

# End-to-End Hybrid Refractive-Diffractive Lens Design with Differentiable Ray-Wave Model: Supplementary Material

XINGE YANG and MATHEUS SOUZA, King Abdullah University of Science and Technology, Saudi Arabia

KUNYI WANG, University of British Columbia, Canada

PRANEETH CHAKRAVARTHULA, University of North Carolina at Chapel Hill (UNC), USA

QIANG FU and WOLFGANG HEIDRICH, King Abdullah University of Science and Technology, Saudi Arabia

## ACM Reference Format:

Xinge Yang, Matheus Souza, Kunyi Wang, Praneeth Chakravarthula, Qiang Fu, and Wolfgang Heidrich. 2024. End-to-End Hybrid Refractive-Diffractive Lens Design with Differentiable Ray-Wave Model: Supplementary Material. In *Proceedings of SIGGRAPH ASIA (SIGA '24)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 DIFFERENTIABLE RAY-WAVE MODEL

### 1.1 Implementation Details

We build our differentiable ray-wave model on an open-sourced differentiable ray tracer DeepLens [Wang et al. 2022; Yang et al. 2023]. Firstly we do coherent ray tracing to the DOE surface, then convert the rays into the wave field, as illustrated in Fig. S1. After that, we apply DOE phase modulation and do free-space propagation to the sensor plane. The free-space propagation is implemented with the angular spectrum method, with safe padding to include off-axis components. On the sensor plane, we calculate the amplitude square, crop the valid point spread function (PSF) region, and normalize it to get the final PSF. Specifically, the PSF center is determined by tracing the chief ray through the refractive lens, since we do not want the DOE to change the focus of the lens system. In our end-to-end lens design, the DOE is learned to correct the optical aberrations or provide information encoding capability for the hybrid lens system. This setting brings another benefit in that the DOE can be designed to have relatively low frequency since it does not need to provide focusing power, therefore reducing the fabrication cost and difficulty.

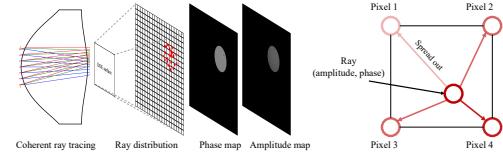


Fig. S1. Coherent ray tracing for accurate simulation of the aberrated wave-front before the DOE surface. During the coherent ray tracing process, the position, direction, and phase information of optical rays are recorded. At the DOE surface, the optical rays are converted into a phase map and an amplitude map by spreading a ray into its neighboring four pixels. The complex wave field is then used for DOE phase modulation and subsequent free-space propagation.

### 1.2 Comparison Between Single Precision and Double Precision in Coherent Ray Tracing

In our experiments, we employ double-precision arithmetic in the hybrid ray-wave model to ensure accurate simulations. Precision considerations are crucial, yet they are often overlooked, particularly in modern deep optics models built upon auto-differentiable frameworks, such as PyTorch and TensorFlow, which default to single-precision arithmetic. Our experimental results reveal that single-precision arithmetic is insufficient for accurately simulating the hybrid ray-wave model, as illustrated in Fig. S2. The simulated phase map obtained using single-precision arithmetic exhibits discontinuities in both amplitude and phase fidelity, whereas the double-precision simulation can successfully overcome such issues.

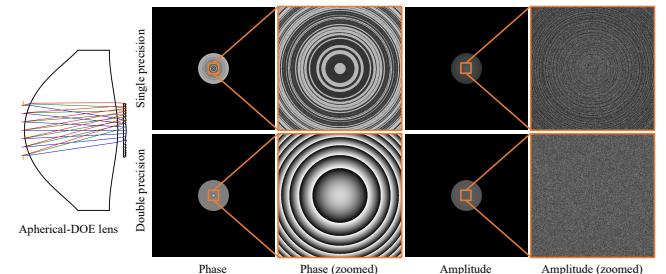


Fig. S2. Comparison of complex wave field calculation results between single and double precision in coherent ray tracing. Single precision introduces significant errors in the wave field calculation, leading to inaccurate PSF simulation results. The precision problem is particularly severe because the wavelength of the light is in the order of micrometers, while the physical size of the wave field is much larger than the wavelength.

---

Authors' Contact Information: Xinge Yang; Matheus Souza, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; Kunyi Wang, University of British Columbia, Vancouver, Canada; Praneeth Chakravarthula, University of North Carolina at Chapel Hill (UNC), Chapel Hill, USA; Qiang Fu; Wolfgang Heidrich, [wolfgang.heidrich@kaust.edu.sa](mailto:wolfgang.heidrich@kaust.edu.sa), King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGA '24, Dec, 2024, Tokyo, Japan

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM  
<https://doi.org/XXXXXXX.XXXXXXX>

### 1.3 Comprehensive Explanation for Table.1

Table S.I. Comparison of different hybrid refractive-diffractive lens simulation models.

	Paraxial optics	Zemax [Zemax LLC 2023]	Chen et al. [Chen et al. 2021]	Zhu et al. [Zhu et al. 2023]	Ours
Optical model	Wave	Ray	Ray	Wave-ray	Ray-wave
Accuracy	✗	✗	✓	✗	✓
Optical aberration	✗	✓	✓	✓	✓
Edge diffraction	✓	✗	✓	✓	✓
Phase modulation	✓	✓	✗	✓	✓
Discontinuous phase	✓	✗	✗	✗	✓
Differentiable	✓	✓	✓	✓	✓
End-to-end design	✓	✗	✗	✓	✓

A comprehensive comparison between different optical simulation models is presented in Table S.I. Specifically, we primarily consider optical simulation models that are or can be designed to be differentiable with reasonable effort. Therefore, although there are several other simulation methods, such as ray-tracing based methods [Gross 2020; Mout et al. 2016] for cascaded diffraction simulation, and wave-optics-based methods liu2024wave decomposing a lens surface into multiple slices, we exclude these methods in our comparison.

As discussed in the main paper, the paraxial wave optics model idealizes the refractive lens as a thin phase plate and fails to simulate optical aberrations. Despite its widespread use in almost all existing end-to-end lens design works [Shi et al. 2022; Sun et al. 2020; Wu et al. 2019] due to its simplicity, this paraxial optical model lacks accuracy and can not provide optimization capability for the refractive lens. Although pupil aberration theory has been proposed to model the optical aberration of the refractive lens [Wyant and Creath 1992], it is challenging to incorporate off-axis aberrations for different wavelengths, since the Zernike coefficients must be calculated again and again for such settings. What is more, it is not possible to convert the optimized pupil function back to lens geometry, which precludes its usage in end-to-end design tasks. A more detailed comparison between the paraxial wave optics model and our ray-wave model is presented in Supplementary Sec. 1.4.

The ray tracing model employed in Zemax can approximate the light bending direction after diffraction, but it fails to model the real diffractive phenomena and can not function for binary and discontinuous phase maps. Additionally, it can not model edge diffraction, as optical rays either pass through or are blocked by the aperture. While this ray tracing model can be designed to support end-to-end optical design, it is unsupported in Zemax. A more detailed comparison between the ray tracing model and our ray-wave model is presented in Sec 1.5.

Chen et al. [2021] proposed a ray tracing model that performs coherent ray tracing to the exit pupil and applies the Rayleigh-Sommerfeld diffraction integral [Born and Wolf 2013; Goodman 2005] to calculate the PSF on the sensor plane. This exit-pupil diffraction approach does not involve wave field conversion and can only function for aperture diffraction. It is unable to model diffractive phase modulation, as the DOE alters the light path after its plane, rendering ray tracing to the exit pupil plane infeasible. While the original literature did not present a differentiable implementation, this exit-pupil diffraction model can be designed to be differentiable; however, this would result in high memory consumption. A more

detailed comparison between Chen et al.’s model and our ray-wave model is presented in Sec 1.6.

In contrast, our ray-wave model can accurately simulate both optical aberrations and diffractive phase modulation, and it is fully differentiable for end-to-end hybrid lens and network design. Our ray-wave model does not require any assumptions, such as the paraxial approximation or local grating approximation, which can reduce simulation accuracy and limit real-world practicality. It can simulate both edge diffraction and discontinuous phase maps, making it a suitable candidate for end-to-end hybrid lens design. Another technical improvement is that both ray tracing sampling ( $10^6$ ) and wave field sampling ( $6,000 \times 6,000$ ) in our experiments exceed the values used in existing works, enhancing the simulation accuracy for hybrid refractive-diffractive lenses.

### 1.4 Comparison Between Pupil Aberration Model and Ray-Wave Model

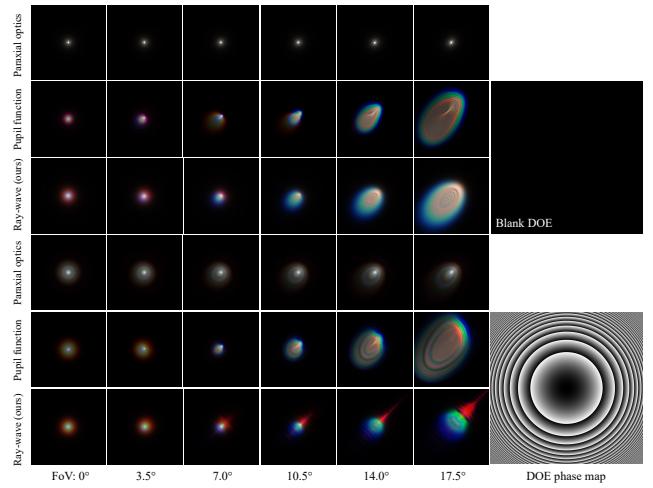


Fig. S3. Comparison of PSF simulations between the paraxial wave optics model (top), pupil aberration model (middle), and our ray-wave model (bottom). The paraxial wave optics model idealizes the refractive lens as a thin phase plate and fails to simulate optical aberrations. The pupil aberration model can model the optical aberrations, but also not accurate enough for off-axis FoVs. In contrast, our ray-wave model can accurately simulate both optical aberrations and diffractive phase modulation, providing a more accurate PSF simulation.

Pupil aberration theory [Wyant and Creath 1992] models the optical aberrations of a refractive lens system by representing them as an aberrated wavefront on the exit pupil plane, i.e., generalized pupil function [Goodman 2005]. Traditionally, the real wavefront error, when compared to an ideal spherical wavefront, is modeled using Zernike polynomials. This approach involves modeling the refractive lenses as an aberrated wavefront function, which affects the corresponding PSF by introducing optical aberrations, making it differ from the ideal Airy pattern. However, one of the challenges with this method is to adequately incorporate off-axis aberrations and it is difficult to convert the optimized pupil function back into an optimized lens geometry, limiting its utility in design tasks.

In this section, we compare the paraxial wave optics model, the pupil aberration model, and our ray-wave model. Using the Wavefront analysis tool in Zemax (OpticStudio 14.2), we obtain the Zernike Standard coefficients (up to the first 37 terms) to compute the aberrated wavefront on the plane where the DOE resides. We repeat the simulation for various field of views ( $0^\circ$ ,  $3.5^\circ$ ,  $7.0^\circ$ ,  $10.5^\circ$ ,  $14.00^\circ$ , and  $17.50^\circ$ ) and various wavelengths. Finally, we compare the PSFs among the three models. The paraxial wave optics model, which simplifies the refractive lens to a thin phase plate, fails to capture optical aberrations effectively. The pupil aberration model can model the optical aberrations, but it is not accurate enough, particularly for off-axis regions. In contrast, our ray-wave model successfully simulates both optical aberrations and diffractive phase modulation, offering a more precise PSF simulation.

### 1.5 Comparison Between Ray-Wave Model and Ray Tracing Model

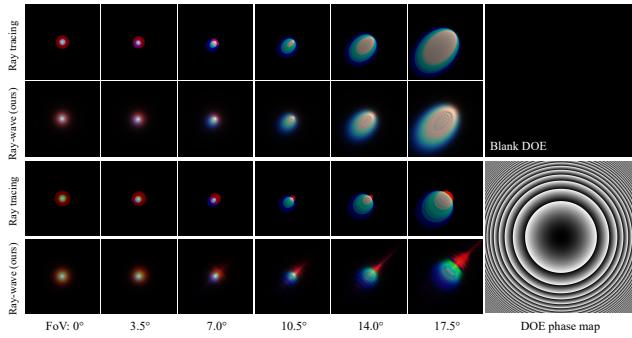


Fig. S4. Comparison of PSF simulations between the ray tracing model and the proposed ray-wave model. For DOE with low-frequency phase patterns, for example, the blank DOE and the center region of the bottom DOE, the ray tracing model and the ray-wave model yield similar results. However, as the spatial frequency of the DOE phase pattern increases, for example, the edge regions of the bottom DOE, significant differences emerge between the PSF simulations of the ray tracing model and the ray-wave model.

Ray tracing is a widely employed method for optical simulations, particularly in the design of refractive lenses. Numerous efforts have been made to simulate diffractive surfaces using ray tracing, with a commonly adopted approach being the introduction of a ray bending term based on the local grating approximation [Fischer et al. 2000; Yu et al. 2011]. This method is utilized in commercial optical design software, such as Zemax [Zemax LLC 2023], as well as in some recent research works [Shih and Renshaw 2024; Zhang et al. 2024; Zhu et al. 2023]. However, this approximation and simulation approach have been proven inaccurate, as they fail to capture the real process of light transport and diffraction effects accurately. Furthermore, the ray tracing model requires calculating the local gradient of the diffractive phase map and cannot handle discontinuous phase maps, as illustrated in the main paper. While the main paper employs an ideal paraxial hybrid lens due to its known ground truth PSF, we present a comparison between our proposed ray-wave model and the ray tracing model using an aspherical lens and different diffractive

optical element (DOE) phase maps, as illustrated in Fig. S4. For DOEs with low-frequency phase patterns, the ray tracing model and the ray-wave model yield similar PSF simulation results. However, as the frequency of the phase pattern increases, the discrepancy between the PSF simulations of the ray tracing model and the ray-wave model becomes more pronounced.

### 1.6 Comparison Between Ray-Wave Model and Chen *et al.*'s Model

Chen *et al.* 2021 introduced a ray tracing model that performs coherent ray tracing up to the exit pupil and employs the Rayleigh-Sommerfeld diffraction integral [Born and Wolf 2013; Goodman 2005] to calculate the PSF on the sensor plane. This method of exit-pupil diffraction bypasses the need for wave field conversion from ray spots, but it is limited to scenarios that only involve clear apertures, making it unsuitable for hybrid refractive diffractive optical systems. Despite this limitation, it is considered a significant contribution to ray tracing-based diffraction simulation models due to its status as one of the latest works in this area and its demonstrated accuracy, validated by real camera captures. In our study, we have compared our ray-wave model to Chen *et al.*'s model, using a clear aperture at the DOE plane, as illustrated in Fig. S5. The results show that both models produce similar PSF simulations in cases with clear apertures, supporting the accuracy of our proposed ray-wave model.

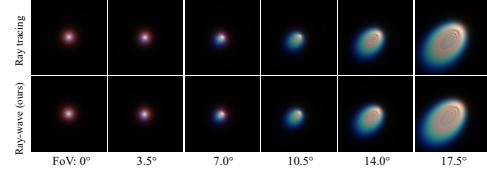


Fig. S5. Comparison of PSF simulations between our proposed ray-wave model and Chen *et al.*'s model [Chen et al. 2021]. The ray-wave model can accurately simulate both optical aberrations and diffractive phase modulation, providing a more accurate PSF simulation. In contrast, Chen *et al.*'s model can only simulate aperture diffraction and is unable to model diffractive phase modulation.

## 2 EXPERIMENTAL SETTINGS AND EXTRA RESULTS

### 2.1 DOE Fabrication

The DOE substrate is a 4-inch fused silica wafer with a thickness of 0.5 mm. We first employ a laser direct writer (Heidelberg DWL66+) to fabricate the feature patterns on 5-inch Soda Lime masks. In the lithography step, the wafer is wet-cleaned by Piranha solution, and coated with a layer of 150 nm Chromium (Cr) by sputtering. We prepare the photoresist (AZ1505) by spin-coating a uniform layer of 0.6  $\mu\text{m}$  before soft-baking for 1 min. The alignment between the mask and the wafer is performed on a contact aligner (EVG6200  $\infty$ ) for a UV exposure dose of 0  $\text{mJ}/\text{cm}^2$ . The photoresist is then developed in AZ726MF for 15 sec. Before reactive ion etching (RIE), the Cr layer is wet-etched with TechniEtch Cr01. The target depth in the fused silica is obtained by plasma etching with 15 sccm of CHF<sub>3</sub> and 5 sccm of O<sub>2</sub> at 10°C. The etching time depends on the target depths, which are 75 nm, 150 nm, 300 nm, and 600 nm, respectively for each fabrication iteration. By 4 iterations of lithography and etching, the 16-level DOE can be fabricated.

### 2.2 Wavelength Selection

In our experiments, we choose three representative wavelengths for each color channel to account for the broadband nature of the spectrum. Specifically, we select 0.62  $\mu\text{m}$ , 0.66  $\mu\text{m}$ , and 0.70  $\mu\text{m}$  for the red channel, 0.50  $\mu\text{m}$ , 0.53  $\mu\text{m}$ , and 0.56  $\mu\text{m}$  for the green channel, and 0.45  $\mu\text{m}$ , 0.47  $\mu\text{m}$ , and 0.49  $\mu\text{m}$  for the blue channel. During training, in each iteration, we randomly select one wavelength from the three choices for each color channel and calculate the corresponding RGB PSF. This wavelength sampling approach is consistent across different fields of view (FoVs). By doing so, we ensure that the entire training process covers the entire broadband spectrum, making both the hybrid lens and the network robust to different wavelengths. For the evaluation of the designed lenses, we calculate the PSF maps by averaging the three wavelengths, weighted by the same sensor response. We did not use the real broadband response because it is unknown for the CMOS sensor (OmniVision OV2710) used in our experiments.

### 2.3 End-to-end training

We employ the simulated PSF map for image simulation and end-to-end training. We find the single A100 GPU can only afford end-to-end optical design for a single RGB PSF, however, in our experiments, we need to consider specially-varying off-axis PSFs. To address the memory issue faced during end-to-end training, we adopt gradient checkpointing and gradient accumulation strategies to reduce memory consumption to an affordable level. During training, we uniformly sample 10×10 PSFs to represent the entire FoV. Specifically, we first calculate the PSF map in a non-differentiable manner. Then, we activate the gradient calculation of the PSF map and perform patch convolution for image simulation. The simulated image is then fed into the network for reconstruction, and the loss function on the final output is calculated. By backpropagating the loss function through the reconstruction network, we can obtain the gradient for each PSF. Then we re-do the forward optical simulation and iteratively backpropagate the gradients for each distinct PSF. When re-calculating the PSF map, we reset the pseudo-random

seed to ensure an identical forward calculation can be replayed, as discussed in existing research [Vicini et al. 2021; Yang et al. 2023].

We employ the DIV2K dataset [Agustsson and Timofte 2017] for training and validation purposes. Initially, we conduct end-to-end training for 50 epochs. During this end-to-end training stage, we do not consider sensor noise as it cannot provide effective gradients for optical design. Subsequently, we fix the optics and fine-tune the network for an additional 50 epochs. During network fine-tuning, we calculate 40×40 PSF map, introduce sensor noise to the simulated images, and average three wavelengths for each color channel to simulate more accurate PSF maps. The learning rates are set to 10<sup>-4</sup> for the refractive lens, 10<sup>-1</sup> for the DOE, and 10<sup>-4</sup> for the network. We utilize the Adam optimizer [Kingma and Ba 2014], along with warm-up and cosine annealing learning rate schedulers, during both the end-to-end design and network fine-tuning stages.

For the aspherical-DOE lens prototype, where the refractive lens is fixed, we pre-calculate the aberrated wave field at the DOE surface to conserve memory and accelerate the process. For each epoch, we sample a group of wavelengths for aberrated wave field calculation. At each iteration, we only need to perform the DOE phase modulation and free space propagation for PSF calculation. For the aberration correction application, we first optimize the size of the PSFs and then perform end-to-end optimization. For the extended depth-of-field (EDoF) application, we first optimize the size and similarity of PSFs at different depths, then perform end-to-end optimization. To further reduce memory consumption and expedite the training for EDoF imaging, we utilize 5×5 PSF maps for training, while the fine-tuning remains unchanged. The loss function for EDoF imaging is defined on both the quality and the similarity of reconstructed images, expressed as

$$\mathcal{L} = \sum_{i \neq j} \mathcal{L} \left( \mathcal{N} (\text{PSF}_{d_i} * \mathbf{I}), \mathcal{N} (\text{PSF}_{d_j} * \mathbf{I}) \right) + \alpha \sum_i \mathcal{L} \left( \mathcal{N} (\text{PSF}_{d_i} * \mathbf{I}), \mathbf{I} \right). \quad (1)$$

In the loss function, the first term minimizes the difference between reconstructed images at different depths, while the second term enhances the quality of the reconstructed images. For both loss terms, we employ the mean-square-root errors. The weight term  $\alpha$  is set to 0.3 in our experiments. Prior to training, we first focus the refractive lens at 30 cm and optimize for three depths (20 cm, 30 cm, and 10 m). After the end-to-end design, we calculate the PSF maps and simulate images at different depths for network fine-tuning.

## 2.4 Experimental Setup

Figure S6 presents the experimental setup of our hybrid aspherical-DOE lens prototype. For the aberration correction application, the designed DOE is rotationally symmetric; therefore, we 3D printed threads on the lens holder and rotated the hybrid lens to focus at different depths. The refractive lens has a “blank” DOE as a square aperture stop for fair comparison. Two lenses are installed on a stage for image capture. Although the captured images are not well aligned, we believe it is sufficient to discern the difference between the images captured by the two cameras. For EDoF imaging, since the DOE is not rotationally symmetric, directly installing the hybrid lens into the camera and rotating it is not a reasonable setting. Additionally, as the EDoF lens is designed for a short distance (minimum imaging distance is 20 cm), the image misalignment caused by camera misalignment can not be neglected. To address this, we use a single CMOS sensor and fix its position in the scene. We install the two hybrid lenses on a translation stage that can move left and right. During the experiments, we first capture the scene with one hybrid lens, then adjust the translation stage to capture the scene with the other hybrid lens. The CMOS sensor is installed on another translation stage that can move forward and backward to adjust the lens focus. We make our best effort to align the two captures.

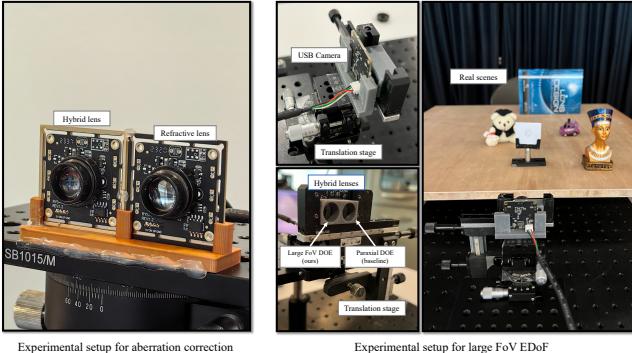


Fig. S6. Experimental setup for the hybrid aspherical-DOE lens prototype. Left: Setup for aberration correction application, where threads are 3D printed on the lens holder to rotate and adjust the focus. Right: Setup for extended depth-of-field imaging, utilizing a single camera sensor and a translation stage to switch between different lenses. The camera sensor can move forward and backward to adjust the focus, and efforts are made to align the two EDoF captures.

## 2.5 Results for Computational Aberration Correction Experiments

Figure S7 presents the simulated PSF maps of the refractive lens and the hybrid lens. The refractive lens incorporates a blank DOE of the same size as the hybrid lens to ensure a fair comparison. The PSF maps are calculated using the proposed ray-wave model. As evident from the figure, the learned DOE successfully corrects optical aberrations, particularly in the off-axis regions where aberrations are less significant. Regarding the DOE phase map, the central region exhibits smoother phase patterns, while the edge regions display higher phase pattern frequencies, thereby providing aberration correction capability.

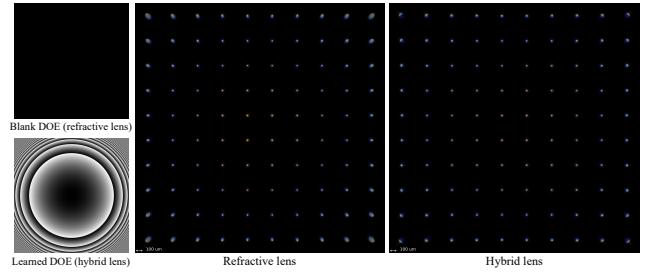


Fig. S7. PSF maps of the refractive lens and the hybrid lens in the aberration correction application. The learned DOE corrects the optical aberrations, especially for the off-axis regions.

Figure S8 presents the simulated raw captures and corresponding reconstruction results of the refractive lens and the hybrid lens. The aberration correction DOE learned by our ray-wave model successfully mitigates the optical aberrations introduced by the imperfections of the original refractive lens, especially after image reconstruction. The results demonstrate the successful adoption of the proposed ray-wave model in hybrid lens design for aberration correction.

Figure S9 presents real-world captures and corresponding reconstruction results of the refractive lens and the hybrid lens. The hybrid lens, incorporating a DOE designed using our proposed ray-wave model, exhibits enhanced image quality, particularly in off-axis regions. Although the hybrid lens introduces some haze to the captured images, the reconstruction network effectively recovers sharp and clear final outputs, preserving object structures and details. In contrast, the refractive lens exhibits significant optical aberrations in off-axis image regions.

In classical lens design, compound refractive elements are stacked to achieve optimal image quality, resulting in a bulky physical form factor. Incorporating a DOE at the back of the refractive lens can significantly improve imaging quality without increasing the physical size of the lens.



Fig. S8. Image quality comparison between the refractive lens and the hybrid lens through simulation. A DOE is optimized to correct the aberrations of the aspherical lens (ours). For comparison, the original aspherical lens with a blank DOE is employed, and an image reconstruction network is optimized for image reconstruction. The DOE optimized by our proposed ray-wave model successfully enhances the image quality, particularly after reconstruction by the network.

## 2.6 Results for Extended Depth-of-Field Imaging

Figure S10 presents simulated PSF maps at different depths for the hybrid lens with the paraxial DOE (baseline) and the DOE optimized using our proposed ray-wave model. The PSF maps are calculated using our ray-wave model, while the DOEs are optimized by different models. The paraxial EDoF DOE is designed using the paraxial optical model, which idealizes the refractive lens as a thin phase

plate, neglecting optical aberrations. Consequently, only the central PSFs at different depths exhibit similarity, while off-axis PSFs exhibit significant optical aberrations. In contrast, our proposed ray-wave model accurately simulates optical aberrations during DOE optimization. Therefore, the DOE optimized by our method demonstrates consistent PSF similarity across the full FoV and mitigates the optical aberrations introduced by the refractive lens.

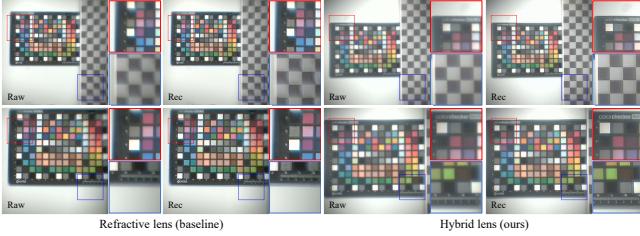


Fig. S9. Real-world image quality comparison between the refractive lens (baseline) and the hybrid lens incorporating a diffractive optical element designed using the proposed ray-wave model. The hybrid lens mitigates optical aberrations caused by imperfections in the refractive lens, enhancing the quality of both captured and reconstructed images.

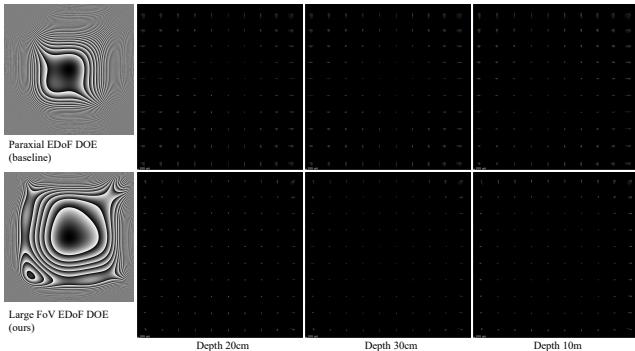


Fig. S10. PSF map comparison of extended depth-of-field imaging between DOEs designed by paraxial wave optics (top) and our proposed ray-wave model (bottom). Our EDoF DOE successfully optimizes the PSFs to be similar across different depths and FoVs, while the paraxial EDoF DOE neglects off-axis regions and introduces more aberrations.

Figure S11 presents a comparison of EDoF imaging quality between DOEs designed using paraxial wave optics and our proposed ray-wave model. For the paraxial EDoF DOE design, both the DOE and the image reconstruction network are optimized using the paraxial optical model. After the end-to-end design process, the network is fine-tuned with images simulated by the paraxial optical model, consistent with existing EDoF works [Pinilla et al. 2022; Sitzmann et al. 2018]. For evaluation, we load the different DOE designs into our ray-wave model and calculate PSF maps using the ray-wave model. For the paraxial DOE, the simulated images (“raw”) at different depths exhibit significant optical aberrations, particularly in off-axis regions. The corresponding image reconstruction network can not successfully recover clear and sharp images due to two reasons: first, the network is trained on degraded images simulated using on-axis PSFs; second, even the on-axis PSFs are not accurately simulated in the paraxial optical model, especially for short imaging distances. Consequently, the network optimized alongside the paraxial optical model can not effectively reconstruct the real simulated images. In contrast, our ray-wave model accurately simulates optical aberrations and DOE phase modulation. During the end-to-end training, the DOE and the network learn to produce clear and sharp

images in the presence of optical aberrations and even correct these aberrations. Additionally, the network is optimized with accurately simulated camera captures, resulting in successful reconstruction of final images.

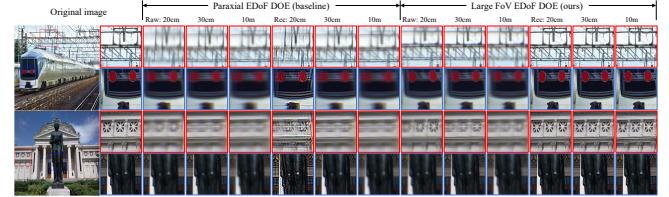


Fig. S11. Image quality comparison of extended depth-of-field imaging between DOEs designed by paraxial wave optics (baseline) and our proposed ray-wave model (ours). Both simulated and reconstructed images using the paraxial optical model show significant optical aberrations compared to our design. Moreover, the reconstruction results of paraxial DOE at 20 cm show strong artifacts because the network is trained with inaccurate PSFs, therefore failing to generalize to realistic PSFs.

Figure S12 presents a comparison of EDoF imaging performance between hybrid lenses incorporating DOEs designed using the paraxial optical model (baseline) and the proposed ray-wave model. The paraxial EDoF DOE demonstrates EDoF capability and clear imaging only for the central region and a narrow FoV. In off-axis regions, both the captured and reconstructed images exhibit significant optical aberrations, resulting in loss of detail. In contrast, the EDoF DOE designed with our ray-wave model enables clear imaging within a large FoV for different depths. The network, optimized with accurately simulated images, can effectively reconstruct the camera captures and produce clear results, preserving object details.

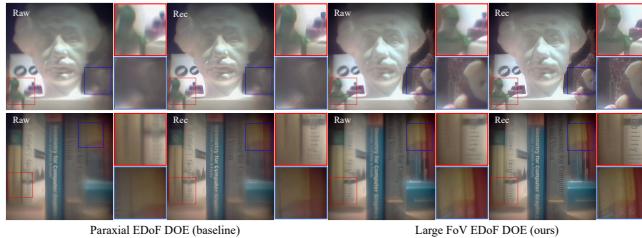


Fig. S12. Comparison of EDoF imaging performance in real-world scenarios between DOEs designed using paraxial wave optics (baseline) and the proposed ray-wave model (ours). For the paraxial DOE, only the central region of the captured images exhibits clarity over the depth of field, while off-axis regions suffer from significant optical aberrations. In contrast, the EDoF DOE designed with our ray-wave model mitigates aberrations across the field of view.

## REFERENCES

- Eirikur Agustsson and Radu Timofte. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 126–135.
- Max Born and Emil Wolf. 2013. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier.
- Shiqi Chen, Huajun Feng, Dexin Pan, Zhihai Xu, Qi Li, and Yueting Chen. 2021. Optical aberrations correction in postprocessing using imaging simulation. *ACM Transactions on Graphics (TOG)* 40, 5 (2021), 1–15.
- Robert Edward Fischer, Biljana Tadic-Galeb, Paul R Yoder, Ranko Galeb, Bernard C Kress, Stephen C McClain, Tom Baur, Richard Plympton, Bob Wiederhold, and Bob Grant Alastair J. 2000. *Optical system design*. Vol. 599. Citeseer.
- Joseph W Goodman. 2005. *Introduction to Fourier optics*. Roberts and Company publishers.
- Herbert Gross. 2020. Cascaded diffraction in optical systems. Part I: simulation model. *JOSA A* 37, 2 (2020), 240–249.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Marco Mout, Michael Wick, Florian Bociort, Jörg Petschulat, and Paul Urbach. 2016. Simulating multiple diffraction in imaging systems using a path integration method. *Applied optics* 55, 14 (2016), 3847–3853.
- Samuel Pinilla, Seyyed Reza Miri Rostami, Igor Shevkunov, Vladimir Katkovnik, and Karen Egiazarian. 2022. Hybrid diffractive optics design via hardware-in-the-loop methodology for achromatic extended-depth-of-field imaging. *Optics Express* 30, 18 (2022), 32633–32649.
- Zheng Shi, Yuval Bahat, Seung-Hwan Baek, Qiang Fu, Hadi Amata, Xiao Li, Praneeth Chakravarthula, Wolfgang Heidrich, and Felix Heide. 2022. Seeing through obstructions with diffractive cloaking. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–15.
- Ko-Han Shih and C Kyle Renshaw. 2024. Hybrid meta/refractive lens design with an inverse design using physical optics. *Applied Optics* 63, 15 (2024), 4032–4043.
- Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.
- Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 2020. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1386–1396.
- Delio Vicini, Sébastien Speirer, and Wenzel Jakob. 2021. Path replay backpropagation: Differentiating light paths using constant memory and linear time. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.
- Congli Wang, Ni Chen, and Wolfgang Heidrich. 2022. dO: A differentiable engine for deep lens design of computational imaging systems. *IEEE Transactions on Computational Imaging* 8 (2022), 905–916.
- Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. 2019. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–12.
- James C Wyant and Katherine Creath. 1992. Basic wavefront aberration theory for optical metrology. *Applied optics and optical engineering* 11, part 2 (1992), 28–39.
- Xinge Yang, Qiang Fu, and Wolfgang Heidrich. 2023. Curriculum learning for ab initio deep learned refractive optics. *arXiv preprint arXiv:2302.01089* (2023).
- Nanfang Yu, Patrice Genevet, Mikhail A Kats, Francesco Aieta, Jean-Philippe Tetienne, Federico Capasso, and Zeno Gaburro. 2011. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *science* 334, 6054 (2011), 333–337.
- Zemax LLC. 2023. *Zemax OpticStudio User Manual*. Zemax LLC, Kirkland, WA, USA. Available from Zemax website: <https://www.zemax.com>.
- Qiango Zhang, Zeqing Yu, Mengguang Wang, Yiyang Liu, Changwei Zhang, Chang Wang, and Zhenrong Zheng. 2024. Centimeter-Scale Achromatic Hybrid Metalens Design: A New Paradigm Based on Differentiable Ray Tracing in the Visible Spectrum. *arXiv preprint arXiv:2404.03173* (2024).
- Ziwei Zhu, Zhaocheng Liu, and Changxi Zheng. 2023. Metalens enhanced ray optics: an end-to-end wave-ray co-optimization framework. *Optics Express* 31, 16 (2023), 26054–26068.