ABSTRACT

Title of Thesis: OPTIMAL POINT-SPREAD-FUNCTION ENGINEERING WITH DYNAMIC OPTICS AND EVENT CAMERAS

Sachin Shah
Masters of Science, 2024

Thesis Directed by: Professor Christopher A. Metzler
Department of Computer Science

Computational imaging systems co-design optics and algorithms to observe phenomena beyond the reach of traditional cameras. Point-spread-function (PSF) engineering is a powerful technique wherein a custom phase mask is integrated into an optical system to encode additional information into captured images. Used in combination with deep learning, such systems now offer state-of-the-art performance at three-dimensional molecule localization, extended depth-of-field imaging, lensless imaging, and other tasks.

Recent hardware breakthroughs are unlocking unprecedented ultrafast capabilities such as micro-electromechanical system based spatial light modulators will allow us to module light at kilohertz rates and neuromorphic event cameras will enable kilohertz lower-power and high-dynamic-range capture. Unfortunately, existing theories and algorithms are unable to fully harness these new capabilities. This work answers a natural question: Can one encode additional information and achieve superior performance by leveraging the ultrafast capabilities of spa-

tial light modulators and event cameras. We first prove that the set of PSFs described by static phase masks is non-convex and that, as a result, time-averaged PSFs generated by dynamic phase masks displayed on a spatial light modulator are *fundamentally more expressive*. We then derive the theoretical limits on three-dimensional tracking with PSF-engineered event cameras. Using these bounds, we design new optimal phase masks and binary amplitude masks. We demonstrate the efficacy of our designs through extensive simulations and validate our method with a simple lab prototype.

OPTIMAL POINT-SPREAD-FUNCTION ENGINEERING WITH DYNAMIC OPTICS AND EVENT CAMERAS

by

Sachin Shah

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Masters of Science
2024

Advisory Committee:
Professor Christopher A. Metzler, Chair/Advisor
Professor Yiannis Aloimonos
Professor Giuliano Scarcelli

Acknowledgments

I would like to thank my advisor, Prof. Christopher A. Metzler, for giving me the opportunity to work on exciting computational imaging problems. My work would not have been possible without his invaluable guidance on navigating prior optics research, working through detailed theoretical analysis, and paper writing. He has always made himself available to discuss both technical minutia and high-level long term goals.

In addition, I would like to thank my co-authors Sakshum Kulshrestha, Matthew Chan, Haoming Cai, Jingxi Chen, Chahat Deep Singh and Yiannis Alomonios for a great collaborative atmosphere for tackling challenging problems. Thanks to Kevin Zhang and Hadi Alzayer for fruitful discussions on state-of-the-art computer vision research. I have had great interactions with the rest of the Intelligent Sensing Lab, Mingyang Xie, Sanjaya Herath, Janith Senanayaka, Roksana Khanom, Isabelle Rathbun, and Matthew Ziemann on a wide range of topics.

I would like to acknowledge financial support from the Joint Directed Energy Transition Office, Office of Naval Research, Dolby, and SAAB.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Computational Imaging	CI
Cramér Rao Bound	CRB
Diffractive Optical Element	DOE
Fisher Information Matrix	FIM
Micro-ElectroMechanical System	MEMS
Multi-Layer Perception	MLP
Peak Signal to Noise Ratio	PSNR
Pixel-wise	PW
Point Light Source	PLS
Point-Spread-Function	PSF
Root Mean Square Error	RMSE
Single Molecule Localization Microscopy	SMLM
Spatial Light Modulator	SLM
Structural Similarity Index	SSIM

Chapter 1: Introduction

Since the introduction of a camera in the 1800s, camera development has largely focused on produced pleasing photos for human enjoyment. With the advent of computer vision, photos are now used a primary signals to perform a wide range of tasks. Unfortunately, many tasks we hope to accomplish are ill-posed, inefficient, or suboptimal with general purpose sensors designed for producing pretty pictures.

Looking towards nature, we can see remarkable development of diverse visual processing systems. Complex eye structures in species like frogs [1], cuttlefish [2], and honey bees [3] demonstrate many possible sensory behaviors. Evolutionary pressures have resulted in these organisms *specializing* their senses to be hyper-effective in their environment. Inspired by these long-term natural trends, computational imaging (CI) systems integrate and co-design optics and algorithms to extract information and observe phenomena beyond the reach of traditional cameras and optics.

Researchers have realized cleverly designed optics can "encode" additional information into these images to have better task performance, such as depth estimation [4]. These systems trade off visually appealing photographs with ones with defocus blur and chromatic aberrations in hopes of uncovering non-visual information.

CI has already had a profound impact: In biomedicine, CI based single-particle cryo-

electron microscopes formed high resolution images of the SAR-CoV-2 protein, helping us to understand and neutralize COVID-19. In astronomy, CI based radar telescopes captured never-before-seen images of black holes, helping us understand our place in the universe.

Recent breakthroughs in material science, machine learning, and nanofabrication are unlocking unprecedented capabilities. For example, micro-electromechanical system (MEMS) based spatial light modulators will allow us to modulate and manipulate light at KHz rates and neuromorphic event cameras will enable low-power, fast, high dynamic range data capture. Unfortunately, existing theories and algorithms cannot effectively harness these capabilities. This dissertation focuses developing and applying new theory to fully harness the capabilities of these next-generation hardware devices. First, we show dynamic phase masks through SLMs unlock a fundamentally new point-spread-function design space. We demonstrate these new dynamic designs can outperform static ones at depth estimation and all-in-focus imaging tasks. Second, we show event cameras can enable ultrafast 3D single-molecule localization microscopy. We derive new fundamental limits for 3D tracking with event cameras to design optimal phase masks specific for event measurements.

Chapter 2: Background

2.1 Foundational Math

Point source. As a warm up consider a point light source (PLS) at location (x, y, z) that we would like to image. An ideal-pin-hole camera would capture a sharp image,

$$I_{\text{ideal}}(u, v) = \delta\left(u - f\frac{x}{z}, v - f\frac{y}{z}\right) \tag{2.1}$$

where (u, v) are coordinates on the sensor plane, δ is the Dirac Delta function and f is the focal length. Because pin-hole cameras are light inefficient, conventional cameras use a focusing lens instead. The improved light efficiency comes at the cost of blurrier images depending on the optical system's point-spread-function (PSF). For a PLS, the wavefront that arrives at the optical plane is given as a spherical wavefront,

$$\phi^{DF}(u, v) = \exp\left(ik\sqrt{(u - x)^2 + (v - y)^2 + z^2}\right). \tag{2.2}$$

where $k = 2\pi/\lambda$ is the wavenumber. The PSF h can be modeled with Fourier optics theory [5],

$$h = |\mathcal{F}\left[A \exp\left(i\phi^{DF} + i\phi^{M}\right)\right]|^2 \tag{2.3}$$

where A is the amplitude modulation caused by blocking light and ϕ^M is the phase modulation caused by glass thickness. Then, a PLS captured by a regular camera is

$$I_{\text{blurry}} = h(x, y, z) * I_{\text{ideal}} \tag{2.4}$$

$$= h(x, y, z). \tag{2.5}$$

Complex Scenes We now extend this "encoding" idea for complex scenes. If the PSF is uniform over the image, $h(x, y, z) = h(z)$, and the scene has constant depth (such as imaging a painting), the image formation model is simply the convolution of the sharp ideal image with the PSF for the depth,

$$I_{\text{blurry}} = h(z) * I_{ideal}. \tag{2.6}$$

However, most real-world scenes have variable depth. Prior works have adopted a simple linear image formation model [6],

$$I_{\text{blurry}} = \sum_{d=1}^{D} h(d) * (I_{ideal} \cdot O_d) \tag{2.7}$$

where $\{1, \ldots, D\}$ represent a set of discrete depth layers and O_d represents the occlusion mask at depth d. This model is fast to compute, but can be inaccurate near depth boundaries. A non-linear image formation model has been proposed [7]:

$$I_{\text{blurry}} = \sum_{d=1}^{D} \frac{h(d) * I_{ideal}}{h(d) * \sum_{d'=1}^{d} O_{d'}} \prod_{d'=d+1}^{D} \left(1 - \frac{h(d) * O_d}{h(d) * \sum_{d'=1}^{d} O_{d'}}\right). \tag{2.8}$$

2.2 Depth Estimation

Extracting 2D information from images tends to be a significantly easier task than extracting depth, hence, monocular depth estimation is often the bottleneck in 3D tracking performance. Structured light projectors [8] or time-of-flight sensors [9] use active illumination to extract depth information. Given these methods' reliance on an internal light source, performance can degrade in adverse lighting conditions. If we allow multiple views, stereo [10] or structure from motion [11] can triangulate 3D position. These methods are sensitive to occlusion and texture-less scenes and require multiple calibrated cameras. Many neural network approaches with all-in-focus CMOS images as input have been proposed [12, 13, 14, 15]. Recently, event-based depth estimation has made significant progress with neural networks [16, 17, 18, 19, 20]. Spiking neural networks have been proposed for spiking cameras, which similar to event cameras, offer asynchronous readout of pixels [21].

2.3 Optical Design

Designed optics generally fall into three buckets: First are heuristic-based such as the Double-Helix PSF [22]. These are designed using human-intuition about what blur pattern may be useful for the task. Second are information optimal designs [23]. These theoretically maximize the information content in the image measurement about the parameters we care about. Lastly are end-to-end designed. Because designed PSFs are only as good as the "decoding" algorithm meant to extract the hidden information, researchers have proposed simultaneously learning optical parameters and the algorithm [6]. By using a differentiable model for light prop-

agation, back-propagation can be used to update optical parameters jointly with neural network parameters.

Optics based approaches for depth estimation use sensors and optical setups to encode and recover depth information. Many methods use the depth-dependent blur induced by the imaging system to estimate the depth of pixels in an image [24]. These approaches compare the blur at different ranges to the expected blur caused by an aperture focused at a fixed distance.

Groups improved on this idea by implementing coded apertures, retaining more high frequency information about the scene to disambiguate depths [4]. Similar to depth estimation tasks, static phase masks have been used to produce tailored PSFs more *invariant* to depth, allowing for extended depth-of-field imaging [25]. However, these optically driven approaches with numerical analysis have been passed in performance by modern deep neural networks, allowing for joint optimization of optical elements and neural reconstruction networks.

Many methods have engineered phase masks with specific depth qualities. By maximizing Fisher information for depth, the coded image theoretically will have the most amount of depth cues as possible [23]. Deep learning techniques can be used to jointly train the optical parameters and neural network based estimation methods. The idea is that one can "code" an image to retain additional information about a scene, and then use a deep neural network to produce reconstructions. By using a differentiable model for light propagation, back-propagation can be used to update phase mask values simultaneously with neural network parameters [6].

Designed optics has had wide success in various applications such as extended depth-of-field imaging [26, 27, 7, 28], depth estimation [6, 29, 7], hyper-spectral sensing [30, 31], high-dynamic-range imaging [32, 33], holography [34, 35], privacy preservation [36], high speed imaging [37], localization microscopy [38, 39], line estimation [40], super resolution [27], and

seeing through occlusions [41, 42].

To the best of our knowledge designed optics with dynamic phase masks or event cameras has been largely unexplored.

Chapter 3: Dynamic Phase Mask Design

Point-spread-function (PSF) engineering is a powerful computational imaging technique wherein a custom phase mask is integrated into an optical system to encode additional information into captured images. Used in combination with deep learning, such systems now offer state-of-the-art performance at monocular depth estimation, extended depth-of-field imaging, lensless imaging, and other tasks. Inspired by recent advances in spatial light modulator (SLM) technology, this work answers a natural question: Can one encode additional information and achieve superior performance by changing a phase mask dynamically over time? We first prove that the set of PSFs described by static phase masks is non-convex and that, as a result, time-averaged PSFs generated by dynamic phase masks are *fundamentally more expressive*. We then demonstrate, in simulation, that time-averaged dynamic (TiDy) phase masks can leverage this increased expressiveness to offer substantially improved monocular depth estimation and extended depth-of-field imaging performance.

3.1 Introduction

Extracting depth information from an image is a critical task across a range of applications including autonomous driving [43, 44], robotics [45, 46], microscopy [47, 48], augmented reality [49, 50], and perspective editing [51, 52]. To this end, researchers have developed engineered

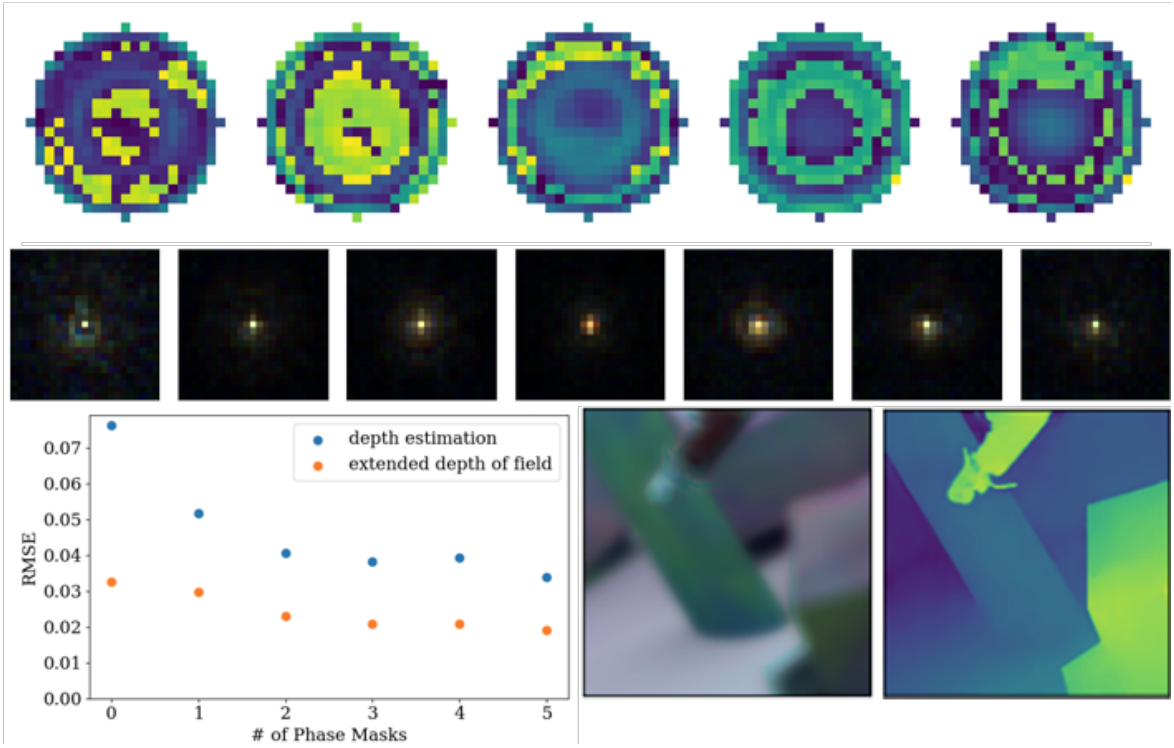


Figure 3.1: **Time-averaged Dynamic PSFs** Top: Phase mask sequence that was optimized to perform simultaneous extended depth-of-field imaging and monocular depth estimation. Middle: Proposed TiDy PSFs at specific depths. Bottom left: Depth estimation and all-in-focus imaging performance improve as one averages over more phase masks. Bottom right: Depth-encoded image and reconstructed depth map.

phase masks and apertures which serve to encode depth information into an image [4, 38]. To optimize these phase masks, recent works have exploited deeps learning: By simultaneously optimizing a phase mask and a reconstruction algorithm "end-to-end learning" is able to dramatically improve system performance [6, 27].

Most existing works have focused on learning or optimizing a single phase mask for passive depth perception. We conjecture that this restriction leaves much room for improvement. Perhaps by using an SLM to introduce a sequence of phase masks over time, one could do much better.

Supporting this idea is the fact, which we prove in Theorem 3.2.3, that the set of PSFs described by a single phase mask is non-convex. This implies that time-averaged PSFs, which

span the convex hull of this set, can be significantly more expressive. In this work, we exploit the PSF non-convexity by developing a multi-phase mask end-to-end optimization approach for learning a sequence of phase masks whose PSFs are averaged over time.

This work's central contributions are as follows:

- We prove the set of PSFs generated by a single phase mask is non-convex. Thus, dynamic phase-masks offer a fundamentally larger design space.
- We extend the end-to-end learning optics and algorithm design framework to design a dynamic set of phase masks.
- We demonstrate, in simulation, that time-averaged PSFs can achieve superior monocular depth estimation and extended depth-of-field imaging performance.

3.2 Theory

Micro-electromechanical system (MEMS) based SLMs offer high framerates but have limited phase precision due to heavy quantization [53]. As [35] noted, intensity averaging of multiple frames can improve quality by increasing effective precision to overcome quantization. Our key insight is that even as SLM technology improves, intensity averaging yields a more expressive design space than a single phase mask. This is supported by the claim that the set of PSFs that can be generated by a single phase mask is non-convex. We provide a rigorous proof for the claim as follows.

Theorem 3.2.1. *The set of PSFs that can be generated by a phase mask with an infinite aperture ($A(x) = 1 \; \forall \; x \in \mathbb{R}^2$) is non-convex.*

Proof. Consider two tilt shift phase masks, $M_1(x) = ax$ and $M_2(x) = -ax$, with respect to a coordinate x with non-zero a. The corresponding averaged PSF is $\frac{1}{2}\delta(a) + \frac{1}{2}\delta(-a)$, where δ denotes a Dirac delta function. To realize this PSF with a single phase mask we would need the field at the aperture to satisfy

$$|\mathcal{F}(E)|^2 = \frac{1}{2}\delta(a) + \frac{1}{2}\delta(-a). \tag{3.1}$$

This implies

$$E(x) = \frac{1}{\sqrt{2}}e^{-iax}e^{i\gamma_1} + \frac{1}{\sqrt{2}}e^{iax}e^{i\gamma_2}, \tag{3.2}$$

$$= \frac{2}{\sqrt{2}}e^{i\frac{\gamma_1+\gamma_2}{2}}\cos\left(ax + \frac{\gamma_1 - \gamma_2}{2}\right) \tag{3.3}$$

for some $\gamma_1, \gamma_2 \in \mathbb{R}$. $E(x)$'s amplitude varies according to a cosine and thus cannot be realized by a phase-only mask. □

We now demonstrate the non-convexity claim holds for finite apertures on a discrete grid.

Definition 3.2.1. $A \in \{0,1\}^{N\times N}$ is some valid aperture with a non-zero region S such that there exists lines L_1 and L_2 where S can be contained between them, and $L_1 \parallel L_2$ and $u = S \cap L_1$ and $v = S \cap L_2$ are single points (Figure 3.2).

This definition of A supports most commonly used apertures including but not limited to circles, squares, and n-sided regular polygons. See supplement for proof for all shapes.

Definition 3.2.2. Let $T_A(N)$ be the set of $N \times N$ matrices in $\mathbb{T}^{N\times N}$ with non-zero support A, i.e. the matrix is supported only where $A = 1$, where $\mathbb{T}$ is the complex unit circle.

The PSF induced by a phase mask M can be modeled as the squared magnitude of the Fourier transform of the pupil function f [6].

Definition 3.2.3. Let $f : \mathbb{R}^{N\times N} \to T_A(N)$ be defined by

$$f(M) = A \odot \exp(iD + icM) \tag{3.4}$$

where $\odot$ denotes entry-wise multiplication, and $D \in \mathbb{R}^{N\times N}$ and $c \in \mathbb{R} - \{0\}$ (the reals except for 0) are fixed constants.

Definition 3.2.4. Let $g : T_A(N) \to \mathbb{R}^{N\times N}$ be defined by

$$g(X) = \frac{|\mathcal{F}(X)| \odot |\mathcal{F}(X)|}{\|\mathcal{F}(X)\|_F^2} \tag{3.5}$$

where $\mathcal{F}$ denotes the discrete Fourier Transform with sufficient zero-padding, $|\cdot|$ denotes entry-wise absolute value, and $\|\cdot\|_F$ denotes the Frobenius norm.

Lemma 3.2.2. *From fourier optics theory [5], any single phase mask's PSF at a specific depth can be written as*

$$PSF = g \circ f.$$

Theorem 3.2.3. *The range of PSF is not a convex set.*

Proof. f is clearly surjective, so it suffices to argue the range of g is not convex. Assume by way of contradiction that the range of g is convex. Then, for all $X^{(1)}, \ldots, X^{(k)} \in T_A(N)$ there exists

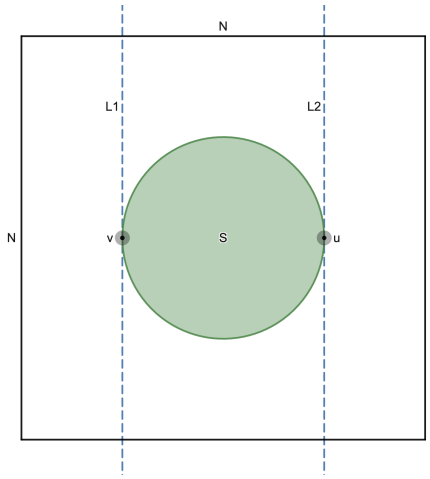


Figure 3.2: **Example aperture that satisfies constraints on A.** The aperture is fitted between parallel lines $L1$ and $L2$, which only intersect the aperture at one point each. Common aperture shapes fit into these constraints.

$Y \in T_A(N)$ such that $g(Y) = \frac{1}{k}\sum_{i=1}^{k} g(X^{(i)})$. By Parseval's Theorem,

$$\|\mathcal{F}(X)\|_F^2 = N^2\|X\|_F^2 = N^2 \sum_{i=0}^{N}\sum_{j=0}^{N} A_{i,j} \tag{3.6}$$

so the condition is

$$|\mathcal{F}(Y)| \odot |\mathcal{F}(Y)| = \frac{1}{k}\sum_{i=1}^{k} |\mathcal{F}(X^{(i)})| \odot |\mathcal{F}(X^{(i)})| \tag{3.7}$$

or equivalently

$$\mathcal{F}(Y) \odot \overline{\mathcal{F}(Y)} = \frac{1}{k}\sum_{i=1}^{k} \mathcal{F}(X^{(i)}) \odot \overline{\mathcal{F}(X^{(i)})}. \tag{3.8}$$

Then the cross-correlation theorem reduces it to

$$\mathcal{F}(Y \star Y) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{F}(X^{(i)} \star X^{(i)}) \tag{3.9}$$

where $\star$ denotes cross-correlation. Because the Fourier Transform is linear we finally have

$$Y \star Y = \frac{1}{k} \sum_{i=1}^{k} X^{(i)} \star X^{(i)}. \tag{3.10}$$

Therefore, the convexity of the range of g is equivalent to the convexity of the set $\{X \star X : X \in T_A(N)\}$. We will show the set's projection onto a particular coordinate is not convex.

$$(X \star X)_{s,r} = \sum_{i=0}^{N} \sum_{j=0}^{N} X_{i,j} \overline{X_{i+s,j+r}} \tag{3.11}$$

where we adopt the convention that $X_{s,r} = 0$ when $s, r > N$ or $s, r < 0$. Take the points u and v from the definition of A (3.2.1). Also observe that correlation can be represented geometrically as shifting $\overline{X}$ over X. In this representation, notice that as the shift (s, r) approaches $v - u$, the non-zero overlap between X and $\overline{X}$ shifted by (s, r) approaches 1 by construction. That is, when L_1 is shifted to overlap L_2, u and v will be the only non-zero overlaps between the shifted and original non-zero points (Figure 3.3). No other non-zero points can overlap above or below L_2 by definition of S. Therefore, $(X \star X)_{v-u}$ becomes

$$X_u \overline{X_v} + \sum_{i=1}^{N^2-1} 0. \tag{3.12}$$

Because $X_u \overline{X_v} \in \mathbb{T}$, $(X \star X)_{v-u} \in \mathbb{T}$ which is a non-convex set. Therefore, the set of correlation's of values on the complex unit circle masked by A is also not convex, and so is PSF.

□

Time-averaged PSFs span the convex hull of the set of static-mask PSFs, meaning there

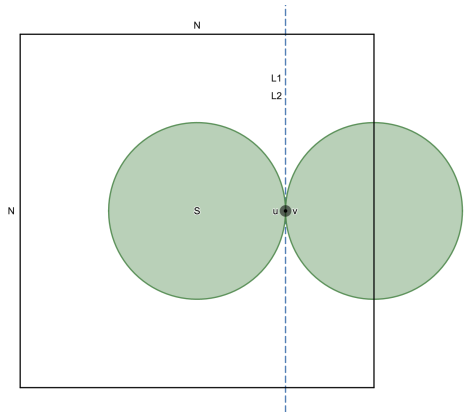


Figure 3.3: **Geometric interpretation of correlation $(\mathbf{X} \star \mathbf{X})_{\mathbf{v-u}}$.** The figure represents the correlation step when the shift is $v - u$. Notice that only u and v overlap once the shift is applied.

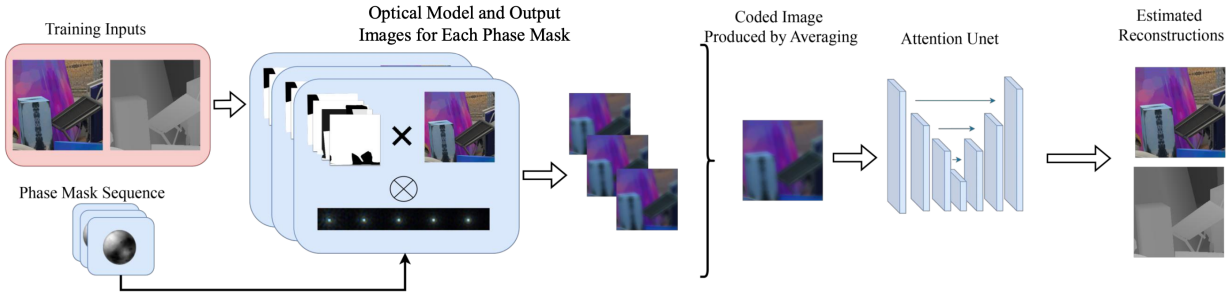


Figure 3.4: **Multi-phase mask forward model overview.** A sequence of phase masks are used to generate a sequence of depth-dependent PSFs. These PSFs are convolved with depth masked clean images to simulate depth dependent convolution. The images produced by each phase mask are averaged to create a coded image which is fed into an attention U-Net. The reconstruction loss is back-propagated end-to-end through the network and the optical model to design phase masks and algorithms capable of performing monocular depth estimation and extended depth-of-field simultaneously.

exists some PSFs achievable only through intensity averaging PSFs from a sequence of phase masks.

3.3 Multi-Phase Mask Optimization

3.3.1 Optical Forward Model

Similar to PhaseCam3D [6], we model light propagation using Fourier optics theory [5]. In contrast to previous work, we compute the forward model (2.7) for multiple phase masks, producing a stack of output images, which form our coded image when averaged. This coded image simulates the recorded signal from imaging a scene using a sequence of phase masks in a single exposure (Figure 3.4).

3.3.2 Specialized Networks

For the monocular depth estimation task, we use the MiDaS Small network [14]. This is a well known convolutional monocular depth estimation network designed to take in natural images and output relative depth maps. The network is trained end-to-end with the phase masks. A mean-squared error (MSE) loss term is defined in terms of the depth reconstruction prediction, $\hat{D}$ and the ground truth depth map D,

$$L_{Depth} = \frac{1}{N}\|D - \hat{D}\|_2^2 \tag{3.13}$$

where N is the number of pixels. This process allows for the simultaneous optimization of the phase masks as well as fine tuning MiDaS to reconstruct from our coded images.

For the extended depth-of-field task, we use an Attention U-Net [54] to reconstruct all-in-focus images. The network is optimized jointly with the phase mask sequence. To learn a

reconstruction $\hat{I}$ to be similar to the all-in-focus ground truth image I, we define the loss term using MSE error

$$L_{AiF} = \frac{1}{N}\|I - \hat{I}\|_2^2 \tag{3.14}$$

where N is the number of pixels.

3.3.3 Joint Task Optimization

We also present an alternative to the specialized networks: a single network jointly trained for monocular depth estimation and extended depth-of-field using a sequence of phase masks. This network has a basic Attention U-Net architecture outputting 4 channels representing depth maps as well as all-in-focus images. Similar to prior works, we use a combined loss function, adding a coefficient to weight the losses for each individual task:

$$L_{total} = \lambda_{Depth} L_{Depth} + \lambda_{AiF} L_{AiF}. \tag{3.15}$$

3.4 Experimental Details

3.4.1 Training Details

We use the FlyingThings3D from Scene Flow Datasets [55], which uses synthetic data generation to obtain all-in-focus RGB images and disparity maps. We use the cropped 278×278 all-in-focus images from [6]. In total, we use 5077 training patches and 419 test patches.

Both the optical layer and reconstruction networks are differentiable, so the phase mask

sequence and neural network can be optimized through back-propagation. Each part is implemented in PyTorch. During training, we use the Adam [56] optimizer with parameters $\beta_1 = 0.99$ and $\beta_2 = 0.999$. The learning rate for the phase masks is 10^{-8} and for the reconstruction network it is 10^{-4}, and the batch size was 32. Finally, training and testing were performed on NVIDIA Quadro P6000 GPUs.

We parameterize 23×23 phase masks pixel-wise as [28] found pixel-wise parameterization to produce the best overall performance. The monocular depth estimation task uses the MiDaS Small architecture pretrained weights for monocular depth estimation downloadable from PyTorch [14]. The extended depth-of-field task pretrains an Attention U-Net with a fixed Fresnel lens for 300 epochs. For the joint task, we set $\lambda_{Depth} = \lambda_{AiF} = 1$ to balance overall performance, and we pretrain the Attention U-Net for 300 epochs with a fixed Fresnel lens. In simulation, the red, blue, and green channels are approximated by discretized wavelengths, 610 nm, 530 nm, and 470 nm respectively. Additionally, the depth range is discretized into 21 bins on the interval $[-20, 20]$, which is larger than previous works.

3.4.2 Evaluation Details

For ablation studies on our method, we used the testing split of the FlyingThings3D set for both monocular depth estimation and extended depth-of-field imaging [55]. For comparisons to existing work, we also tested our monocular depth estimation network on the labeled NYU Depth v2 set [57]. The ground truth depth maps were translated to layered masks for the clean images by bucketing the depth values into 21 bins, allowing us to convolve each depth in an image with the required PSF. We use root mean squared error (RMSE) between ground truth and estimated

depth maps for depth estimation and RMSE between ground truth and reconstructed all-in-focus images for extended depth-of-field imaging. We also use peak signal-to-noise ratio (PSNR) and structural similarity index [58] (SSIM) for extended depth-of-field imaging.

3.5 Results

We compare our time averaged dynamic PSF method to the state-of-the-art methods for both extended depth-of-field imaging and monocular depth estimation. The relevant works we compare to are as follows:

1. PhaseCam3D [6] used a 23×23 phase mask based on 55 Zernike coefficients. The phase mask parameters were then end-to-end optimized with a U-Net reconstruction network to perform depth estimation.

2. Chang et al. [29] used a singlet lens introducing chromatic aberrations with radially symmetric PSFs. Similar to [6], the lens parameters were also then end-to-end optimized.

3. Ikoma et al. [7] used a radially symmetric diffractive optical element (DOE). The blurred image was preconditioned with an approximate inverse of the PSF depth dependent blur. The RGB image stack was fed into a U-Net to produce both an all-in-focus image and a depth map. The DOE and U-Net parameters were optimized in an end-to-end fashion.

4. Liu et al. [28] used various phase mask parameterizations with the same U-Net architecture as [7]. One method used pixel-wise height maps (PW) and the other introduced orbital angular momentum (OAM).

5. Sitzmann et al. [27] implements a single DOE based on Zernike coefficients, and solves

Method	FlyingThings3D	NYUv2
PhaseCam3D [6]	0.521	0.382
Chang et al. [29]	0.490	0.433
Ikoma et al. [7]	0.184	-
MiDaS [14]	-	0.357
ZoeDepth [59]	-	0.277
TiDy (1)	0.026	0.259
TiDy (5)	**0.019**	**0.175**

Table 3.1: **RMSE comparison of monocular depth estimation methods.** We present quantitative results on two datasets to compare to state of the art optical and single shot monocular depth estimation methods. Our method performs best with our 5 phase mask system achieving the lowest error on both datasets.

the Tikhonov-regularized least-squares problem to reconstruct an all-in-focus image.

6. MiDaS [14] and ZoeDepth [59] are state of the art single shot monocular depth estimation methods with all-in-focus images as inputs.

Because both [7] and [28] simultaneously learn all-in-focus images and depth maps, when comparing against our specialized methods, we take their best performing weighting of each task.

Individual Tasks. For monocular depth estimation, our specialized method using a sequence of 5 phase masks trained for 300 epochs outperforms prior work on FlyingThings3D (Table 3.1). Additionally, our approach performs significantly better and achieves lower error than previous methods on NYUv2 without any additional fine tuning. For extended depth-of-field, our specialized method using a sequence of 5 phase masks outperforms prior work on FlyingThings3D (Table 3.2). This demonstrates the benefit of multi-phase mask learning on computational imaging tasks.

Method	RMSE↓	PSNR↑	SSIM↑
Liu et al. [28]	-	29.80	-
Ikoma et al. [7]	0.1327	31.88	0.905
Sitzmann et al. [27]	-	32.44	-
TiDy (1)	0.0148	37.33	0.968
TiDy (5)	**0.0092**	**41.11**	**0.989**

Table 3.2: **Comparison of extended depth-of-field imaging methods.** We present quantitative results on FlyingThings3D to compare to state-of-the-art. Our method performs best with our 5 phase mask system achieving the best PSNR.

Method	All-in-focus PSNR↑	Depth RMSE↓
Ikoma et al. [7]	31.88	0.191
Liu et al. [28] - PW	29.80	0.056
Liu et al. [28] - OAM$_t$	25.86	0.053
TiDy (1)	31.20	0.052
TiDy (5)	**34.79**	**0.034**

Table 3.3: **Comparison of multi-objective optimization of extended depth-of-field imaging and depth estimation methods.** We compare quantitative results on FlyingThings3D to the state-of-the-art. Our method performs best with our 5 phase mask system achieving the best balance between objectives.

Multi-Objective Optimization. We also evaluate our method against other joint all-in-focus and depth map learning approaches. This problem is challenging because good depth cues to produce depth maps is antithetical to producing an all-in-focus image. Our combined 5 phase mask trained for 300 epochs approach outperforms prior jointly trained approaches (Table 3.3).

3.6 Ablation Studies

3.6.1 Effect of Phase Mask Sequence Length

For both all-in-focus imaging and depth estimation, we vary the phase mask count that the end-to-end system is trained with to gauge the benefits of using multiple phase masks. The

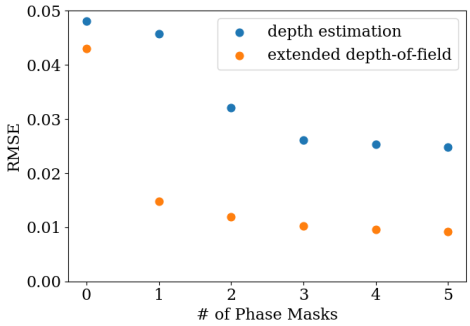


Figure 3.5: **RMSE for specialized tasks** for each phase mask sequence length. RMSE decreases with respect to phase mask sequence length for both specialized extended depth-of-field imaging and monocular depth estimation tasks. 0 phase masks refers to a reconstruction neural network with a fixed Fresnel lens.

forward model and initial phase masks were held standard while the phase mask count was varied. The resulting networks were evaluated at convergence. For the extended depth-of-field task, the masks were all initialized with random noise uniform from 0 to 1.2×10^{-6}. For the depth estimation task, the masks were initialized with the Fisher mask with added Gaussian noise parameterized by a 5.35×10^{-7} mean and 3.05×10^{-7} standard deviation.

End-to-end optimization on each task with a specialized network yielded improved performance as the phase mask count increased, visualized in Figure 3.5. This result implies that sequences of phase masks are successful in making the PSF space more expressive. Additionally, even for the more complex joint task, learning a system that can produce both all-in-focus images and depth maps, error decreases with phase mask count until a plateau is reached (Figure 3.6).

3.6.2 All-in-Focus without Reconstruction Networks

A phase mask generating a PSF of the unit impulse function at every depth would be ideal for extended depth-of-field as each depth would be in focus. If possible, this phase mask would

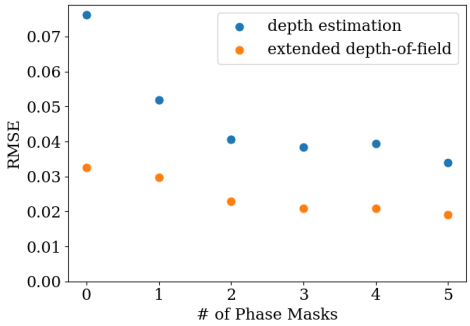


Figure 3.6: **RMSE for joint optimization of monocular depth estimation and extended depth-of-field imaging** for each phase mask sequence length. RMSE decreases with respect to phase mask sequence length for this complex joint task, demonstrating the benefit of multi-phase mask learning. 0 phase masks refers to a reconstruction neural network with a fixed Fresnel lens.

not require any digital processing. We optimize phase mask sequences of varying lengths to produce an averaged PSF close to the unit impulse function for all depths. For each sequence length, phase masks are optimized using MSE loss between the unit impulse function and the averaged PSF at each depth until convergence. We ran 1000 trials of random phase mask initialization for each length. Observe that a side-effect of longer phase masks is training stability. The range of RMSE between the simulated capture image and ground truth all-in-focus image decreases as the sequence length increases (Figure 3.7). This indicates training longer sequences is more resilient to initialization.

3.6.3 Phase Mask Initialization for Depth Perception

Deep optics for depth perception can be very dependent on the initialization of optical parameters before training [6]. To find the extent of the effect of mask initialization on performance, we varied the the initial phase masks while keeping number of masks, the optical model, and duration of training fixed. We trained for 200 epochs. We tested four initializations of sequences

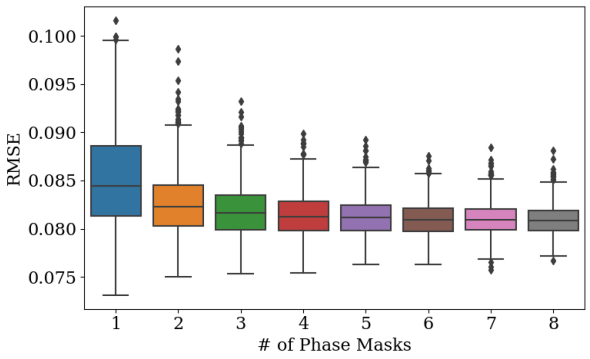


Figure 3.7: **All-in-focus imaging RMSE distribution** for each phase mask length without a reconstruction network. The best RMSE for each phase mask count has low correlation with respect to phase mask sequence length, but the variance of RMSE decreases.

Initialization	RMSE↓
1 Fisher + All noise	0.0329
1 Fisher + Fisher w/ Noise	0.0271
All noise	0.0254
3 Fisher + Fisher w/ Noise	**0.0207**

Table 3.4: **Quantitative evaluation of phase mask initializations.** Four sequence initializations are evaluated on the monocular depth estimation task. Ultimately, 3 Fisher masks and 2 noisy Fisher masks have the best performance after training.

of 5 phase masks as shown in Figure 3.8. The first was uniformly distributed noise from 0 to 1.2×10^{-6}. The second was the first mask in the sequence set to a Fisher mask while the rest are uniform noise. The third is setting each mask to a rotation of the Fisher mask and adding Gaussian noise parameterized by a 5.35×10^{-7} mean and 3.05×10^{-7} standard deviation to 4 masks. Lastly, we set each mask to a rotation of the Fisher mask and added noise to only the last two masks in the sequence. Of the four initializations, it is clear that the 3 Fisher masks and 2 Fisher masks with noise performed the best (Table 3.4).

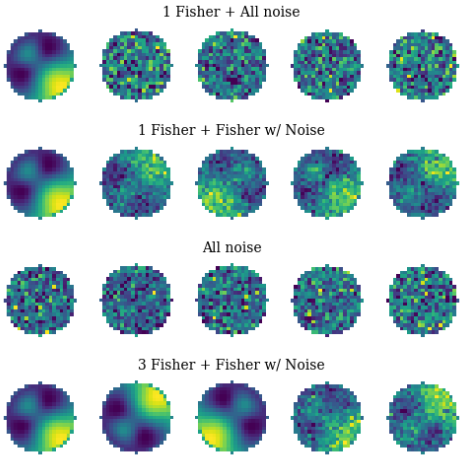


Figure 3.8: **Visualization of phase mask initializations.** Each row represents a different initial phase mask sequence.

3.6.4 Modeling SLM Imperfections

State Switching. Our optical forward model assumes an SLM can swap between two phase patterns instantly. In practice, however, some light will be captured during the intermediate states between phase patterns. These phase patterns, in the worst case, could be random phase patterns, effectively adding noise to our coded images. We model these intermediate states by averaging output images produced by phase masks and the randomized phase patterns weighted by the time that they are displayed for. We model the total exposure time as 100ms, with various durations of switching times from 1 to 16ms per swap. We evaluate our joint optimized network on these new, more noisy, coded images without any additional training (Figure 3.12). Observe that because the 5 phase mask system includes more swaps, performance degrades faster than fewer phase mask systems. However, for short switching times, 5 phase masks still outperform the others without needing any fine tuning.

Quantization. Current high-speed MEMS based SLM technology experience heavy quantization

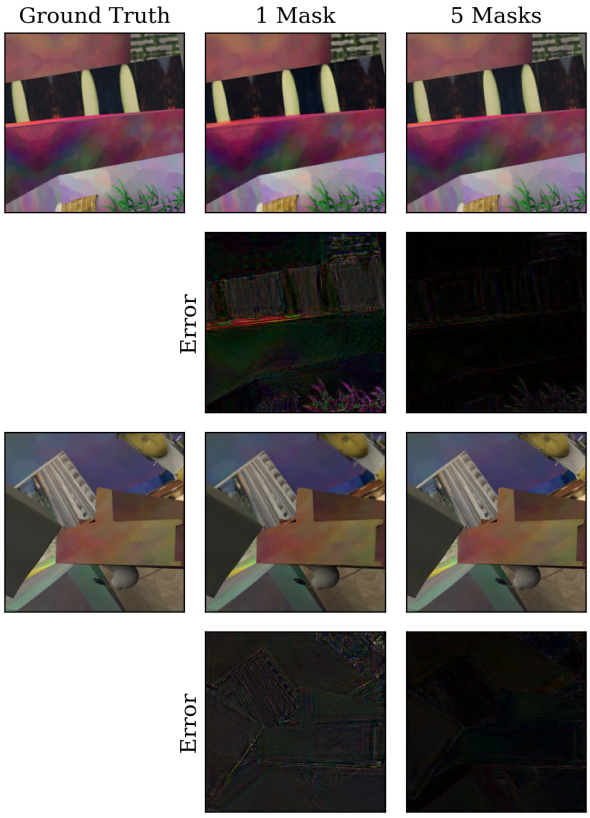


Figure 3.9: **Qualitative results of a specialized network on extended depth-of-field imaging.** Both 1 and 5 phase mask systems are evaluated on FlyingThings3D. Error is computed pixel wise between the ground truth all-in-focus image and the reconstructed output and is boosted by a factor of 3. Notice that the 5 phase mask system introduces minimal error.

[53]. We model the effects of quantization error by adding varying amounts of noise during inference. As demonstrated in Figure 3.13, time-averaged PSFs are robust to quantization noise: 5 masks quantized to 4-bits outperform one mask quantized to 32-bits. This is intuitive as multiple masks can leverage averaging to achieve better precision and remove noise [35].

3.6.5 A Path Towards Better Single-Mask Designs

The performance of end-to-end designed phase masks is highly sensitive to how they are initialized [6]. Can we use our 5-mask design to provide a better initialization for a 1-mask

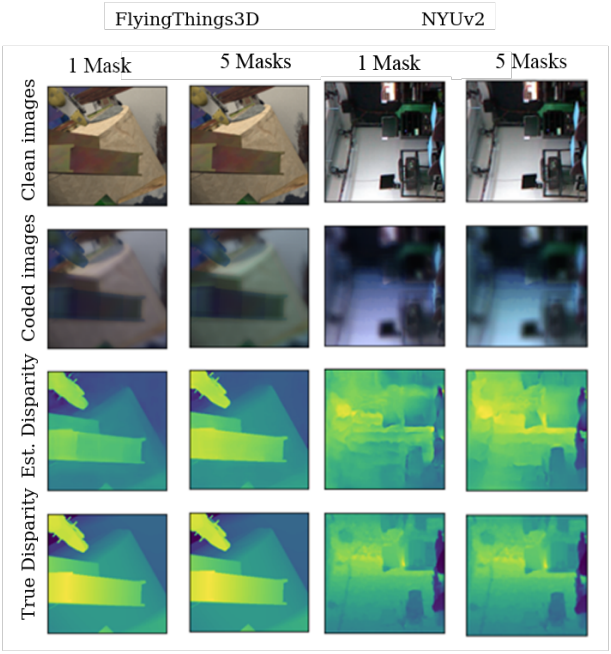


Figure 3.10: **Qualitative results of a specialized networks on monocular depth estimation.** Performance using the five phase mask method outperforms one phase mask on both datasets.

design?

Starting from random initializations, we optimized 200 single phase masks to minimize the mean squared error between each of their PSFs and the time-averaged PSF formed by our optimized 5-mask system. We then took the single mask whose PSF was closest to the 5-mask design and fine-tuned our reconstruction network using its PSF. This static design improved performance over a randomly initialized end-to-end trained 1-mask system; $+2.67$ PSNR for all-in-focus imaging and -0.003 RMSE for depth estimation. However, its performance was still strictly worse than a 5-mask system.

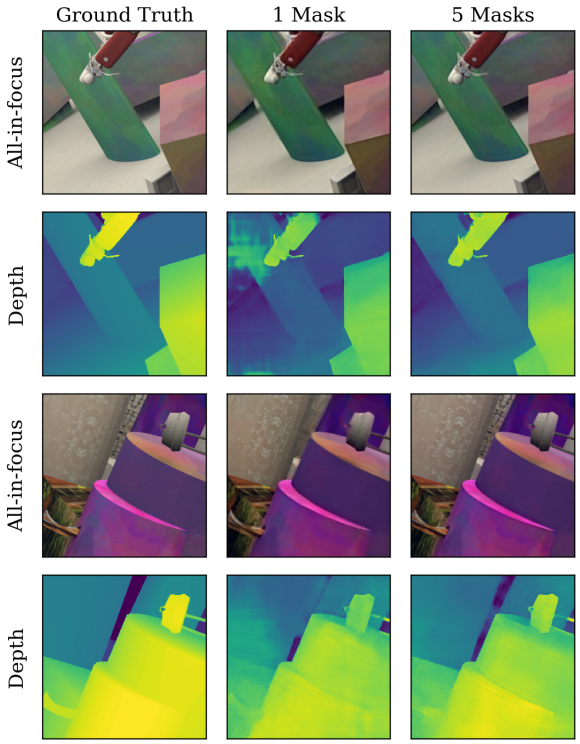


Figure 3.11: **Qualitative results of a joint optimized system for extended depth-of-field imagining and monocular depth estimation.** Both one and five phase mask networks are evaluated on the FlyingThings3D datasets. Notice that five masks has fewer artifacts than a single mask.

3.6.6 Modulating Amplitude and Phase

As illustrated in our proof of Theorem 3.2.1, "complex" masks which modulate both amplitude and phase are more expressive than phase-only masks. Using end-to-end learning, we optimized a pixel wise complex mask (initialized with a random phase mask and all ones amplitude mask) on the joint depth estimation and all-in-focus imaging task. The complex mask offered substationally improved performance; $+2.49$ PSNR for all-in-focus imaging and -0.015 RMSE for depth estimation. However, its performance too was still strictly worse than a 5-mask system. We conjecture this is due to improved training stability of the 5-mask system.

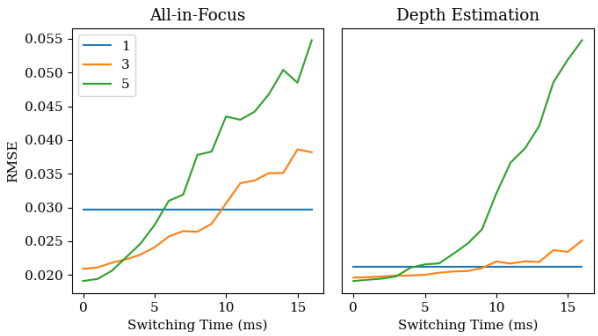


Figure 3.12: **Effect of switching time on joint system performance.** Reconstruction error across phase mask counts as a function of switching time with 100ms overall exposure. Performance of the jointly optimized system degrades as the switching time between phase masks increases, as expected. Our system still performs well when the time spent switching is less than 25% of the overall exposure.

3.6.7 Time Averaging Compared to Multi-Shot Sequences

Our optical model images a static scene through multiple phase masks which we switch between over the course of single exposure (Figure 3.14a). A natural question, then, is why limit ourselves to a single exposure. Why not capture a burst of images, each with a different phase mask (Figure 3.14b)?

While it is true that superimposing the outputs of multiple PSFs creates challenges in disambiguating outputs from phase masks, it also offers several benefits. First, because we only capture a single frame, our system uses less memory due to less I/O required. Second, imaging in a single exposure is more light efficient. Over a fixed time interval, a single exposure allows you to capture the entirety of the light from the scene. Multi-shot, alternatively, would miss photons during readout between shots.

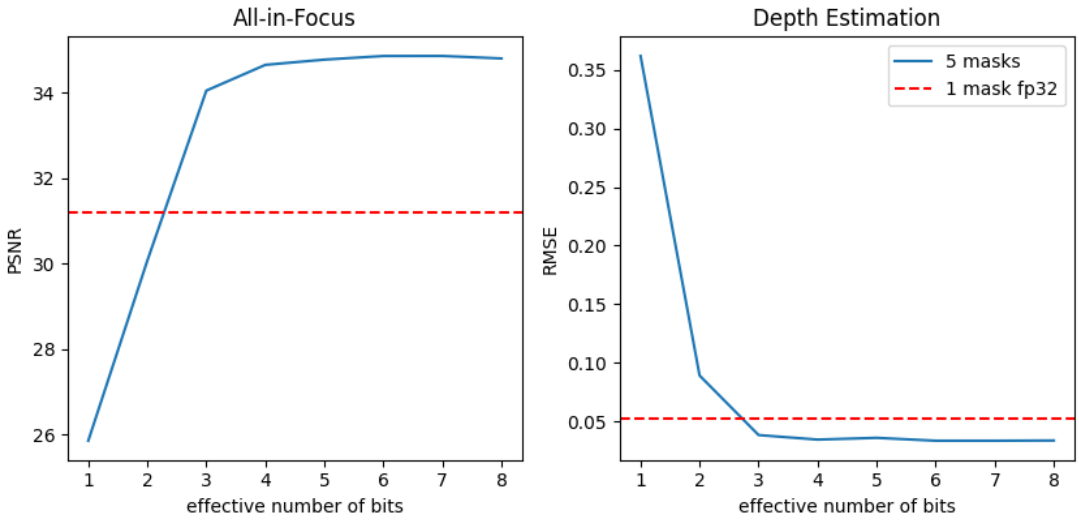
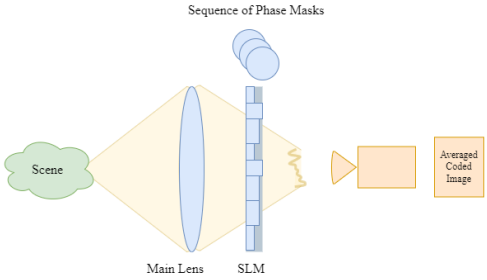


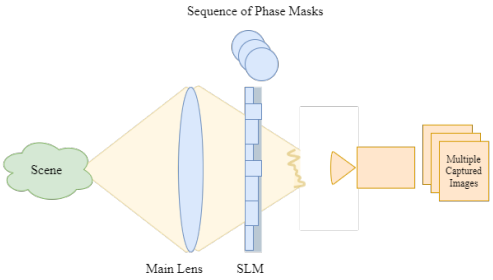
Figure 3.13: **Effect of quantization on joint system performance.** Performance of the jointly optimized system degrades as the number of effective bits decrease; however, even under heavy quantization, our system outperforms single mask setups.

3.7 Limitations

While we were successful in learning dynamic phase masks to improve state-of-the-art performance on imaging tasks, our method carries some limitations. First, our optical model assumes perfect switching between phase masks during training. While evaluation with non-zero switching times showed little degradation of performance, accounting for state switching while training could produce phase masks that are more performant. Our optical model also simulates depths as layered masks over an image, which does not account for blending at depth boundaries. Additionally, our method assumes that scenes are static for the duration of a single exposure. Lastly, though their prices are falling, SLMs are still quite expensive and bulky.



(a) Time averaging phase masks



(b) Multi-shot phase masks

Figure 3.14: **Time averaging and multi-shot optical systems.** Observe that multi-shot systems capture multiple coded images, while time averaging only captures one. This means our system is more light and memory efficient.

3.8 Conclusion

This work is founded upon the insight that the set of PSFs that can be described by a single phase mask is non-convex and that, as a result, time-averaged PSFs are fundamentally more expressive. We demonstrate that one can learn a sequence of phase masks that, when one dynamically switches among them over time, can substantially improve computational imaging performance across a range of tasks, including depth estimation and all-in-focus imaging. Our work unlocks an exciting new direction for PSF engineering and computational imaging system design.

Chapter 4: Optical Design for Event Cameras

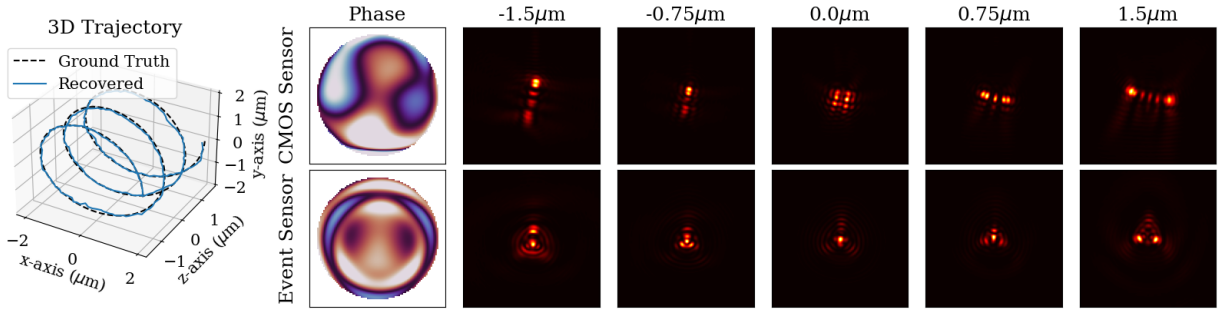


Figure 4.1: **CodedEvent Tracking.** Left: example recovered trajectory using designed optics for an event camera. Right: top row, optimal phase mask design and PSFs for a CMOS sensor, bottom row, our optimal phase mask design and PSFs for an event sensor.

Point-spread-function (PSF) engineering is a well-established computational imaging technique that uses phase masks and other optical elements to embed extra information (e.g., depth) into the images captured by conventional CMOS image sensors. To date, however, PSF-engineering has not been applied to neuromorphic event cameras; a powerful new image sensing technology that responds to changes in the log-intensity of light. This chapter establishes theoretical limits (Cramér Rao bounds) on 3D point localization and tracking with PSF-engineered event cameras. Using these bounds, we first demonstrate that existing Fisher phase masks are already near-optimal for localizing static flashing point sources (e.g., blinking fluorescent molecules). We then demonstrate that existing designs are sub-optimal for tracking moving point sources and proceed to use our theory to design optimal phase masks and binary amplitude masks for this task. To overcome the non-convexity of the design problem, we leverage novel implicit neural

representation based parameterizations of the phase and amplitude masks. We demonstrate the efficacy of our designs through extensive simulations. We also validate our method with a simple prototype.

4.1 Introduction

Single-molecule localization microscopy (SMLM) is a vital tool for resolving nano-scale structures with applications in analysis of protein clusters [60], cell dynamics [61], and electromagnetic effects [62]. Traditional SMLM experiments are limited by the slow capturing process of frame-based CMOS sensors, preventing use in capturing high-speed, dynamic interactions. Recently, [63] showed event cameras are key to enabling high-speed 2D SMLM.

In contrast to traditional CMOS cameras, event cameras are an emerging class of bio-inspired neuromorphic sensors that operate with a high temporal resolution on the order of μs. These sensors are comprised of an asynchronous pixel array, where each pixel records an event when the log intensity change exceeds a set threshold. In addition to having kilohertz time resolution, these sensors are low-power, resistant to constant background noise, and can operate over a high dynamic range [64]. Already, these sensors have proven useful in a range of applications including object tracking [65, 66], gesture recognition [67, 68], and robotics [69, 70].

Just as PSF-engineering allows one to extract additional information using conventional CMOS sensors [23], we believe that event-camera-specific PSF engineering will be the key to enabling high-speed 3D SMLM with event cameras. Unfortunately, existing PSF design theory is not equipped for the event space. In this work, we bridge this gap by developing Cramér Rao Bounds on 3D position estimation for event camera measurements. Leveraging these bounds,

we subsequently develop a novel implicit neural representation for optical elements to design components with improved 3D particle localization capabilities.

Specifically, our principal contributions are as follows:

- We derive the Fisher Information and Cramér Rao Bounds for event camera measurements parameterized by 3D spatial positions.
- We develop novel implicit neural representations for learning both amplitude and phase masks.
- We identify new phase and amplitude designs for optimally encoding 3D information with event cameras.
- We demonstrate in simulation that our designs outperform existing methods at 3D particle tracking.

4.2 Related Microscopy Tracking Work

Originally, single-particle localization was limited to 2D dimensions, where only the x, y coordinates of an emitter are recovered [71]. Similar to works on depth from defocus, the depth of an emitter can be recovered from 2D measurements by considering a microscope's PSF. A standard microscope typically has a PSF resembling the circular Airy pattern; however, because it spreads out quickly its depth resolving range is limited. A few engineered PSFs—such as the double-helix PSF [22]—have since been proposed to improve the imaging range. In particular, Shechtman et al. finds the optimally informative PSF (dubbed the Fisher PSF) for a CMOS sensor to localize the 3D position of a single emitter [23]. A few other techniques for resolving

the 3D location of particles have been proposed such as light-field-microscopy [72] and lensless imaging [73].

Unfortunately, these techniques are limited by the sub-kilohertz readout of conventional CMOS sensors. This hinders their use in imaging fast, dynamic processes such as blood flow [74] and voltage signals [75]. A few ultrafast imaging methods have also been proposed [76, 77, 78, 79] but require high-power illumination which can be phototoxic to certain organic samples. Recently, event cameras have been proposed as an alternative to CMOS sensors for 2D SMLM [63]. Another work proposes extending light-field-microscopy to event cameras to resolve 3D position but requires complex optical setups and sacrifices spatial resolution [80]. By designing optics to encode depth information into event streams, we can enable high-speed 3D SMLM.

4.3 Theory

4.3.1 Event Camera Simulation

Event cameras trigger events with respect to the log of photocurrent $L = \log(I_{\text{blurry}})$ [64] where a pixel's photocurrent is linearly related to the wave intensity at that pixel. Specifically, an event is triggered when the absolute difference between the current intensity at $t + \tau$ and the reference intensity from t, $\Delta L(u, v) = L_{t+\tau}(u, v) - L_t(u, v)$, is greater than some threshold T.

$$O_t(u,v) = \begin{cases} +1 & \Delta L(u,v) > T \\ -1 & \Delta L(u,v) < -T \\ \text{none} & \text{otherwise} \end{cases} \tag{4.1}$$

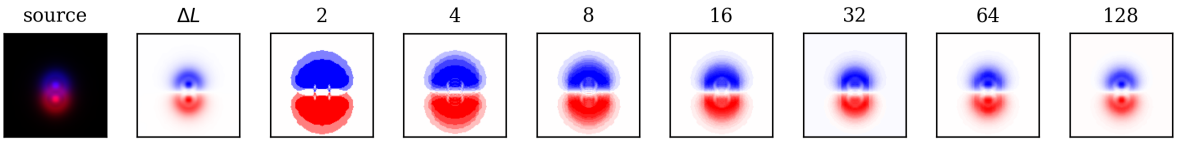


Figure 4.2: **Binning events approximates the log difference as the number of accumulated frames increases.** Consider a point source moving from the blue location to the red location at depth plane 1μm over a fixed time interval in the first image. The second image illustrates the direct access to the difference in (4.2), while the subsequent images demonstrate the effect of accumulating N event frames across the time interval. Observe how large N nearly recovers ΔL, demonstrating the validity of the approximation.

In isolation, each event contains little information; however, a sequence of events can be highly informative [81, 82, 83]. Notably the inceptive event time-surfaces representation suggests the trailing events that occur after the first event correspond to the log-intensity change [84]. Therefore, by binning events over time, one can approximately recover the change in log intensity ΔL. Visuallly, we show the accumulated event frame approaches ΔL as the number of intermediate frames accumulated increases in Figure 4.2. We prove this approximation is at most off by 1 for an idealized event camera in Section 4.9 of the supplement. Therefore, our event measurement (4.1) can be simplified as,

$$O_t = \log\left(I_t\right) - \log\left(I_{t-\tau}\right). \tag{4.2}$$

4.3.2 Information

In the field of statistical information theory, the Fisher Information (FI) reports the amount of information gained about the parameters of a distribution, given a measurement. As such, we can use FI to express the effectiveness of PSFs at encoding depth information. The multi-parameter FI is represented as an $N \times N$ matrix where the i, j entry is defined as the variance of

the score:

$$\mathcal{I}(\theta)_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\theta)\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\theta)\right) \mid \theta\right] \tag{4.3}$$

where θ is the set of parameters, θ_i is the ith parameter, and $f(X;\theta)$ is a probability density function for the distribution observation X is drawn from.

For traditional CMOS sensors, FI has been used to compare coded apertures and phase masks for a wide range of tasks such as depth estimation [85], hyper-spectral imaging [30], and detecting linear structures [40]. Those works have shown that the intrinsic photon shot noise in I^b can be modeled as a Poisson random variable with mean $\lambda = h(x, y, z)$. We derive the FI matrix for an event sensor.

Flashing light. As a warm-up, consider the SMLM technique for event cameras presented in [63], which assumes a blinking labeling model similar to STORM (stochastic optical reconstruction microscopy) [86], PALM (photoactivated localization microscopy) [87] and DNA-PAINT (DNA point accumulation for imaging in nano-scale topography) [88]. With this idealized model of an event camera, $\log I_{t-\tau} = 0$, so (4.2) reduces to

$$O_t = \log I_t. \tag{4.4}$$

By applying e^x to the measurement, we can indirectly measure I_t. Moreover, by applying standard results for FI of a Poisson distribution [89, 90], we can write the FI matrix for an event

camera capturing a blinking particle as:

$$\mathcal{I}(\theta)_{i,j} = \sum_{n}^{N} \frac{1}{h(n) + \beta} \left(\frac{\partial h(n)}{\partial \theta_i} \right) \left(\frac{\partial h_z(n)}{\partial \theta_j} \right) \tag{4.5}$$

where N is the number of pixels, $h(n)$ is the PSF intensity at pixel n, β is background noise, and $\theta = \{x, y, z\}$ corresponds to the 3D location of a point source. Notice that this is the same result as in [85], suggesting that — in the context of blinking particles — the Fisher mask found in [23] for a traditional CMOS camera is also optimal for an event-based sensor.

Generalization. We now derive the positional information content for any event measurement. Rewriting (4.2) with logarithmic rules, we obtain,

$$O_t = \log \frac{I_t}{I_{t-\tau}}. \tag{4.6}$$

The inner expression is drawn from the ratio of Poisson random variables with means λ_t and $\lambda_{t-\tau}$. This can be approximated as a single Normal distribution [91]:

$$\frac{I_t}{I_{t-\tau}} \sim \mathcal{N}\left(\frac{\lambda_t}{\lambda_{t-\tau}}, \frac{\lambda_t}{\lambda_{t-\tau}^2} + \frac{\lambda_t^2}{\lambda_{t-\tau}^3} \right). \tag{4.7}$$

Similar to the flashing light example, we can exponentiate the measurement to recover this ratio. Using the symbolic mathematics solver SymPy [92], we evaluate the expectation in (4.5) with

$\theta = \{x_t, y_t, z_t, x_{t-\tau}, y_{t-\tau}, z_{t-\tau}\}$ and $f(X;\theta)$ as the PDF of the normal distribution, yielding

$$\mathcal{I}(\theta) = \sum_{n}^{N} \frac{\mathcal{D}^T \mathcal{D}}{2\left(\mu+\nu\right)^2} \odot \begin{bmatrix} a & a & a & b & b & b \\ a & a & a & b & b & b \\ a & a & a & b & b & b \\ b & b & b & c & c & c \\ b & b & b & c & c & c \\ b & b & b & c & c & c \end{bmatrix} \tag{4.8}$$

where

$$\mu = \lambda_{t-\tau} = h(x_{t-\tau}, y_{t-\tau}, z_{t-\tau}) + \beta \tag{4.9}$$

$$\nu = \lambda_t = h(x_t, y_t, z_t) + \beta \tag{4.10}$$

$$\mu_i = \frac{\partial}{\partial \theta_i}\mu \tag{4.11}$$

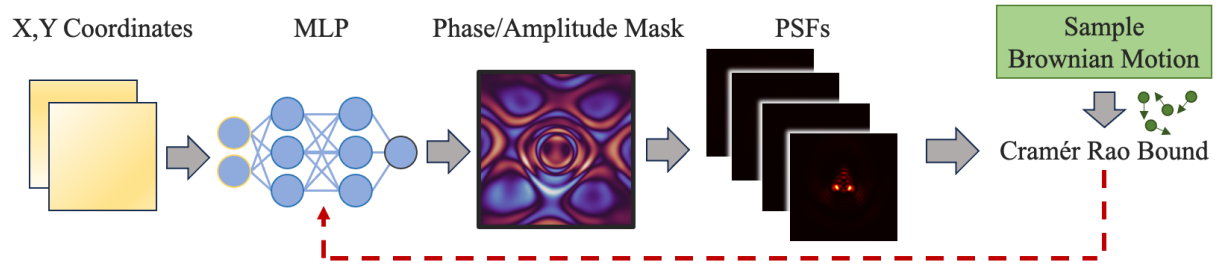
$$\nu_i = \frac{\partial}{\partial \theta_i}\nu \tag{4.12}$$

$$\mathcal{D} = \begin{bmatrix} \mu_x/\mu & \mu_y/\mu & \mu_z/\mu & \nu_x/\nu & \nu_y/\nu & \nu_z/\nu \end{bmatrix} \tag{4.13}$$

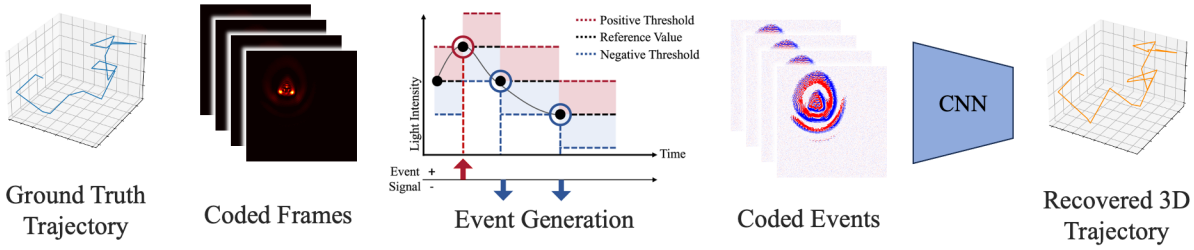
$$a = 2\mu^2\nu + 4\mu^2 + 2\mu\nu^2 + 12\mu\nu + 9\nu^2 \tag{4.14}$$

$$b = -\left(2\mu^2\nu + 2\mu^2 + 2\mu\nu^2 + 7\mu\nu + 6\nu^2\right) \tag{4.15}$$

$$c = 2\mu^2\nu + \mu^2 + 2\mu\nu^2 + 4\mu\nu + 4\nu^2 \tag{4.16}$$



(a) Optical component optimization.



(b) Coded event 3D tracking.

Figure 4.3: **System overview.** (a) An MLP produces a phase or amplitude mask based on a grid of x, y coordinates. The weights are updated through back-propagation of the CRB computed with Brownian Motion. (b) In simulation, coded events are generated by first rendering high-frame-rate coded CMOS frames and converting them to event frames. These measurements are passed to a 3D-tracking algorithm.

4.4 Method

4.4.1 Objective Function

Similar to existing work on 3D tracking for CMOS sensors, we can leverage the FI matrix to optimize optical parameters that efficiently encode depth information [23, 6]. Specifically, we compute the Cramér Rao Bound (CRB), which provides a fundamental bound on how accurately parameters can be estimated given a measurement. If $T(X)$ is the unbiased estimator for parameters θ, then the CRB is

$$CRB_i \equiv \left[\mathcal{I}(\theta)^{-1}\right]_i \leq \mathrm{cov}_\theta \left(T(X)\right)_i . \tag{4.17}$$

Then, the objective function we wish to minimize is

$$\mathcal{L}_{CRB} = \sum_{z \in Z} \sum_{i \in \theta} \sqrt{[\mathcal{I}(\theta)^{-1}]_{i,i}} \tag{4.18}$$

where Z is a set of depth planes.

4.4.2 Optical Parameter Representation

PSF manipulation is typically achieved through designed optical elements such as phase and amplitude masks. In general, phase masks are preferred over binary amplitude masks for their photon efficiency and continuous parametric representation, allowing for optimization via standard gradient descent methods. Inspired by [42], we demonstrate that implicit neural representations can model phase masks in such a way that results in more stable optimization and better-optimized mask designs. We use an architecture similar to the sinusoidal representation network (SIREN) presented in [93] to predict the phase delay caused by the mask at each location (u, v). Input data in $\mathbb{R}^2$ is processed by a four-layer multi-layer perceptron (MLP) with hidden feature size 128, and $\sin$ activation. We refer to this method as *Neural Phase Mask (NPM)*.

Phase masks offer many degrees of freedom and excellent light throughput, but can be relatively expensive to manufacture and are only effective for some frequencies. Meanwhile binary amplitude masks are cheap to manufacture (such as with consumer-grade 3D printers) and can operate across all frequencies (including x-ray), but offer fewer degrees of freedom.

Historically, methods for designing optimal binary apertures have been fundamentally limited due to the lack of optimization techniques for discrete binary parameters. As a result, prior works [4, 94, 95] walk over a restricted search space, leaving ample room for improvement. To

solve this issue, we propose a novel implicit neural representation for binary amplitude masks. We use an MLP to predict the percent of photons blocked at each mask location (u, v). The input in $\mathbb{R}^2$ is processed by a four-layer MLP with hidden feature size 128 and SoftPlus [96] activation. The output to the network is passed through a sigmoid. We refer to this method as *Neural Amplitude Mask (NAM)*.

4.5 Experimental Details

PSFs are simulated for a microscope imaging system with NA$= 1.4$, index of refraction $n = 1.518$, wavelength $\lambda = 550$nm, magnification $M = 111.11$, 4f lens focal length $f = 150$mm, pixel pitch of 49.58μm, and resolution of 256×256. Each phase and amplitude mask is optimized using $\mathcal{L}_{CRB}$ for 10,000 epochs. Because particle motion influences FI, we leverage Monte Carlo sampling while training to maximize information content for all motion directions. For each epoch, we compute the total CRB for 3 random orthogonal motions across 11 depth planes. We use the Adam [97] optimizer with parameters $\beta_1 = 0.99$, $\beta_2 = 0.999$, and a learning rate of 10^{-3}. Training and testing were conducted on NVIDIA RTX A5000 GPUs.

To validate our design's ability to track point sources, we train a Convolutional Neural Network (CNN) to map binned event frames to 3D locations. Events are accumulated over 16 refresh cycles to produce an accumulated event frame. These 256×256 single-channel images are processed by a CNN with 5 convolutional blocks and a linear output head. Each block is followed by batch normalization, ELU activation [98], and max pooling. The output is a normalized length 3 vector representing the position of the particle at a given time step. The CNN is trained on 3 Brownian motion trajectories. Each trajectory is sampled at 16,000 time steps. A 'coded'

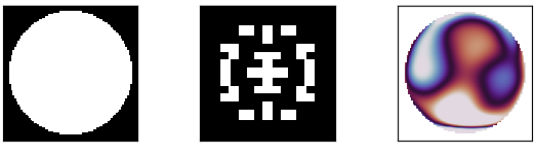


Figure 4.4: **Visualization of non-event camera-specific optical components.** Each component is placed in the same plane as a 150mm focal length lens.

CMOS video frame is simulated by blurring a 300nm emitter with the optical component's PSF for the location and adding Gaussian noise (to simulate other noise sources such as thermal). Next, we generate a 'coded-event-stream' from the high-speed video using standard event camera simulator methods by tracking the per-pixel reference signal [99]. Finally, we bin every 16 frames to produce a 1000-frame 'coded-event-video'. The particle location at the end of the 16-frame bin is considered the ground truth position. We supplement this training with 2000 random starting positions and corresponding motion vectors. Each motion is scaled to have magnitude drawn from $\mathcal{N}(100\text{nm}, 20\text{nm})$. For each position-motion pair, we generate a 16 frame 'coded' CMOS video to accumulate into a 'coded-event-frame'. The CNN is trained for 100 epochs with the Adam optimizer.

We also manufacture a lab prototype to the demonstrate practical benefits of coded apertures for event cameras (see Section S1 in the supplementary materials for details).

4.6 Results

Because designed optics for event cameras is an emerging field, we compare our optimized phase and amplitude mask designs to components designed for traditional CMOS sensors: open aperture/Fresnel lens, Fisher phase mask [23] and Levin et al.'s amplitude mask [4] (Figure 4.4).

	Component	CRB (nm) ↓
	Open Aperture	80.8
Amplitude	Levin et al.	263.3
	NAM (Ours)	50.5
Phase	Fisher	36.3
	NPM (Ours)	**33.1**

Table 4.1: **Average CRB for each optical component** across a 3μm depth range for all 6 position parameters. Phase masks outperform amplitude masks due to higher light efficiency, and our neural-designed phase mask is best.

4.6.1 Cramér Rao Bound

We simulate Brownian motion by sampling 1000 unit direction vectors and independently scaling them by a magnitude drawn from $\mathcal{N}\left(100\text{nm}, 20\text{nm}\right)$. The speed is relative to the event camera refresh rate, with a 1000 accumulated-event-frame per second system, this motion simulates a range of biological processes such as molecular diffusion [100]. We then evaluate the average CRB over the 1000 motions at 30 depth planes spaced evenly on a 3μm range around the focal plane. For all 6 position parameters, we plot the CRB trend with respect to depth (Figure 4.5). Observe that each optical system performs worse as a point source moves away from the focal plane as the defocus change decreases. Although an open-aperture lens is slightly better around the focal plane, its bound increases at a higher rate than the other designs. We also report the average CRB over all parameters and depth slices to demonstrate our neural-based phase mask is best overall (Table 4.1).

4.6.2 3D Tracking

We validate our theoretical results in simulation by tracking a 3D moving emitter across a $8\mu\text{m} \times 8\mu\text{m} \times 4\mu\text{m}$ volume. After training a CNN to decode 3D position from coded event

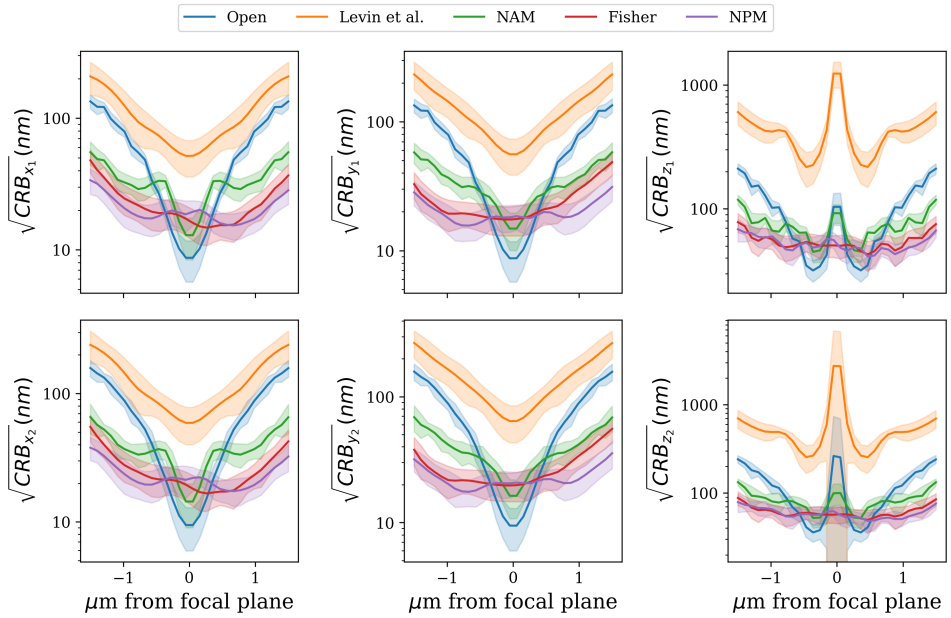


Figure 4.5: **3D localization CRB with respect to depth**. First row: particle's x, y, z position at time $t - \tau$. Second row: particle's x, y, z position at time t. Observe the bound increases as the source drifts from the focal plane.

frames, we evaluate our network tracking performance on 5 sequences of Brownian motion, each consisting of 1000 binned frames. Table 4.2 shows our event camera-specific optical designs minimize 3D tracking error more than conventional designs. Additionally, our method is substantially better at depth plane recovery. Qualitative results in Figure 4.6 demonstrate that 3D positions recovered using our designs more tightly fit ground-truth trajectories.

4.7 Ablation Studies

4.7.1 Optical Representations

Additionally, we compare 3D tracking results using two different amplitude mask representations: pixel-wise and neural amplitude mask (Figure 4.7) and three different phase mask representations: pixel-wise, Zernike basis, and neural phase mask (Figure 4.8). As shown in

	Component	RMSE (nm) ↓ 3D	L_1 (nm) ↓ z
	Open Aperture	617	936
Amplitude	Levin et al.	764	1036
	NAM (Ours)	66.0	49.2
Phase	Fisher	52.6	44.2
	NPM (Ours)	**51.2**	**39.2**

Table 4.2: **Tracking accuracy comparison.** We present quantitative results on 3D trajectory recovery for known optical designs. Our event CRB loss function found the best-performing design. Although only slightly improved in overall 3D tracking, our design noticeably improves depth recovery.

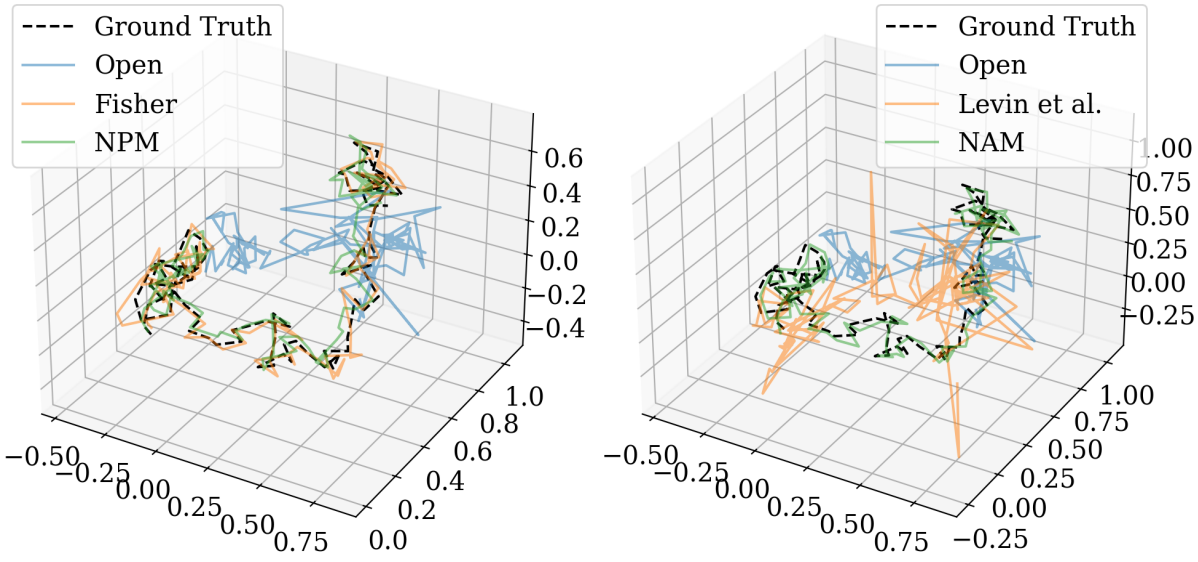


Figure 4.6: **Recovered 3D position over Brownian motion sequence with coded event frames.** Left: phase mask methods, right: amplitude mask methods. Observe trajectories reconstructed from phase mask-coded events more closely align with ground-truth positions. Units in microns.

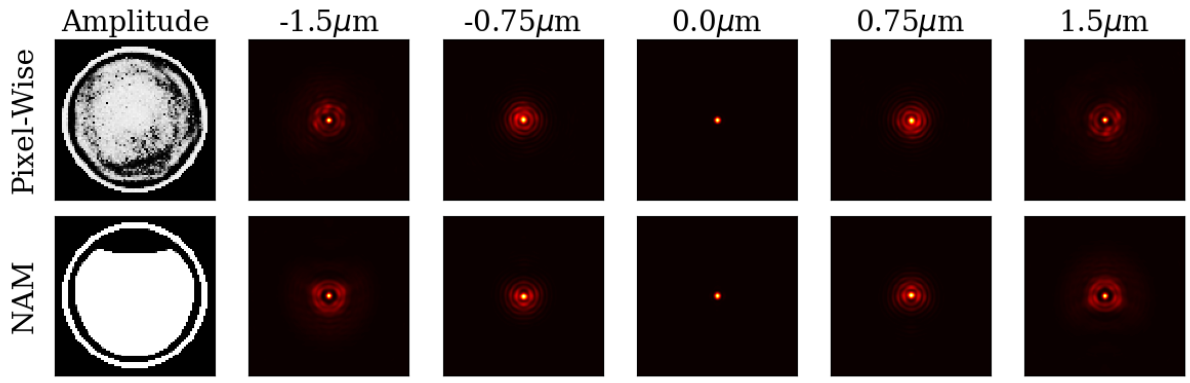


Figure 4.7: **Designed amplitude masks and corresponding PSFs.** Top: pixel-wise representation. Bottom: implicit neural representation.

	Representation	CRB (nm) ↓
Amplitude	Pixel-Wise	65.5
	NAM	50.5
Phase	Pixel-Wise	34.2
	Zernike	34.8
	NPM	**33.1**

Table 4.3: **Average CRB of different optimized representations across a 3μm depth range**. Notice the neural representations outperform their pixel-wise counterparts.

Table 4.3, our implicit neural representation-based methods achieve a lower average error bound than alternative representations, despite being two times smaller than pixel-wise representations with respect to the number of parameters. As expected, phase mask results generally outperform the amplitude mask results (Figure 4.9). However, our novel neural binary aperture makes optimizing amplitude masks more tractable. We observe that pixel-wise representations not only yield difficult-to-manufacture apertures but also suboptimal performance. In terms of 3D tracking, the implicit neural representations produce a smaller error on average (Table 4.4) and more accurately match sampled 3D trajectories (Figure 4.10).

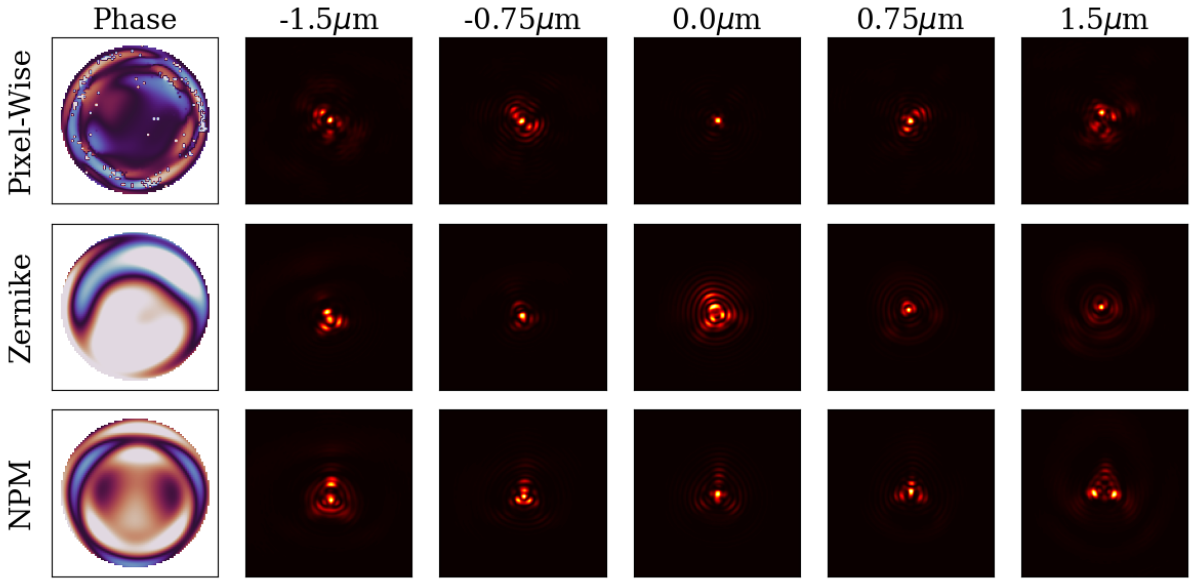


Figure 4.8: **Designed phase masks and corresponding PSFs.** Top: pixel-wise representation. Middle: first 55 Zernike coefficients representation. Bottom: implicit neural representation.

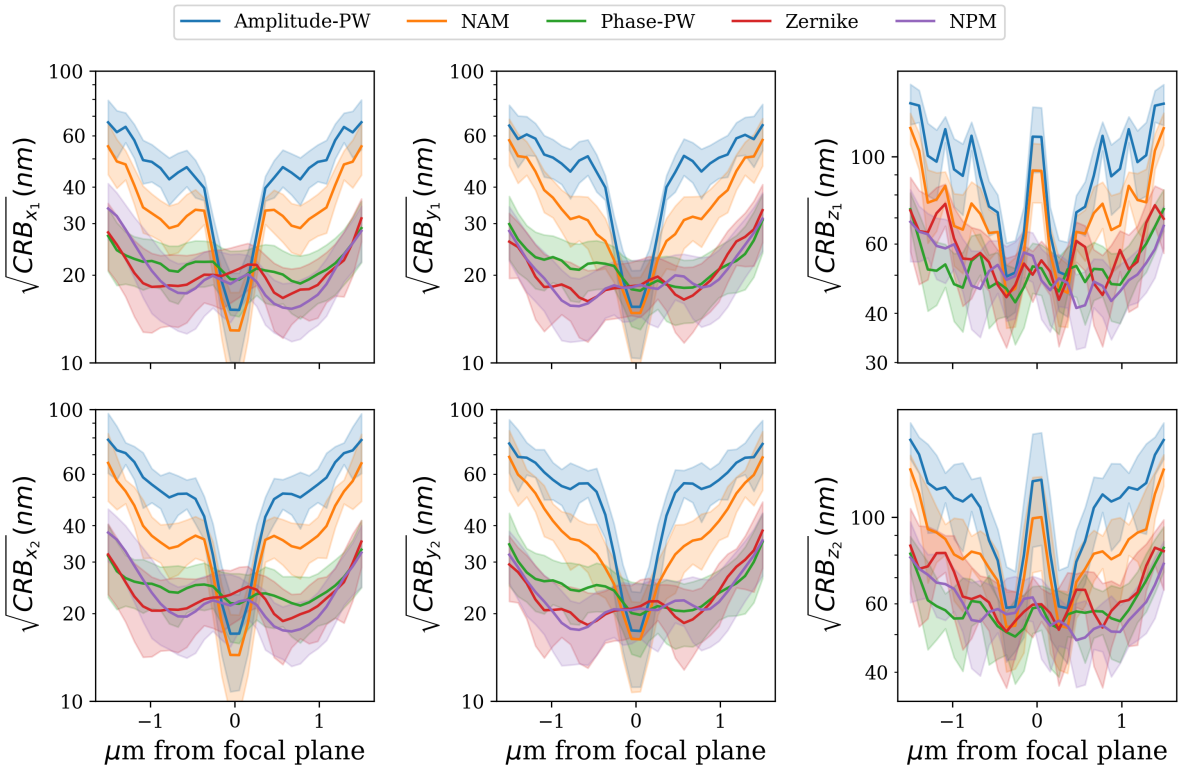


Figure 4.9: **Effect of optical parameterization on 3D localization CRB**. First row: particle's x, y, z position at time $t - \tau$. Second row: particle's x, y, z position at time t. Our implicit neural representations are particularly advantageous for amplitude masks.

	Component	RMSE (nm) ↓ 3D	L_1 (nm) ↓ z
Amplitude	Pixel-Wise	120	103
	NAM	66.0	49.2
Phase	Pixel-Wise	56.5	45.9
	Zernike	51.3	50.2
	Our NPM	**51.2**	**39.2**

Table 4.4: **Effect of optimized mask parameterization on tracking accuracy.** Average distance between ground-truth Brownian motion and the recovered 3D position is minimized with our neural-based designs.

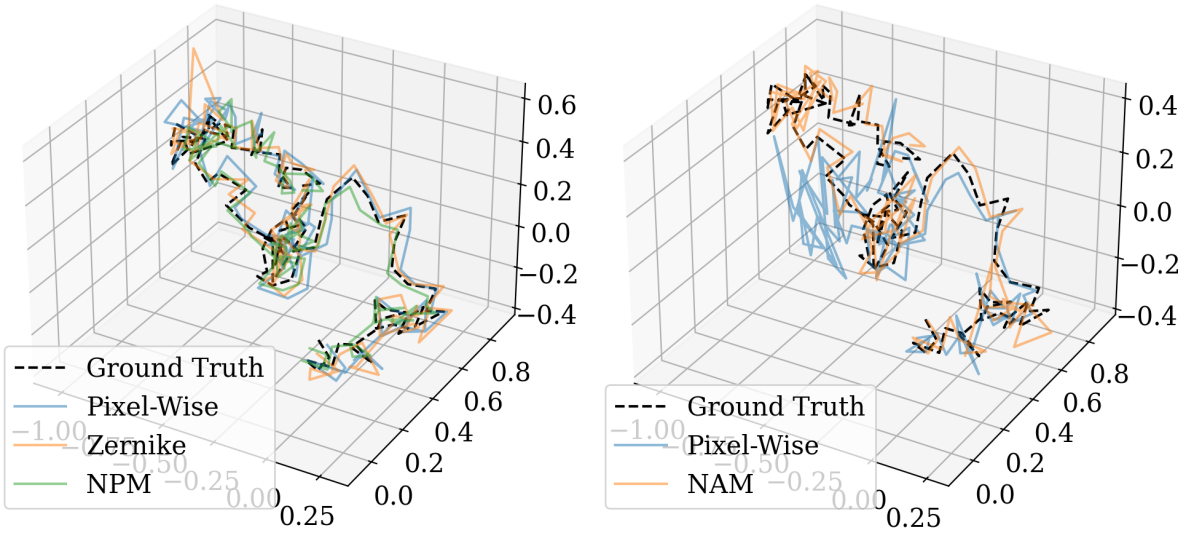


Figure 4.10: **Effect of optical representation on 3D trajectory recovery.** Left: phase mask methods, right: amplitude mask methods. Observe that neural representations produce tighter reconstructions. Units in microns.

4.7.2 Tracking Limits

In this section, we explore the limits of 3D tracking with variable external factors. For each experiment, we compute the average CRB over 30 depth slices and 6 parameters for 3 orthogonal unit directions (x, y, and z). First, as the number of available photons increases, the lower bound on 3D position estimation monotonically decreases (Figure 4.11). More available photons equate to a higher signal-to-noise ratio. Additionally, this result helps explain why phase masks outperform amplitude masks. Second, we show extremely slow-moving particles (less than nanometers per refresh rate) experience a significantly higher CRB (Figure 4.12). Minimal movement indicates smaller intensity changes and thus an event camera would trigger fewer events. On the other side, as a particle moves faster, the number of events will decrease as there is a non-zero delay between when an event camera can trigger sequential events. Our learned phase mask is more robust to speed changes than an open aperture and our learned amplitude mask. Third, when the percentage of photons due to background noise increases, the bound on error also increases (Figure 4.13). We design our masks with 1% of captured photons attributable to the background, but the learned designs are more resistant to degraded conditions than an open aperture.

4.7.3 Accumulation Time

Cutting-edge event cameras offer 10kHz fresh rates; even with 16-frame accumulation, the camera effectively operates at 625FPS — much faster than conventional CMOS sensors. We also retrained our CNN-based tracking algorithm on 'pure' event frames with no accumulation. Overall performance degraded: NPM by $+45\%$ RMSE and NAM by $+54\%$ RMSE. Alternative

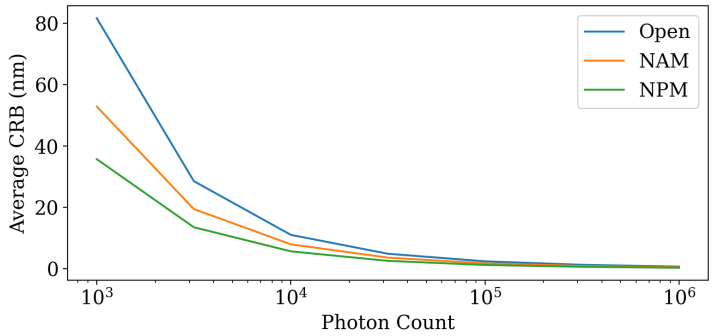


Figure 4.11: **Flux effect on CRB.** With more available photons, the signal-to-noise ratio increases, so the 3D information content is more reliable, and the bound on 3D tracking error decreases.

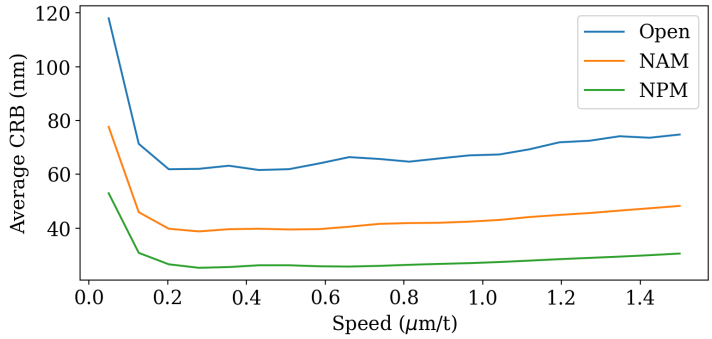


Figure 4.12: **Speed effect on CRB.** Too-slow moving particles trigger fewer events yielding a worse CRB. Similarly, as a particle moves faster the delay between triggers leads to fewer events.

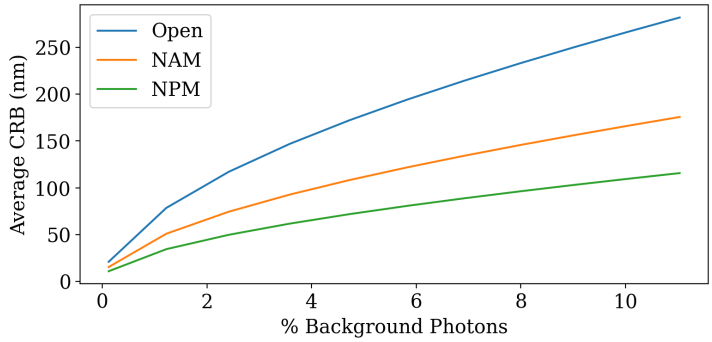


Figure 4.13: **Background photon effect on CRB.** As the percentage of photons hitting the sensor due to background noise increases, CRB also increases. The impact is minimal in our method.

architectures such as Spiking Neural Networks designed for sparse binary measurements may be better suited for processing 'pure' events.

4.7.4 Particle Speed

We have shown CRB depends on particle speed; a natural question is does the optimal design change with respect to speed. We optimize our neural phase mask using the CRB objective function with fixed particle speeds—$\{50, 100, 500, 1000\}$nm per time step. Our learned designs are shown in Figure 4.14. When a particle moves quickly relative to the binned interval, the optimal design resembles the Fisher phase pattern found for traditional CMOS sensors.

One can explain this collapse to the original Fisher mask design as follows. As a particle moves faster, the captured binned event frame looks more similar to the composition of a negative PSF at the start location and a positive PSF at the end location (Figure 4.15). This suggests that single-point event tracking mirrors two-point CMOS tracking.

4.8 Hardware Prototype

We performed a real-world experiment for tracking a point light source at meter scale using a binary amplitude mask and a Prophessee EVK3 event camera. Specifically, we fabricated the NAM mask at 20mm diameter scale on a Creality Ender 3 S1 Pro using 1.75mm PLA filament (see Figure 4.16). Then, we captured an event dataset by moving a point source at discrete depth planes ranging between 75cm and 125cm with and without our coded aperture. For all measurements, the camera was focused at 100cm. We binned events in 1ms intervals to achieve an effective frame rate of 1000 FPS and trained a CNN to estimate the event frame's depth.

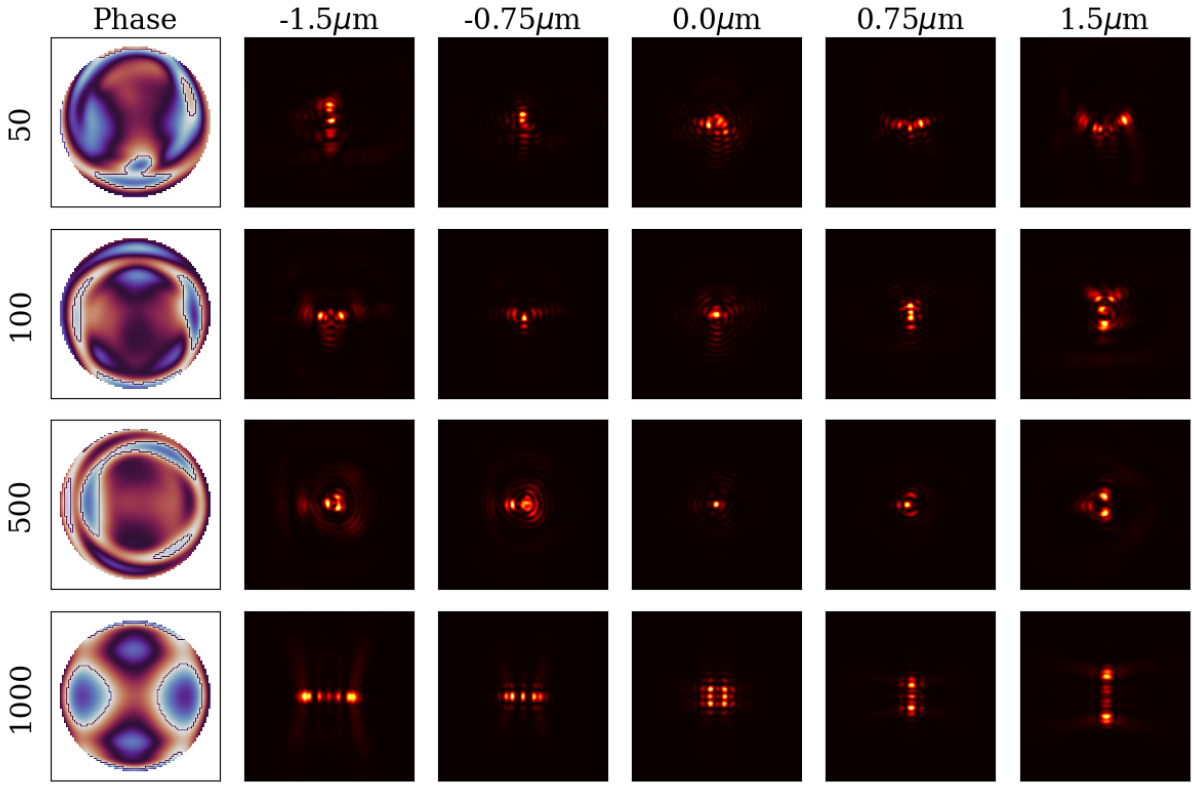


Figure 4.14: **Designed Phase Masks and corresponding PSFs for specific speeds.** Each row visualizes the neural phase mask designed for tracking particles moving at N nanometers per time interval. Observe that the optimal design for 'fast' moving particles is the Fisher design.

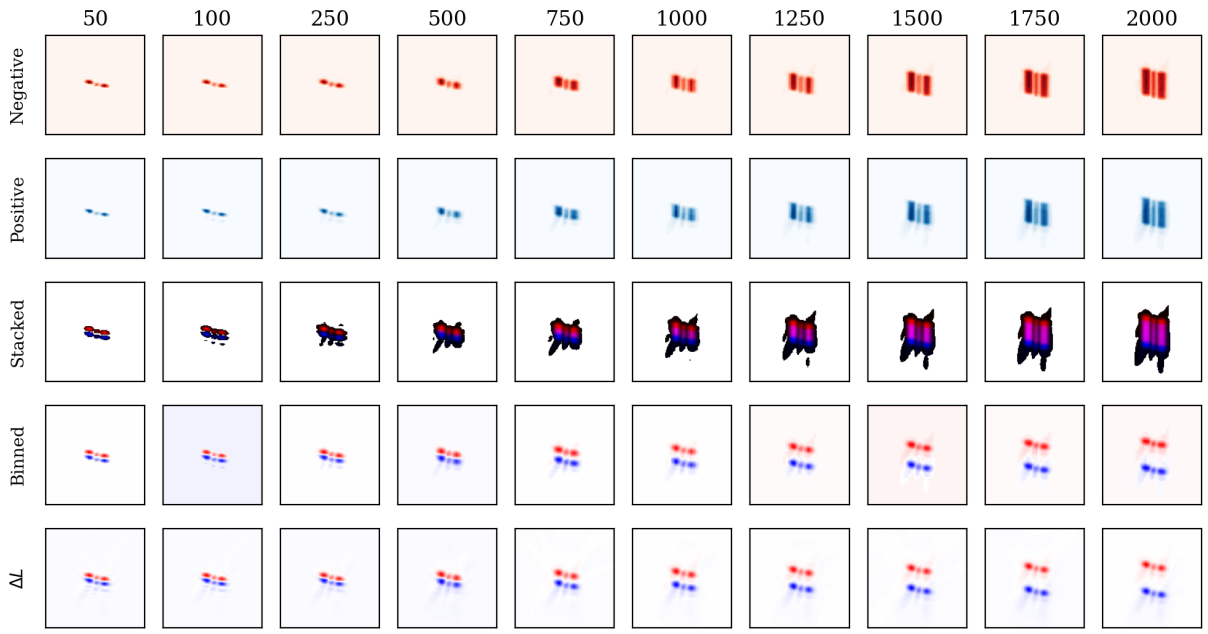


Figure 4.15: **Event camera measurements of a moving particle with the Fisher mask.** Motion is simulated over a fixed time interval with 100 event samples. Observe a 'fast' moving particle produces an event frame with two copies of a regular PSF: a negative copy at the start location, and a positive copy at the end location. *Row 1*: negative event count over the time interval. *Row 2*: positive event count over the time interval. *Row 3*: the red channel visualizes negative events and the blue channel visualizes positive events. The pink regions represent where the events cancel in a binned measurement. *Row 4*: binned event frame $pos - neg$. *Row 5*: log-intensity difference ΔL.

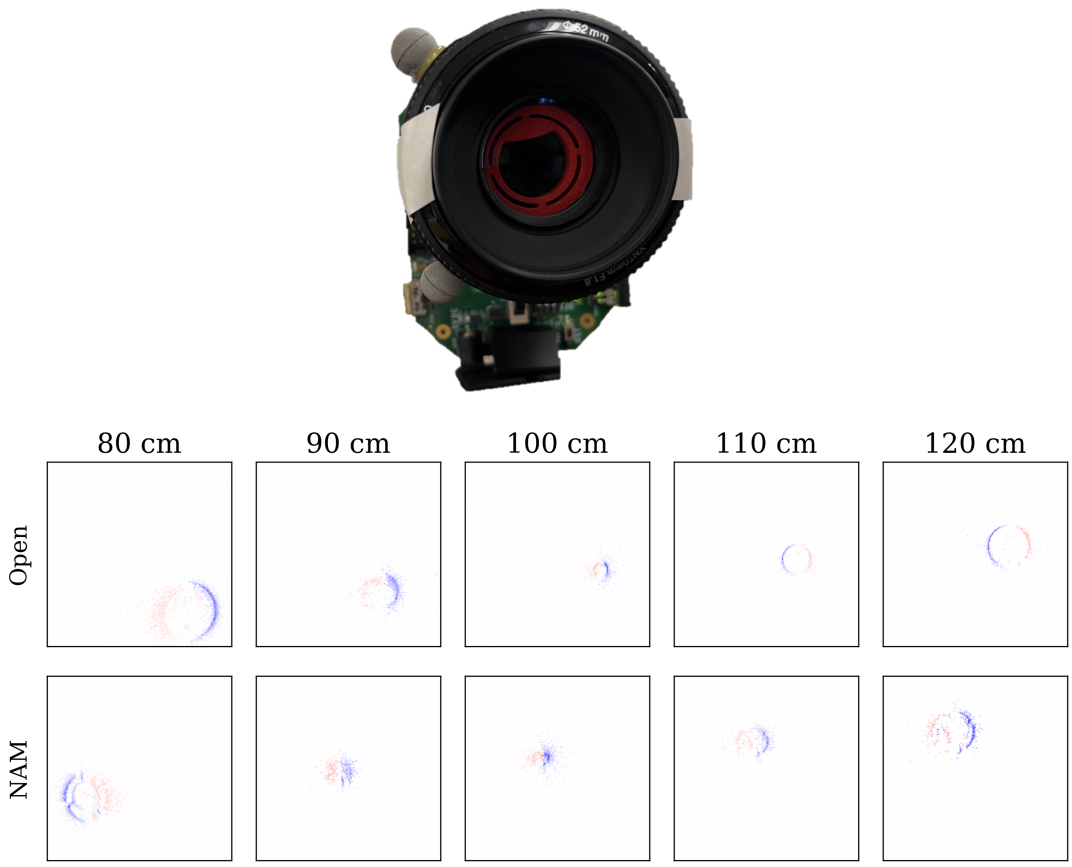


Figure 4.16: **Prototype.** Top: The fabricated mask is placed at the aperture plane of an event camera with a 50mm focal length lens. Bottom: Sample captured event frames for a point source.

Results in Figure 4.17 demonstrate improved tracking performance compared to an open aperture, particularly at depths where the point source is defocused.

4.9 Log-Intensity Difference Approximation

In this section, we prove the log-intensity difference approximation we consider when deriving the Cramér Rao Bound is proportional to binned event frames.

Assume an idealized event camera model, where an event is triggered as soon as the log-intensity change between the reference and the current intensity equals some threshold, $\mathcal{T}$. Consider producing a binned event frame for a time interval $[t_{\text{start}}, t_{\text{end}}]$. For a single pixel, let the se-

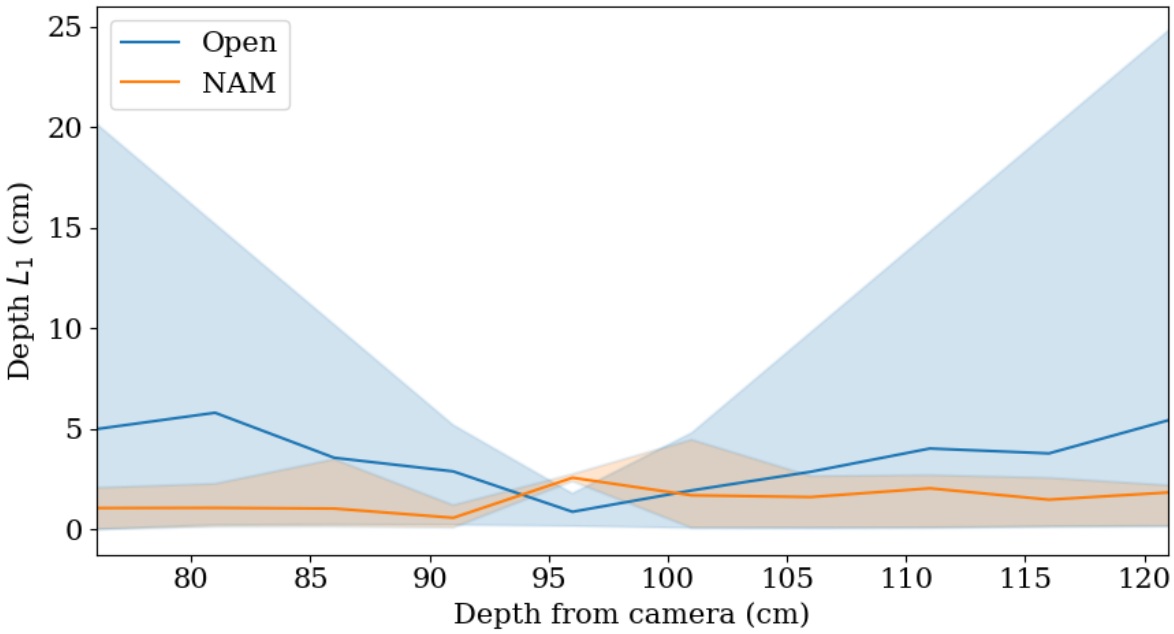


Figure 4.17: **Real-world 3D tracking.** Comparison between NAM and Open apertures for depth estimation at 1000FPS. Error bars show the 90% interquartile range.

quence of events over this interval occur at times $t_1, t_2, \ldots t_n$ and have polarities $p_1, p_2, \ldots, p_n \in \{-1, 1\}$. Let $f(t)$ be the log-intensity at time t for the same pixel and be continuous over the interval.

Lemma 4.9.1. *The log-intensity difference, $f(t_{end}) - f(t_{start})$, is proportional to the binned event pixel value, $\sum_{i=1}^{n} p_i$, with error $|\epsilon| < 1$.*

$$f(t_{end}) - f(t_{start}) \propto \epsilon + \sum_{i=1}^{n} p_i \tag{4.19}$$

Proof. By assumption, the magnitude of the change corresponding to each event is $\mathcal{T}$. Notice that $\mathcal{T} p_i$ is the log-intensity difference between the previous event time (the reference) and the current event time.

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \sum_{i=1}^{n} f(t_i) - f(t_{i-1}) \tag{4.20}$$

The right-hand side is a telescoping sum,

$$\sum_{i=1}^{n} f(t_i) - f(t_{i-1}) = f(t_n) - f(t_0). \tag{4.21}$$

$t_0 = t_{\text{start}}$ because the first event must occur $t_1 - t_0$ after the start of the interval. Then, the binned event frame is

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \left(f(t_n) - f(t_{\text{start}}) \right). \tag{4.22}$$

Finally, $|f(t_n) - f(t_{\text{end}})| = |\delta| < \mathcal{T}$ because if the quantity exceeded the threshold, an additional event would be triggered. Substitute t_{end} for t_n.

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \left(f(t_{\text{end}}) - f(t_{\text{start}}) + \delta \right) \tag{4.23}$$

$$= \frac{1}{\mathcal{T}} \left(f(t_{\text{end}}) - f(t_{\text{start}}) \right) + \epsilon \tag{4.24}$$

Thus, a binned event frame can be approximated as log-intensity difference divided by $\mathcal{T}$ with error $|\epsilon| < 1$. □

As an event camera becomes more sensitive to change ($\mathcal{T}$ decreases), the approximation's percent error decreases because the magnitude of the binned event frame increases but the total absolute error is fixed at most 1.

4.10 Limitations

While we were successful in designing optics to improve performance on 3D tracking with event cameras, our method carries some limitations. First, although our binned event frames can

be obtained at kHz refresh rates, they do not take full advantage of the asynchronous nature of event cameras. Second, our bounds are for an idealized event camera model with no read-noise. It would be impossible to outperform these bounds, but there might exist a tighter bound that accounts for these hardware imperfections. Lastly, we only consider single-emitter images. With multiple point sources, the resolving accuracy between single points may be more limited.

4.11 Conclusion

This work introduces PSF-engineering to neuromorphic event-based sensors. We first derive information theoretical limits on 3D point localization and tracking. We demonstrate that existing amplitude and phase mask designs are suboptimal for tracking moving emitters and design new optical elements for this task. Additionally, to overcome the non-convexity of this optimization problem, we introduce a novel implicit neural representation for optical components. Finally, we validate the effectiveness of our designs in simulation and compare against state-of-the-art mask designs. Our work unlocks not only highly performant optics for event cameras but also the ability to design highly expressive elements for other sensors.

Chapter 5: Concluding Remarks

5.1 Summary

While optics development has traditionally emphasized creating clear representations for human comprehension, optical systems can also be engineered to encode additional data consumption by computers. For instance, depth-dependent defocus patterns can be used to encode high-quality depth information through object blurring. In scientific imaging, such specialized optics enable precisely localizing molecules in three-dimensional space [22, 23]. However, these designs aren't suitable for next-generation sensors, given fundamental differences in underlying image acquisition process.

Next-generation ultrafast hardware such as SLMs and event cameras will allow us to observe phenomena beyond the reach of conventional designs. In this work, we prove the design space for point-spread-functions generated with dynamic SLM phase masks are *fundamentally more expressive* than their static counterparts unlocking an exciting new direction for PSF engineering. We demonstrated the ability to learn time-averaged dynamic PSFs (TiDy-PSFs) that can substantially improve computational imaging performance. Then, we derive the information theoretical bounds for 3D tracking with event cameras. We show existing designs for conventional cameras are already optimal for static flashing point sources, then demonstrate the designs are sub-optimal for moving constant point sources. Using these bounds, we design new optimal

phase masks for this task and demonstrate improved 3D tracking performance.

5.2 Future Work

In the context of event-cameras, recent developments in neuromorphic computing chips by Intel show promise in asynchronously processing raw event streams without event accumulation. Leveraging spiking neural networks with these new circuits could uncover even more efficient optical designs.

Additionally, nanofabrication techniques for designing sub-wavelength optical elements show promise for a brand new point-spread-function design space — one where we have strong wavelength and polarization dependence. We can derive new theory for the limits of imaging with such flat 'meta-optics' and develop new algorithms for processing those measurements.

Bibliography

[1] Nadia G Cervino, Agustín J Elias-Costa, Martín O Pereyra, and Julián Faivovich. A closer look at pupil diversity and evolution in frogs and toads. *Proceedings of the Royal Society B*, 288(1957):20211402, 2021.

[2] Lydia M Mäthger, Roger T Hanlon, Jonas Håkansson, and Dan-Eric Nilsson. The w-shaped pupil in cuttlefish (sepia officinalis): functions for improving horizontal vision. *Vision Research*, 83:19–24, 2013.

[3] Mandyam V Srinivasan. Honey bees as a model for vision, perception, and cognition. *Annual review of entomology*, 55:267–284, 2010.

[4] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70–es, 2007.

[5] Joseph W. Goodman. *Introduction to Fourier Optics*. Freeman, 2017.

[6] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d — learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2019.

[7] Hayato Ikoma, Cindy M. Nguyen, Christopher A. Metzler, Yifan Peng, and Gordon Wetzstein. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. *IEEE International Conference on Computational Photography (ICCP)*, 2021.

[8] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, Jun 2011.

[9] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[11] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[12] Jaime Spencer, Richard Bowden, and Simon Hadfield. Defeat-net: General monocular depth via simultaneous unsupervised representation learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[13] Zunzhi You, Yi-Hsuan Tsai, Wei-Chen Chiu, and Guanbin Li. Towards interpretable deep networks for monocular depth estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12879–12888, October 2021.

[14] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1623–1637, 2022.

[15] Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3d: Towards zero-shot metric 3d prediction from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9043–9053, October 2023.

[16] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[17] Daniel Gehrig Javier Hidalgo-Carrio and Davide Scaramuzza. Learning monocular dense depth from events. *IEEE International Conference on 3D Vision.(3DV)*, 2020.

[18] Mohammad Mostafavi, Kuk-Jin Yoon, and Jonghyun Choi. Event-intensity stereo: Estimating depth by the best of both worlds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4258–4267, October 2021.

[19] Yeongwoo Nam, Mohammad Mostafavi, Kuk-Jin Yoon, and Jonghyun Choi. Stereo depth from events cameras: Concentrate and focus on the future. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6114–6123, June 2022.

[20] Peilun Shi, Jiachuan Peng, Jianing Qiu, Xinwei Ju, Frank Po Wen Lo, and Benny Lo. Even: An event-based framework for monocular depth estimation at adverse night conditions, 2023.

[21] Jiyuan Zhang, Lulu Tang, Zhaofei Yu, Jiwen Lu, and Tiejun Huang. Spike transformer: Monocular depth estimation for spiking camera. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 34–52, Cham, 2022. Springer Nature Switzerland.

[22] Sri Rama Prasanna Pavani, Michael A. Thompson, Julie S. Biteen, Samuel J. Lord, Na Liu, Robert J. Twieg, Rafael Piestun, and W. E. Moerner. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*, 106(9):2995–2999, 2009.

[23] Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W. E. Moerner. Optimal point spread function design for 3d imaging. *Phys. Rev. Lett.*, 113:133902, September 2014.

[24] Alex Paul Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):523–531, 1987.

[25] Edward R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34(11):1859–1866, April 1995.

[26] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69–es, jul 2007.

[27] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.*, 37(4), July 2018.

[28] Xin Liu, Linpei Li, Xu Liu, Xiang Hao, and Yifan Peng. Investigating deep optics model representation in affecting resolved all-in-focus image quality and depth estimation fidelity. *Opt. Express*, 30(20):36973–36984, September 2022.

[29] Julie Chang and Gordon Wetzstein. Deep optics for monocular depth estimation and 3d object detection. In *Proc. IEEE ICCV*, 2019.

[30] Seung-Hwan Baek, Hayato Ikoma, Daniel S. Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H. Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2651–2660, October 2021.

[31] Daniel S. Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Transactions on Graphics (Proc. SIGGRAPH 2019)*, 38(4):117:1–13, 2019.

[32] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385, 2020.

[33] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 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*, pages 1386–1396, 2020.

[34] Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, and Gordon Wetzstein. Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. *ACM Trans. Graph.*, 40(6), December 2021.

[35] Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, Matthew O'Toole, and Gordon Wetzstein. Time-multiplexed neural holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. In *Proceedings of the ACM SIGGRAPH*, pages 1–9, 2022.

[36] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. Learning privacy-preserving optics for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2573–2582, October 2021.

[37] D. Chan, M. Sheinin, and M. O'Toole. Spincam: High-speed imaging via a rotating point-spread function. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10755–10765, Los Alamitos, CA, USA, oct 2023. IEEE Computer Society.

[38] Yoav Shechtman, Lucien Weiss, Adam Backer, Steffen Sahl, and William Moerner. Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions. *Nano letters*, 15, May 2015.

[39] Hayato Ikoma, Takamasa Kudo, Yifan Peng, Michael Broxton, and Gordon Wetzstein. Deep learning multi-shot 3d localization microscopy using hybrid optical–electronic computing. *Opt. Lett.*, 46(24):6023–6026, Dec 2021.

[40] Bhargav Ghanekar, Vishwanath Saragadam, Dushyant Mehra, Anna-Karin Gustavsson, Aswin C. Sankaranarayanan, and Ashok Veeraraghavan. Ps ^{2} f: Polarized spiral point spread function for single-shot 3d sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12, 2022.

[41] Zheng Shi, Yuval Bahat, Seung-Hwan Baek, Qiang Fu, Hadi Amata, Xiao Li, Praneeth Chakravarthula, Wolfgang Heidrich, and Felix Heide. Seeing through obstructions with diffractive cloaking. *SIGGRAPH*, 41(4), 2022.

[42] Brandon Y. Feng, Haiyun Guo, Mingyang Xie, Vivek Boominathan, Manoj K. Sharma, Ashok Veeraraghavan, and Christopher A. Metzler. Neuws: Neural wavefront shaping for guidestar-free imaging through static and dynamic scattering media. *Science Advances*, 9(26):eadg4671, 2023.

[43] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Weinberger. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[44] Feng Xue, Guirong Zhuo, Ziyuan Huang, Wufei Fu, Zhuoyue Wu, and Marcelo H Ang. Toward hierarchical self-supervised monocular absolute depth estimation for autonomous driving applications. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2330–2337. IEEE, 2020.

[45] Anupa Sabnis and Leena Vachhani. Single image based depth estimation for robotic applications. In *2011 IEEE Recent Advances in Intelligent Computational Systems*, pages 102–106, 2011.

[46] Menglong Ye, Edward Johns, Ankur Handa, Lin Zhang, Philip Pratt, and Guang-Zhong Yang. Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery, 2017.

[47] Robert Fischer, Yicong Wu, Pakorn Kanchanawong, Hari Shroff, and Clare Waterman-Storer. Microscopy in 3d: A biologist's toolbox. *Trends in cell biology*, 21:682–91, October 2011.

[48] Luca Palmieri, Gabriele Scrofani, Nicolò Incardona, Genaro Saavedra, Manuel Martínez-Corral, and Reinhard Koch. Robust depth estimation for light field microscopy. *Sensors*, 19(3), 2019.

[49] Woontack Woo, Wonwoo Lee, and Nohyoung Park. Depth-assisted real-time 3d object detection for augmented reality. In *International Conference on Artificial Reality and Telexistence (ICAT)*, 2011.

[50] Yawen Lu, Sophia Kourian, Carl Salvaggio, Chenliang Xu, and Guoyu Lu. Single image 3d vehicle pose estimation for augmented reality. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5, 2019.

[51] Sean J. Liu, Maneesh Agrawala, Stephen DiVerdi, and Aaron Hertzmann. ZoomShop: Depth-Aware Editing of Photographic Composition. *Computer Graphics Forum*, 2022.

[52] Hadi Alzayer, Zhihao Xia, Xuaner Zhang, Eli Shechtman, Jia-Bin Huang, and Michael Gharbi. Magic fixup: Streamlining photo editing by watching dynamic videos. In *arXiv*, pages –, 2024.

[53] Terry A. Bartlett, William C. McDonald, and James N. Hall. Adapting Texas Instruments DLP technology to demonstrate a phase spatial light modulator. In Michael R. Douglass, John Ehmke, and Benjamin L. Lee, editors, *Emerging Digital Micromirror Device Based Systems and Applications XI*, volume 10932, page 109320S. International Society for Optics and Photonics, SPIE, 2019.

[54] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2018.

[55] Nikolaus Mayer, Eddy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4040–4048, 2016.

[56] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[57] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.

[58] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[59] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth, 2023.

[60] Stephanie A Maynard, Philippe Rostaing, Natascha Schaefer, Olivier Gemin, Adrien Candat, Andréa Dumoulin, Carmen Villmann, Antoine Triller, and Christian G Specht. Identification of a stereotypic molecular arrangement of endogenous glycine receptors at spinal cord synapses. *eLife*, 10:e74441, dec 2021.

[61] Hippolyte Verdier, François Laurent, Alhassan Cassé, Christian L. Vestergaard, Christian G. Specht, and Jean-Baptiste Masson. A maximum mean discrepancy approach reveals subtle changes in α-synuclein dynamics. *bioRxiv*, 2022.

[62] Anika Kinkhabwala, Zongfu Yu, Shanhui Fan, Yuri Avlasevich, Klaus Müllen, and W. E. Moerner. Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photonics*, 3(11):654–657, 2009.

[63] Clément Cabriel, Tual Monfort, Christian G. Specht, and Ignacio Izeddin. Event-based vision sensor for fast and dense single-molecule localization microscopy. *Nature Photonics*, 2023.

[64] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020.

[65] Anastasios N. Angelopoulos, Julien N.P. Martel, Amit P.S. Kohli, Jorg Conradt, and Gordon Wetzstein. Event based, near-eye gaze tracking beyond 10,000 hz. *IEEE Transactions on Visualization and Computer Graphics (Proc. VR)*, 2021.

[66] Luc Tinch, Nitesh Menon, Keigo Hirakawa, and Scott McCloskey. Event-based detection, tracking, and recognition of unresolved moving objects. *Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference*, 2022.

[67] Jun Haeng Lee, Tobi Delbruck, Michael Pfeiffer, Paul K J Park, Chang-Woo Shin, Hyunsurk Eric Ryu, and Byung Chang Kang. Real-time gesture interface based on event-driven processing from stereo silicon retinas. *IEEE Trans Neural Netw Learn Syst*, 25(12):2250–2263, December 2014.

[68] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza,

Jeff Kusnitz, Michael Debole, Steve Esser, Tobi Delbruck, Myron Flickner, and Dharmendra Modha. A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[69] Craig Iaboni, Himanshu Patel, Deepan Lobo, Ji-Won Choi, and Pramod Abichandani. Event camera based real-time detection and tracking of indoor ground robots. *IEEE Access*, 9:166588–166602, 2021.

[70] Juan Pablo Rodríguez-Gómez, Raul Tapia, Maria del Mar Guzmán Garcia, Jose Ramiro Martínez-de Dios, and Anibal Ollero. Free as a bird: Event-based dynamic sense-and-avoid for ornithopter robot flight. *IEEE Robotics and Automation Letters*, 7(2):5413–5420, 2022.

[71] Alex Small and Shane Stahlheber. Fluorophore localization algorithms for super-resolution microscopy. *Nature Methods*, 11(3):267–279, 2014.

[72] A. Llavador, J. Sola-Pikabea, G. Saavedra, B. Javidi, and M. Martínez-Corral. Resolution improvements in integral microscopy with fourier plane recording. *Opt. Express*, 24(18):20792–20798, Sep 2016.

[73] Fanglin Linda Liu, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Fourier diffuserscope: single-shot 3d fourier light field microscopy with a diffuser. *Opt. Express*, 28(20):28969–28986, Sep 2020.

[74] Matthew B Bouchard, Brenda R Chen, Sean A Burgess, and Elizabeth M C Hillman. Ultra-fast multispectral optical imaging of cortical oxygenation, blood flow, and intracellular calcium dynamics. *Opt Express*, 17(18):15670–15678, Aug 2009.

[75] Ahmed S. Abdelfattah, Jihong Zheng, Amrita Singh, Yi-Chieh Huang, Daniel Reep, Getahun Tsegaye, Arthur Tsang, Benjamin J. Arthur, Monika Rehorova, Carl V. L. Olson, Yichun Shuai, Lixia Zhang, Tian-Ming Fu, Daniel E. Milkie, Maria V. Moya, Timothy D. Weber, Andrew L. Lemire, Christopher A. Baker, Natalie Falco, Qinsi Zheng, Jonathan B. Grimm, Mighten C. Yip, Deepika Walpita, Martin Chase, Luke Campagnola, Gabe J. Murphy, Allan M. Wong, Craig R. Forest, Jerome Mertz, Michael N. Economo, Glenn C. Turner, Minoru Koyama, Bei-Jung Lin, Eric Betzig, Ondrej Novak, Luke D. Lavis, Karel Svoboda, Wyatt Korff, Tsai-Wen Chen, Eric R. Schreiter, Jeremy P. Hasseman, and Ilya Kolb. Sensitivity optimization of a rhodopsin-based fluorescent voltage indicator. *Neuron*, 111(10):1547–1563.e9, 2023/11/12 2023.

[76] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V. Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014.

[77] Jianglai Wu, Yajie Liang, Shuo Chen, Ching-Lung Hsu, Mariya Chavarha, Stephen W Evans, Dongqing Shi, Michael Z Lin, Kevin K Tsia, and Na Ji. Kilohertz two-photon fluorescence microscopy imaging of neural activity in vivo. *Nat Methods*, 17(3):287–290, Mar 2020.

[78] Yayao Ma, Youngjae Lee, Catherine Best-Popescu, and Liang Gao. High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells. *Proceedings of the National Academy of Sciences*, 118(3):e2004176118, 2021.

[79] Sheng Xiao, John T. Giblin, David A. Boas, and Jerome Mertz. High-throughput deep tissue two-photon microscopy at kilohertz frame rates. *Optica*, 10(6):763–769, Jun 2023.

[80] Ruipeng Guo, Qianwan Yang, Andrew S. Chang, Guorong Hu, Joseph Greene, Christopher V. Gabel, Sixian You, and Lei Tian. Eventlfm: Event camera integrated fourier light field microscopy for ultrafast 3d imaging, 2023.

[81] Xavier Lagorce, Garrick Orchard, Francesco Galluppi, Bertram E Shi, and Ryad B Benosman. Hots: A hierarchy of event-based time-surfaces for pattern recognition. *IEEE Trans Pattern Anal Mach Intell*, 39(7):1346–1359, Jul 2017.

[82] Amos Sironi, Manuele Brambilla, Nicolas Bourdis, Xavier Lagorce, and Ryad Benosman. Hats: Histograms of averaged time surfaces for robust event-based object classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[83] Ignacio Alzugaray and Margarita Chli. Asynchronous corner detection and tracking for event cameras in real time. *IEEE Robotics and Automation Letters*, 3(4):3177–3184, 2018.

[84] R. Wes Baldwin, Mohammed Almatrafi, Jason R. Kaufman, Vijayan Asari, and Keigo Hirakawa. Inceptive event time-surfaces for object classification using neuromorphic cameras. In Fakhri Karray, Alfred Yu, and Aurélio Campilho, editors, *Image Analysis and Recognition - 16th International Conference, ICIAR 2019, Proceedings*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pages 395–403, Germany, 2019. Springer Verlag.

[85] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, 2004.

[86] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature Methods*, 3(10):793–796, 2006.

[87] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[88] Alexey Sharonov and Robin M. Hochstrasser. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences*, 103(50):18911–18916, 2006.

[89] Donald L. Snyder and Michael I. Miller. *Random Point Processes in time and space*. Springer, 2 edition, 1991.

[90] Steven M. Kay. *Fundamentals of Statistical Signal Processing*, volume 1. Prentice-Hall, 1 edition, 1993.

[91] Tralissa F Griffin. Distribution of the ratio of two poisson random variables. Master's thesis, Texas Tech University, 1992.

[92] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103, January 2017.

[93] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Proc. NeurIPS*, 2020.

[94] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded aperture pairs for depth from defocus. In *2009 IEEE 12th International Conference on Computer Vision*, pages 325–332, 2009.

[95] Changyin Zhou and Shree Nayar. What are good apertures for defocus deblurring? In *2009 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2009.

[96] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 807–814, Madison, WI, USA, 2010. Omnipress.

[97] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, 2015.

[98] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2016.

[99] Y Hu, S C Liu, and T Delbruck. v2e: From video frames to realistic DVS events. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2021.

[100] Nicole J Yang and Marlon J Hinner. Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods Mol Biol*, 1266:29–53, 2015.