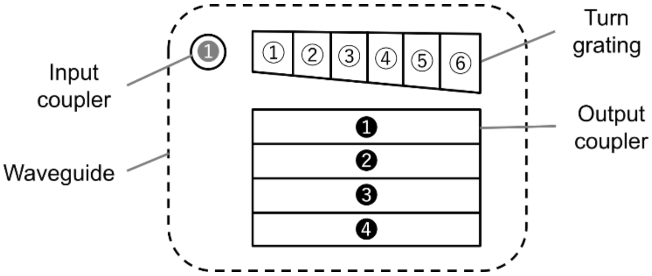

Figure 1. Example of a conventional waveguide using a total of 11 segments (all grating area considered).

2. A Concept Based on Hyper-parametrization

Designing SRG waveguide is challenging for multiple reasons. To become part of a wearable product, the waveguide must comply with various specifications, from maximizing optical performance to achieving an appealing form-factor and avoiding any potential cosmetic issues (e.g., stray-light artifacts).

To balance these specifications, we opted for an increase of the DOE design-space. New parameters are considered to define the potential grating candidates, expanding the number of dimensions. We often exceed 3 parameters, with extensive ranges, to parametrize the DOE used in our designs, resulting in numerous grating candidates (Figure 2).

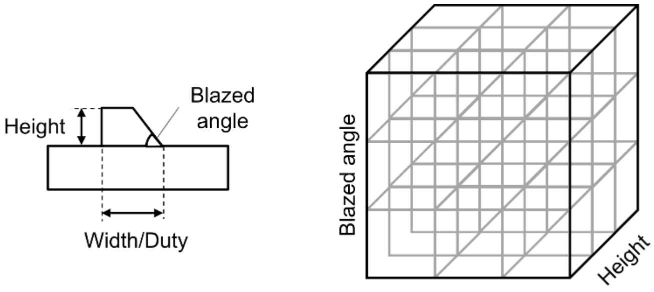

Figure 2. Left: Example of DOE with its parameters; right: Hypercube representing the DOE parameter space.

ISSN 0097-996X/25/5603-1579-$1.00 © 2025

In addition to the DOE definition, we decided to increase the resolution of the grating area segmentation and always use at least 20 uniquely defined segments. A unique segment will be associated with a DOE design that is not reused within another segment. Combined with the free geometrical definition (shape and position) of the segments themselves, the potential designs become vast and allows for more flexibility in how to adjust the waveguide specifications. The newly unlocked design-space also generally provides room for better optical performance optimization.

However, this concept has a major trade-off coming from its core idea. The design complexity grows exponentially as the number of DOE and segments rises (Figure 3). Third-party software and conventional methods very easily reach their computational limits when the number of design candidates increases greatly. Expanding hardware capabilities may be another solution but will only partially absorb the growing complexity due to its linear improvements (number of PC, faster CPU, use of GPU…), and will also greatly rise the computational cost, and consequently, the waveguide design cost.

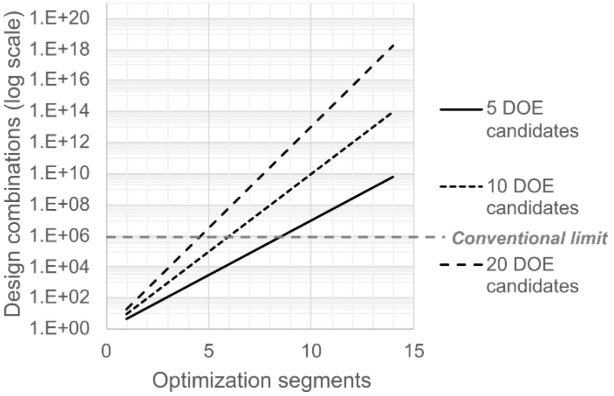

Figure 3. Evolution of the design complexity as a function of number of segments and DOE candidates.

3. Our Design Software Capabilities

As existing software solutions suffer from early computational limitation, we developed our own in-house simulation software suite, particularly tailored for the design of SRG waveguides.

The core of our design software suite is not unique and uses conventional processes to simulate the optical performances of a given waveguide definition. A microscopic solver is used to characterize the DOE optical characteristics such as their diffractive efficiencies. A macroscopic solver is also used to couple in and propagate the light, from its source, within the waveguide, to then reach the area or property being studied (e.g., the users' eye-box, stray-light analysis, etc.). To parametrize and link both processes, an optimizer is integrated allowing the exploration of the selected design-space.

Due to the design complexity of our hyper-parametrization concept, we additionally developed a process to convert the design data (obtained after optimization) into a fabrication-compatible dataset. This automated operation can handle a significant number of segments and associate DOE definition. To avoid any human intervention-based error, the translation is directly made from the design optimization dataset combined with fabrication bias information. This technique also reduces considerably the required time to create the fabrication dataset compared to conventional methods.

To support innovative ideas and novel methods, the software suite was architected with a particular attention to its modularity. The processes within the optimization loop are disconnected from the software they use. This is achieved by using multiple API "layers", avoiding any dependency with the programs used to operate the corresponding process (Figure 4). An example of the benefits from this architecture could be the integration of a new solving software, due to increase input capabilities or a faster calculation method. This integration can then be made faster, with limited effort and without disrupting the overall process flow.

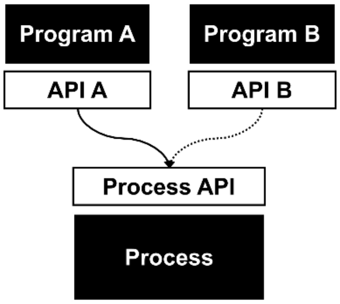

Figure 4. API-based software architecture allowing to disconnect the process and the programs.

With these software capabilities, we were able to confidently work on innovative solutions to face the complexity generated by our hyper-parametrized SRG waveguide concept.

4. A Unique Human-AI Design Collaboration

Before starting any design or optimization process, it is essential to understand the design-space one is about to explore. Minimizing the parameter-space based on design and fabrication constraints is a good start. Over-optimizations can also be avoided if the sensitivity of each input parameter is known in advance, and if tolerances are considered from the very beginning. Local minima and maxima tend to slow down the optimization convergency (at best) and their frequency must be reduced for good performance. This preliminary work must be done carefully, based on human knowledge and experience, and sets up the collaboration between the engineers and the software.

The optimization setup phase is a common form of human-software interaction but allowing such communications during the optimization process is not common nor trivial: most conventional methods and third-party software would not permit it. To tackle this issue, we collaborate with Sony Research Inc. and introduce in-house AI-based algorithms, achieving quicker iterations and smarter decision making. Sony Research Inc's expertise and experience in AI-based R&D, for different applications and industries, enables us to explore beyond conventional methods. Our joint development is tailored to the SRG waveguide design and is seamlessly integrated within our design process.

With an optimization method able to react to engineers' requests, we break the global optimization problem into smaller problems, called "strategies". Those strategies are solved (partially or fully) sequentially and require specific objective definition and merit functions defined by our engineering team. Combined with our improved optimization algorithms, we can dramatically reduce the need for many iterations and for long simulation time. This human-AI collaboration guarantees us to reach the optimal SRG waveguide performance, in a limited amount of time compared to conventional optimization methods (Figure 5).

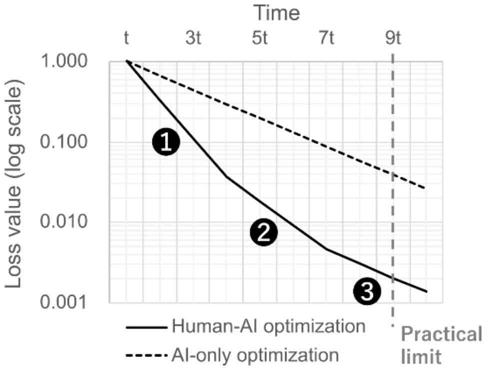

Figure 5. Example of a speed-up improvement due to the use of 3 different strategies within the optimization process. The performance increase depends on the strategies used.

5. Concept Validation and Achievements

To validate our hyper-parametrization concept and to test its flexibility, we successfully designed a set of prototypes, ranging from small to large field-of-view capable waveguides, including the use of one or multiple stacked waveguides.

We are also running several benchmarks to evaluate the benefits of both the design software capabilities and the strategy-based collaboration between engineers and the AI-based optimization. The results will be shared in a second version of this document.

As a conclusion, we would like to point out that segmentation is often responsible for the projected image quality degradation and reduced contrast, which we evaluate by analyzing the modulation transfer function (MTF). Fabricating segments with different DOE designs tend to increase dramatically the probability of defects, particularly at their border connections. However, we demonstrated that we could control the MTF degradation due to this phenomenon, and our latest prototypes show comparable MTF performances to non-segmented designs.

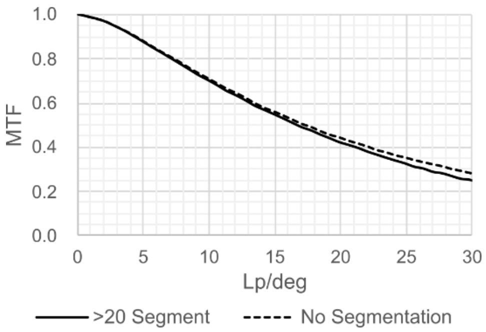

Figure 6. MTF, as a function of line pairs per degree, for both segmented and non-segmented waveguides.

6. Impact and Next Steps

We proved that the hyper-parametrization offers both improved optical performances and design flexibility compared to more conventional designs. We did not face any sign showing that specification limits have been reached (especially the achieved efficiency), so we plan to continuously enlarge our design-space and increase the number of segments used in our waveguides. Now that the SRG waveguide design boundaries have been pushed further away, we aim to achieve designs that will significantly reduce the need for difficult specification trade-offs.

7. Acknowledgements

We, authors, would like to thank all the Sony Semiconductor Solutions collaborators and researchers, both from Japan and Europe, who are taking part of this project and paper. Also, we would like to express our gratitude to Sony Research Inc. for actively participating in the development of our AI-based algorithms and tools, and for their valuable expertise within the field of AI., in particular Zhao When Chow and Lorenzo Servadei from Sony AI.