MEMS Mirror-Based High-Speed Spatial Light Modulators



Cem Yalcin

Electrical Engineering and Computer Sciences University of California, Berkeley

Technical Report No. UCB/EECS-2025-18 http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-18.html

May 1, 2025

Copyright © 2025, by the author(s). All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

MEMS Mirror-Based High-Speed Spatial Light Modulators

By

Cem Yalcin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Rikky Muller, Chair Professor Ming C. Wu Professor Laura Waller Professor Nicolas Pégard

Spring 2023

MEMS Mirror-Based High-Speed Spatial Light Modulators

Copyright 2023 by Cem Yalcin

Abstract

MEMS Mirror-Based High-Speed Spatial Light Modulators

by

Cem Yalcin

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Rikky Muller, Chair

In this work, optical, computational, and electronic considerations for high-speed MEMS micromirror-based spatial light modulators (SLMs) are presented with a focus on holographic systems that utilize these elements for point cloud generation.

High-speed 3-D holographic patterning of light into point clouds is a prominent technique in a variety of applications ranging from AR/VR displays and holographic projectors to 3-D printing and biology. All-optical neural interfaces utilizing this technique stand to provide a minimally-invasive pathway to circumvent the fundamental limitations of electrical interfaces. However, the refresh rate and temporal capabilities of such holographic systems are heavily bottlenecked by both the computationally intensive computer-generated holography (CGH) algorithms used to compute the hologram and the slow-settling phase-modulating elements in the SLMs used to project the hologram.

I first present a computationally light CGH algorithm for point cloud patterning that closely matches the performance of state-of-the-art algorithms at 2-6 orders of magnitude faster computation times. Its non-iterative and memory-light architecture allows for CPUbased computation in ms timescales. Fast computation easily lends itself to time-multiplexingbased approaches for target throughput increase and speckle reduction. Experimental verification confirms the simulation results across SLM formats, target counts, and refresh rates.

I then present the analysis, design, and verification of two generations of MEMS mirrorbased SLMs. Firstly, a reduced-degree-of-freedom SLM built from high-speed piston-motion micromirrors is discussed. This device consists of an annular array comprising >23000 micromirrors arranged into 32 concentric rings, and a custom-designed driver ASIC capable of correcting for the global process variations of the MEMS fabrication. The array was used in random-access varifocal operation, demonstrating optical functionality. A secondgeneration family of piston-motion micromirror-based SLMs is presented next, with a pathway to achieving high-degree-of-freedom SLM operation with array sizes of up to 64x64 individually-addressable mirrors. The integration scheme and actuation voltage requirements for these devices necessitated the design of a second-generation ASIC, which is also presented. To my family: my brother Cihan & my parents, Berrin & Fatih

Contents

Co	onter	ats	ii
Lis	st of	Figures	iv
Lis	st of	Tables	vi
1	Intr 1.1 1.2 1.3 1.4 1.5	oduction Optical Neural Interfaces SLM-based Optical Neural Interfaces SLM Technologies Thesis Contribution Thesis Organization	1 2 6 8 10
2	A L 2.1 2.2 2.3 2.4 2.5 2.6	ightweight Algorithm for 3D Point-Cloud HolographyIntroductionDescription of NIMBLE-PATCH Algorithm and Evaluation FrameworkSingle-frame Algorithm Performance ComparisonExperimental Evaluation of Single-Frame Algorithm PerformanceTime-multiplexing Performance EvaluationDiscussion and Conclusions	 13 13 16 22 25 30 34
3	A N Inte 3.1 3.2 3.3 3.4 3.5	IEMS-based Optical Scanning System for Precise, High-speed Neural orfacing Introduction Introduction MEMS Mirror Actuation Requirements Driver ASIC Implementation Measurement Results Summary and Discussion	36 36 37 41 45 48
4	A 3 4.1 4.2 4.3	2×32, 29-V Driver ASIC for High-Parasitic Micromirror Arrays Introduction	51 51 51 57

	4.4Measurement Results	•	•	$\begin{array}{c} 61 \\ 65 \end{array}$
5	Conclusion			67
Bi	bliography			69

iii

List of Figures

1.1 1.2 1.3 1.4	Simplified diagrams of light delivery systems for all-optical neural interfaces Simplified diagram of an SLM-based 2f holographic point-cloud projection system. Conceptual diagram of an all-optical full-duplex interface to the neural tissue An overview of some alternative technologies for fast spatial light modulation	4 6 7 9
2.1	Illustration of regional variations in phase mask efficiency under 3D point steering	18
2.2	SLM partitioning principle underlying NIMBLE-PATCH algorithm	20
2.3	Efficiency roll-off profiles along the X axis and Z axis	21
2.4	Sweep results comparing NIMBLE-PATCH to GS for single-frame simulations .	24
2.5	Sweep results comparing NIMBLE-PATCH to GS for single-frame simulations in	
	terms of the two auxiliary metrics	25
2.6	Experimental setup and lateral and axial projections of measured intensity dis-	
	tributions for NIMBLE-PATCH, GSx1 and GSx	27
2.7	Simulated and measured FWHM spot sizes for NIMBLE-PATCH, GSx1 and GSx3	28
2.8	Intended target depth plotted against recorded peak intensity depth from on-axis	
	focus tuning calibration for experimental acquisitions	29
2.9	Percentage and position error of successfully identified spots in simulation and	
0.10	experiment.	31
2.10	Diagrams illustrating the decomposition principle employed in NIMBLE-PATCH	00
0.11	and GS	32
2.11	Sweep results comparing NIMBLE-PATCH to decomposed GS for excess refresh	പ
	rate scaling evaluation	33
3.1	Structural diagram and operating principles of the micromirror-based varifocal	
	element	38
3.2	Diagram illustrating the effects of resolution in phase on the hologram quality	
	visually and in terms of the two metrics α and η	39
3.3	Voltage vs. displacement curve for a simulated MEMS micromirror	40
3.4	Simplified block diagram of the mirror driver ASIC.	41
3.5	The schematic and the operation of the nonlinear DAC	42
3.6	Principle of operation of the DRAM write chain with 4 phases of configuration	
	shown	44

3.7	Chip micrograph with the inset showing MEMS pad openings, chip specifications and power breakdown.	45
3.8	Measured precision and maximum residual error of the nonlinear DAC vs. DAC	
	code	46
3.9	Static and dynamic measurements of the ASIC-MEMS system performed under a DHM	47
3.10	Optical measurement setup and results for the tunable lens system formed by the	
	ASIC and 32-channel MEMS array	49
4.1	The structure, static response, and dynamic response of the second-generation	
	mirror structures	52
4.2	Simulated voltage vs. displacement curve for the two variants of the second- generation micromirror structures, with annotated maximum voltage require-	
	ments and smallest step sizes	54
4.3	Simulated two-step drive profiles of the 38% extended structures	54
4.4	MEMS die lavouts of two micromirror array variants	56
4.5	Simplified block diagram of the second-generation ASIC.	58
4.6	The schematic and simulation results of the pixel circuitry	59
4.7	A microphotograph of the second-generation ASIC, the summary of the chip	
	specifications, and the power breakdown of major power consumers on the ASIC.	61
4.8	Measurement results from one of the ASIC pixels driving a square wave between	
	two V_{out} values (~6.25 V and ~24.75 V)	62
4.9	Measurement results for a C_{load} of 100 pF $\ldots \ldots \ldots$	63
4.10	Measured transduction, INL, and DNL curves of a pixel output versus digital	
	input code	64
4.11	Response of a MATLAB-based model of the 38% extended structure to the mea-	
	sured output of the pixel	65

v

List of Tables

3.1 Varifocal Element Comparison Table .		50
--	--	----

Acknowledgments

First and foremost, I extend my deepest thanks to my research advisor Prof. Rikky Muller. Her trust, support, and guidance allowed for an incredibly fruitful PhD, personally and professionally. The cheerful and safe culture of her lab fostered many close friendships, and her thorough and critical review of every piece of research made me into a much more rigorous engineer. I never felt lost or lacked guidance through my PhD, which is a privilege not every student has. I thank you from the bottom of my heart.

I would like to thank the rest of my dissertation committee as well. Prof. Laura Waller and Prof. Ming Wu stepped in whenever we needed cross-disciplinary input and lent us their valuable lab space and even more valuable researchers' time. It would not have been possible to do nearly as much as we did without their collaboration. Prof. Nicolas Pégard has been involved in this project since its conception and provided so much knowledge and guidance to me that when I look back at our meeting notes, I realize that we're still implementing his suggestions from 6 years ago. I would also like to acknowledge Prof. Daniel Lopez for his contributions to the project. I also want to extend my gratitude to Prof. Nick Antipa for his guidance in building the optical setup for our first prototype, it was a unique experience seeing our devices in action for the first time and wouldn't have been possible without his help.

During my PhD, I was blessed with the best possible project partner in the world, Nathan Tessema Ersaro, who promptly became my best friend. I dedicate this paragraph to him. Needless to say, there is no chance I could have done a fraction of what we did without you. A true friend who saw me through some of the hardest times of my life, I cannot thank you enough.

I also feel extremely fortunate to have met many exceptional people during my time at Berkeley. Burak Eminoğlu is one of the kindest, most humble, and technically knowledgeable people I have ever met. I am grateful to Burak for his continuous support and friendship, invaluable technical input, and nuanced opinions on every subject we talk about. I am also very grateful to Effhymios Papageorgiou, truly a complete engineer and a friend with whom I enjoy all technical and non-technical debates over weekend brunches. I also want to thank M. Meraj Ghanbari, who was always there when I needed help with anything, from IC design and 3-D modeling to plot coloring and Iranian food recipes. I am also very grateful to Sina Faraji, I don't think there is a single good IC idea I came up with that wasn't preceded by his technical input. He is also one of the best people to co-GSI a class with. I am really grateful to have Liz Murray as the successor for this project, I hope you don't mind the messy business of continuing our work. I want to thank Aviral Pandey, he is not only one of the best IC designers I know but also an immensely ethical and conscientious person. I would like to extend my thanks to my other fellow graduate students and postdocs – Ryan Kaveh, George Bocchetti, Justin Doong, Adelson Chua, Mauricio Bustamante, Jade Pinkenburg, Maarten Schelles, Chun-Yen Yao, Changuk Lee, Leyla Kabuli.

I am grateful to Shirley Salanio and the staff at BWRC – Candy Corpus, Mikaela Cavizo-Briggs, Jeffrey Anderson-Lee, James Dunn and Brian Richards – for their incredible support over the course of my PhD.

The research presented in this dissertation was financially supported by Chan-Zuckerberg Biohub, McKnight Technological Innovations in Neuroscience Award, NIH, Apple PhD Fellowship, UC Berkeley EECS department fellowship, the sponsors of BWRC, and TSMC. I am grateful to all these funding sources and agencies for supporting this research.

Before joining Berkeley, I had the privilege to work with Dr. Selim Eminoğlu, one of the most complete engineers I have ever met in my life. I strive to adopt his work ethic and relentless pursuit of technical progression, never becoming complacent. It is an understatement to say that he enabled me at the very beginning of my career, and I have no doubt I would not be writing this dissertation without his initial guidance. I would also like to thank Prof. Fatih Koçer, Prof. Tufan Karalar, and Prof. Tayfun Akın for encouraging me to pursue a PhD.

I'm grateful to my lifelong friends Suhip Tuncer Soyer and Sertan Demirdağ. Your friendship shaped me to become the person that I am.

I am most grateful to my wonderful family, who have sacrificed so much for me to have a life of my choosing, for their unconditional love and endless support. I dedicate this dissertation to them. Mom, your endless energy and unmatched empathy are unreachable standards that I aspire to live up to. You make it seem so easy to be an amazing parent and a hard-working professional at the same time. Dad, you've instilled the love and curiosity of science in me ever since my earliest memories. You taught me to stand by what is right and never to succeed at others' expense. Both of you are true inspirations to me for the person I want to be. My brother, Cihan, you are the most intelligent and tasteful person I know. I don't think it is a coincidence we share so many things we enjoy, I owe a huge chunk of my personality to you. You shaped my hobbies, sense of humor, interest in life, and analytical thinking skills, and even nudged me toward EE from CS, a nudge that I never questioned twice throughout my professional career.

Millie, you really set impossible standards for how empathetic, sweet, and thoughtful a person can be. You taught me so much about how to pronounce English words, even the ones that make no sense. I am really grateful to you, and I look up to you a lot.

Chapter 1 Introduction

Holography is an optical technique used to faithfully recreate 3-D sceneries from a wide range of viewing angles or even sculpt light into arbitrary 3-D shapes. In its most basic form, holography is performed by illuminating an object with a coherent light source and interfering the reflection with a reference beam. The resulting interference pattern, called the hologram, is recorded, and the original wavefront can then be recreated by illuminating a physical realization of the hologram with the reference beam. Reconstructing the light field instead of the image of an object allows for the capability to view the recorded object from a range of angles (limited by the recording and projection capabilities), making holography widely accepted as the ultimate method to reproduce 3-D sceneries [1]. Most modern realizations of this technique utilize digital holography, sampling the interference pattern on a digital imager such as a CMOS or CCD camera, which is widely used across many applications, including digital holographic microscopes (DHMs), optical security, bio-medicine, and augmented and virtual reality [2, 3, 4].

The advent of computing capabilities led to the development of computer-generated holography (CGH) techniques, which allow the computation of holograms that correspond to approximations of any specified 3-D distribution of light intensity instead of recording an existing object or scenery. CGH algorithms aim to converge to a hologram that sufficiently corresponds to the desired light intensity distribution as described by the user. While it is possible to display these holograms through static physical media such as silver halide or photopolymer elements [5], most applications require configurable holographic surfaces to adjust the hologram that is being projected dynamically. The core tools powering such CGH systems are dynamically configurable arrayed surfaces, known as spatial light modulators (SLMs), that impart pixel-level phase/amplitude modulation as defined by the hologram to incident beams in order to produce desired illumination patterns via downstream interference. SLMs are diffractive by nature; therefore, the discrete sampling grid formed by the finite set of analog-drive (and typically phase-only) modulation pixels determines the finite space-bandwidth product (SBP) available for modulation (i.e., higher pixel count in the SLM corresponds to a higher available SBP). Furthermore, the finite settling time of SLM unit elements usually limits the temporal bandwidth of the optical system. The finite SBP and temporal bandwidth of SLMs form a set of non-idealities that limit the quality of the generated 3-D light intensity distribution.

A very useful sub-class of CGH is the dynamic patterning of 3-D point cloud distributions, which entails the generation and precise placement of spots across the volume of interest. In the context of biological microscopy, 3-D point cloud patterning is employed for non-invasive all-optical interfacing with cell ensembles such as neuronal populations via bioengineered optical read/write probes [6, 7, 8, 9]. In augmented and virtual reality (AR/VR), near-eye display systems can incorporate virtual 3-D point cloud-based objects into real-world scenes [10, 11, 12, 13]. In the realm of material processing, point cloud patterning can be mobilized for 3-D nanofabrication via multiphoton or ultraviolet lithography [14, 15, 16, 17]. More broadly, any arbitrary 3-D light intensity distribution can generally be decomposed and treated under a point cloud basis [1, 18, 19, 20]. For the rest of this dissertation, mainly the SLM-facing requirements of 3-D point cloud patterning CGH systems will be considered, with an emphasis on the optical neural interface application.

1.1 Optical Neural Interfaces

One of the hardest reverse engineering problems in existence is the pursuit of understanding how cognitive and behavioral features of humans arise from the relatively understood individual operation of neurons. Operation of massive networks of "simple" unit elements can quickly become cryptic, as seen by the pursuit to understand how human-made neural networks work. Specifically, decoding the operation of a neural network is an extremely challenging problem even with the knowledge of the full connectivity profile and states of all unit elements (infinite spatial resolution of interrogation), the ability to introduce changes at every element at every step of computation (infinite spatiotemporal resolution of manipulation), and observe what happens after a single step of computation (infinite temporal resolution of interrogation). Present-day neuroscientific tools come nowhere near full observability or controllability, in time or space. Furthermore, an ideal neural interface would not only have complete access to the volume in terms of spatiotemporal resolution but also be noninvasive, as one of the greatest benefits of our increased understanding of the human brain is to be able to treat as-yet untreatable conditions ranging from blindness (e.g., degeneration of retinal tissue or optical nerve) to partial or complete paralysis (e.g., trauma to the sensorimotor pathway from the central nervous system to the peripheral nervous system).

Electrical neural interfaces have been used to study the brain for close to a century and are still the most common way to study neural behavior [21]. Modern embodiments of electrical interfaces typically utilize electrode or microelectrode arrays capable of reading out electroencephalography (EEG), intracranial electroencephalography (iEEG), or electrocorticography (ECoG) signals, and are used to observe individual neuron activity via action potential monitoring as well as aggregate activity of groups of neurons such as local field potentials [22]. The same electrodes can be used to stimulate neurons through charge delivery into the neural tissue for bidirectional interfacing [23]. Electrical interfaces face extremely challenging trade-offs, however, as in order to achieve cellular-level spatial resolution, electrodes themselves need to be scaled to the size and number of neurons of interest and invasively implanted into the neural tissue. In order to achieve the required temporal resolution to discern action potentials, the readout front-end needs to be able to record neural signals at >kHz bandwidth across all of the electrodes while meeting electrical requirements of the interface such as noise, linearity, and input impedance [24]. While the specifications of these systems are not impossible, every step of the way is riddled with fundamental trade-offs such as the choice of electrode size to balance spatial resolution, prevention of scar tissue formation, and safe charge delivery for stimulation [25, 26]. While it has been theorized that it could be possible to record electrical activity from every neuron in the cortex simultaneously, the effects of delivering stimulus current through a finite microelectrode area limit the stimulation capability of electrical systems [27]. Optical interfaces can provide a scalable modality by circumventing these trade-offs and providing natural solutions to some of the practical challenges.

In optical neural interfaces, the recording of neural activity is most commonly achieved through fluorescence imaging of key markers (indicators) that correlate with cell activity. The most common usage of this technique is in calcium imaging, utilizing genetically encoded calcium indicators (GECIs) to express Ca^{+2} -selective fluorescent markers in neurons, that modulate fluorescent light emission in response to intracellular calcium concentration. While GECIs themselves bottleneck imaging throughput due to their long time constants (10s to 100s of ms), the development of faster imaging agents is an active research area [28]. For instance, emerging genetically encoded voltage indicators (GEVIs) can encode single cell potentials down to mV levels into fluorescence signals, with response times of 100s of μ s to 10s of ms [28, 29]. For optical stimulation, neurons can be virally or genetically modified (through the technique called optogenetics) to express light-sensitive proteins (opsins) that excite or inhibit neural activity in response to light at specific wavelengths [6, 8]. State-of-the-art opsins reliably respond with exposure times on the order of a few ms and with sub-ms jitter performance [30].

The capabilities of an optical neural interface are heavily determined by the optical system and the method of light delivery. While there exists no formal classification for optical interface types, Figure 1.1 shows the three main approaches to light delivery into neural tissue, namely (1) direct delivery of broad static illumination, such as an LED or optical fiber delivering light to a population of neurons, (2) scanning methods, in which a single spot of light (either diffraction-limited or matched to the dimensions of the neuron's soma) is sequentially placed onto target neurons, and (3) holography, in which the stimulation or fluorescence excitation pattern is sculpted into a hologram to simultaneously target multiple neurons of interest. While (1) allows for a simple optical system, lack of precise spatiotemporal control over illumination limits the use case of these systems to bulk optogenetics applications in which genetically identical populations of neurons that express the optogenetics encoders are always stimulated simultaneously as a unique ensemble. For imaging applications, broad illumination entirely places the burden of reconstructing the 3-D scene on the imaging system, either through a scanner located in the imaging path or through



Figure 1.1: Simplified diagrams of light delivery systems for all-optical neural interfaces. (a) Direct illumination systems with no scanning elements provide non-specific illumination. (b) Scanned systems where lateral (XY) and varifocal (Z) elements provide 3-D positioning of a spot of light to perform sequential light delivery to individual cells. (c) Holographic systems where a spatial light modulator is configured to project light in parallel to multiple neurons, with single-cell precision.

computational imaging methods wherein the 3-D scene is reconstructed from a single 2-D image at the cost of higher computation resources [31].

On the other hand, scanning and holographic light delivery systems allow for the arbitrary placement of cell-sized spots of illumination in a millimeter-scale field-of-view (FoV). Scanbased systems most commonly utilize resonant lateral scanners (XY scanning) and varifocal elements (Z scanning) to perform a raster scan of the volume, in which the diffraction-limited spot is positioned through precise timing of the laser firing. Such systems are capable of scanning a wide FoV (mm-scale) while preserving neuron-sized (~10 µm) spot sizes. However, the finite dwell time required to excite the florescent markers or stimulate the neurons through optogenetic agents imposes a constraint on the throughput of these systems, as each cell needs to be sequentially addressed [7]. Furthermore, resonant scanners prevent random access of cells of interest, forcing a full scan of the volume even when regions/cells of interest may not occupy the full FoV. Despite these limitations and the added optical complexity, scan-based systems are widely adopted as they are a large step in spatial resolution with minimal computational complexity over flood-illuminated systems.

Holographic systems aim to resolve the spatiotemporal limitations of scan-based systems through patterning point-cloud holograms that target multiple cells simultaneously. The most common approach to building a holographic neural interface is through the use of SLMs to generate arbitrary 3-D light intensity distribution profiles. Holographic patterning efforts in neuroscience have demonstrated the ability to target up to 750 neurons with a single frame of a 0.5-megapixel SLM [8]. Several studies have also employed computer-generated holography patterning jointly with raster scanning for expanded system capabilities, including an extended field of view [32] and spiral beam tracing across neuron soma for stronger photostimulation [6]. The capability of generating multiple, cell-sized spots across a large FoV allows not only precise activity modulation in individual neurons amongst thousands of neurons, but also selective interrogation and fluorescence excitation of different locations in the volume, enabling time-multiplexed readout of the volume, and greatly simplifying scene reconstruction to the point where a single photodetector can serve as the imaging element [33].

A typical example of single neuron targeting in the cerebral cortex involves target sizes of down to 10 µm, within a FoV of 1mm x 1mm (lateral) x 300 µm (axial), using wavelengths that range from 450nm to 1500nm. For the optical system to not be a significant bottleneck to the overall throughput of the system, its components must have refresh rates of at least several kHz (10%-90% settling times of 100s of µs), as the settling time of optical elements is added to the exposure time of opsins and GEVIs to determine the overall throughput of the system. For scanned systems, a high optical system refresh rate directly translates to higher throughput as targets have to be addressed sequentially [7]. For holographic systems speckle noise, which is a high spatial frequency artifact usually encountered in coherent systems, can also be reduced through the utilization of high refresh rates. Time averaging of multiple holograms suppresses speckle noise, improving the accuracy of the resulting light distribution as the refresh rate increases beyond the regime in which opsins operate [34, 35, 36, 37]. Furthermore, as will be discussed later in this dissertation, complex holograms composed of hundreds of targets can be broken down into simpler, higher-contrast holograms and timemultiplexed at sufficiently high speeds to generate an effective 3-D intensity distribution when averaged over the timescale of the application.

1.2 SLM-based Optical Neural Interfaces

Operating Principles

SLM-based optical neural interfaces are most commonly constructed as 2f holographic projection systems where the SLM is located on the front focal plane (FFP) of a Fourier transforming lens or objective, and the target volume is centered around the back focal plane (BFP). A point-cloud hologram is projected into the target volume, such that bright spots of light are concentrated onto cells of interest for imaging or stimulation. Figure 1.2 depicts a simplified diagram of such a system, with x', y' denoting the spatial coordinates in the input domain of the system, and x, y, z denoting the spatial coordinates in the target domain. Such a system utilizes the 2-D Fourier Transform (2DFT) relationship between the two focal planes, where the 2DFT of the complex amplitude distribution at the BFP is projected onto the FFP in coherent systems such that [38]

$$U(x,y)|_{z=0} = \frac{1}{j\lambda f} \iint_{-\infty}^{+\infty} U'(x',y') e^{-j\frac{2\pi}{\lambda f}(x'x+y'y)} dx' dy'$$
(1.1)

where U'(x', y') is the complex amplitude profile of the reflecting wave at the FFP after modulation by the SLM, and $U(x, y)|_{z=0}$ is the complex amplitude profile at the BFP (usually the center plane of the target volume), λ is the wavelength and f is the focal length of the lens. As we are interested in the 3-D intensity distribution (or 2-D intensity profiles at multiple discrete depth planes), U(x, y, z) for any given z can be calculated through the Fresnel diffraction (or Fresnel propagation) relation [38]



Figure 1.2: Simplified diagram of an SLM-based 2f holographic point-cloud projection system.

$$U(x,y,z) = \frac{e^{jkz}}{j\lambda z} \iint_{-\infty}^{+\infty} U(p,q) \Big|_{z=0} e^{\frac{jk}{2z}[(x-p)^2 + (y-q)^2]} dp dq$$
(1.2)

where p, q are substitute variables for x, y at z = 0, and k is the wavenumber $2\pi/\lambda$. This can also be represented as a convolution between U(x, y, 0) and the radiation of a point source [38]. This relation is usually used in CGH algorithms to efficiently perform the projection of the hologram to depth planes of interest in the Fourier domain.

By specifying U(x, y, z) such that the volume light intensity profile is that of a cloud of points located on the neurons of interest (for imaging or stimulation), it is possible to recover a U'(x', y') to be configured onto the SLM surface through CGH. This process will be explained in more detail in Chapter 2, but for most SLMs, U'(x', y') is constrained such that |U'(x', y')| = 1 for all points on the SLM aperture. The resulting angles $\angle U'(x', y')$ then yield the phase element configurations across the array.

A conceptual diagram of a bidirectional, holographic, all-optical cortical neural interface is shown in Figure 1.3. In such a scheme, an optical channel to cortical neurons can be formed through the use of a transparent window in the skull. High-speed (high throughput and low latency) optical readout of neuron activity can be established by GEVIs while optical manipulation of neural activity is performed through optogenetics. Through the co-engineering of GEVIs and opsins, it is possible to achieve a full-duplex interface by multiplexing the read (fluorescence excitation and readout) and write (optogenetic stimulation)



Figure 1.3: Conceptual diagram of an all-optical full-duplex interface to the neural tissue. A fluorescence excitation laser is shaped by an SLM, and delivered onto cells of interest through the combination of a lens and an optical window into the cortex. The resulting fluorescent signal is collected in a different wavelength and focused onto the camera for processing. Optogenetic stimulation is delivered through a third wavelength band. The two excitation wavelengths are operated in a time-multiplexed manner to reuse the same SLM.

operations to non-overlapping optical wavelength bands [7]. The same SLM can be used to deliver both the fluorescence excitation and optogenetic stimulation lasers, operated in a time-multiplexed fashion. As discussed earlier, for the SLM to not impose a significant burden on the dynamic response of the optical pipeline, settling times <100 µs are required. However, this settling time cannot come at the expense of significant optical performance degradation, as optogenetic stimulation also necessitates high-contrast holograms, with such systems often aiming for contrast values of >20 to minimize off-target firing [8, 30]. While such metrics of hologram quality will be discussed in more detail in Chapter 2, they play a large role in the technology selection for the unit element of the SLM.

1.3 SLM Technologies

Most SLMs are realized as diffractive arrays of phase or amplitude-modulating elements. While it is not in the scope of this work to review all existing methods for spatial light modulation, in this section we give an overview of existing technologies that are either in active use in commercial products or have demonstrated scalability through academic works that involve arrayed elements.

The most widespread method to perform spatial light modulation is through a class of devices known as phase-only SLMs, which provide control over the phase of the light on a per-pixel basis. Such devices reconstruct a hologram that has uniform intensity across the array, restricting the hologram to only the phase of the complex amplitude at the SLM plane. This is, in fact not a severe constraint, and a variety of methods have been developed since the 1960s to either convert a complex amplitude hologram to a phase-only one or to generate a phase-only hologram given a 3-D intensity distribution [39]. Phase-modulating elements of SLMs are most commonly realized through the use of birefringent liquid crystal (LC) materials. These devices operate on the principle that by applying an electric field across a layer of LC, its molecules can be rotated to modulate the refractive index of the layer. To provide the pixel-specific electric fields, the LC layer is placed on a CMOS driver (referred to as liquid crystal on silicon, LCoS), forming a reflective phase-only SLM. LCoS SLMs are a mature class of commercially available devices, with megapixel-level array formats with pixel pitches of $\sim 10 \ \mu m$ and refresh rates (as determined by 10%-90% settling) of up to 500 Hz [6]. One major limitation of these devices is their refresh rate, as the layer thicknesses necessary to realize 2π phase shifts for longer wavelength ranges in the red and infrared part of the electromagnetic spectrum force longer relaxation settling times for the LC molecules [40]. Furthermore, LCoS devices require polarization of the incident beam. incurring further efficiency and optical component penalties. Finally, LCoS SLMs are often optimized for a specific wavelength range, making it difficult to reuse the same SLM for onephoton fluorescence excitation (~ 490 nm) and two-photon optogenetic stimulation (~ 1000 nm).

There are a variety of technologies that have been proposed and deployed to alleviate the speed bottleneck for applications that require higher temporal bandwidth or resolution. Some



Figure 1.4: An overview of some alternative technologies for fast spatial light modulation. On the left, various MEMS-based approaches including the phase light modulator (PLM) [41], continuous deformable mirrors [42], and grating-based OPAs [43, 44]. On the right, non-MEMS technologies including photonic IC-based thermo-optic [45], PN junction [46] and electro-optic phase shifters [47], as well as an emerging class of photonic crystal-based phase modulators [48].

of these technologies, and their advantages and disadvantages, are depicted in Figure 1.4. Piston-motion micromirrors displaced vertically can be used to modulate the travel distance of locally incident light for reflection off a planar mirror, creating phase differences across the array [49]. In this scheme, which will be discussed in more detail in Chapters 3 and 4, a suspended structure attached to the micromirror is displaced vertically, most commonly through electrostatic force between two parallel plates, usually referred to as top and bottom electrodes. By adjusting the voltage across the two electrodes, the displacement can be controlled as governed by the force balance equation between the recovery force of the springs formed by the suspension arms of the mirror structure, and the attractive electrostatic force between the plates. The Texas Instruments Phase Light Modulator (PLM) uses a similar approach but instead utilizes a constant voltage of 10 V, instead configuring the area of the bottom plate of the structure through digital drive [41]. While this drive scheme allows for the development of piston-motion micromirrors atop digital micromirror device (DMD) drivers, which are commercially available, it results in a highly nonlinear actuation profile, resulting in significant deterioration in image quality. DMD technology itself is not suitable for general-purpose SLM applications due to being limited to binary amplitude modulation only, causing severe efficiency penalties and image artifacts [50]. Continuous deformable mirrors (CDMs) are scaled to thousands of actuators, but their pixel-to-pixel coupled actuation and extremely high drive voltages of up to 200 V pose a challenge in scaling to 10-µm pitches and megapixel-scale array formats of LCoS SLMs [42].

Other structures that do not rely on vertical displacement have also been demonstrated.

For example, grating phase shifter-based MEMS optical phased arrays (OPAs) have been demonstrated to operate at settling times of down to 5.7 μ s, corresponding to a >175 kHz operation, with fabricated array formats of up to 160×160 [43]. Applicability of these devices to visible wavelengths requires advancements in MEMS fabrication techniques, however, as the features of the gratings need to be at the level of the wavelength of operation, posing fabrication challenges for wavelengths lower than 1000nm. Photonic IC (PIC)-based OPAs are another promising class of devices, distributing the light through integrated waveguides and achieving per-pixel phase shifts through thermo-optic [45], PN-junction [46], or electrooptic [47] phase shifters. While PIC-based OPAs are a very promising class of devices as alternatives to SLMs, pixel-level custom phase shift introduction has been a challenge for PICs. For thermal phase shifters, electrical efficiency becomes a scalability issue as the power required to introduce a phase shift of π is on the 1-10 mW order of magnitude range with existing approaches, and higher power efficiency usually sacrifices operating speed by thermally isolating the phase shifter from the substrate. Reverse-biased PN-junction phase shifters are electrically efficient but suffer from coupled phase-amplitude modulation, an undesirable characteristic for phase-only SLMs. Finally, while electro-optic modulation is much faster than other methods of modulation (>10 MHz), this technology has not been scaled to reliable array-scale fabrication yet, and existing materials also possess issues such as loss of poling [51]. More exotic solutions have also been proposed, such as photonic crystal-based unit elements modulated via out-of-plane illumination with µLEDs to control the phase [48]. Using this approach, arrays consisting of 64 elements were demonstrated to achieve nanosecond-level response times alongside high optical efficiency.

In this work, piston-motion MEMS micromirrors have been chosen as the unit element of the SLM, as this class of actuators has the combination of being commercially available through a multitude of MEMS foundries, open-loop and efficient electrical drive characteristics and well-understood mechanical and optical properties. Response times of these structures are fast enough to shift the bottleneck from the SLM to modern GEVIs and opsins and provide an excess refresh rate above their response times to allow for throughput improvement and speckle noise reduction as will be discussed in Chapter 2. High reflectivity achievable through the gold finish (>95% for 1000nm wavelength with 250nm layer thickness) allows for high-optical-efficiency operation. Finally, the well-understood relation between the applied voltages and the resulting dynamic and static responses allows for analysis-driven, trade-off-aware development of the driver electronics.

1.4 Thesis Contribution

This work's main objective is to investigate the optical, computational, and electrical requirements of piston-motion MEMS micromirror-based phase-only SLMs in the context of high-speed 3-D volumetric point cloud patterning for optical neural interfaces. In particular, we present the development and verification of a lightweight 3-D point cloud patterning CGH algorithm that matches the state-of-the-art algorithm in performance at 2-6 orders of magnitude faster computation times. We also present the design, implementation, and verification of two generations of array-scale micromirror drivers: 1. an 8-V mirror driver featuring a nonlinear DAC capable of correcting for global variations in micromirror fabrication in an area- and power-efficient manner and 2. a 29-V mirror driver capable of correcting local pixel variations and driving high-parasitic loads for interposer-based integration schemes between the MEMS and CMOS devices. The first-generation driver is verified with a MEMS mirror array operating as a varifocal element, tuning the focal point of a companion offset lens. The second-generation driver is designed for a MEMS array that was in fabrication at the time of writing this dissertation.

1.5 Thesis Organization

The rest of this thesis is organized as follows:

- Chapter 2: The theory behind 3-D point-cloud holography is explored in terms of its implications for SBP and the temporal bandwidth of the SLM. A computationally and optically efficient algorithm is presented for use in emerging high-speed SLM systems where closed-loop operation is desired and real-time computation of phase masks becomes limiting factors for the systems. The results are compared to the industry-standard Gerchberg-Saxton algorithm in terms of optical and computational performance.
- Chapter 3: A varifocal element (axial scanner) comprising an array of micromirrors and an accompanying 8-V driver ASIC is presented. The drive requirements, and specifically the number of bits in drive resolution, are explored, and the design process of the nonlinear DAC-based driver ASIC is described. This system demonstrates the utility of piston-motion micromirrors in applications requiring high-speed, dwellcapable axial scanners and presents an area- and power-efficient method to calibrate global process variations in micromirror fabrication.
- Chapter 4: A second-generation ASIC capable of driving the next iteration of MEMS micromirrors through high-parasitic traces is presented. The driver ASIC-facing effects of modifications to the mirror structure for added local and global process resilience, such as the drive voltage and two-step drive for faster settling, are discussed. Similar to Chapter 3, the resolution requirements for the ASIC are extracted, and the circuit implementation of the high-parasitic-interface-capable driver pixel is presented.
- Chapter 5: This chapter concludes the thesis with a summary of the results and important future research directions.

This chapter includes adaptations from the following articles:

N. T. Ersaro, C. Yalcin, and R. Muller, "The future of brain-machine interfaces is optical," Nature Electronics, 2023.

C. Yalcin, N. T. Ersaro, M. M. Ghanbari, G. Bocchetti, S. F. Alamouti, N. Antipa, D. Lopez, N. C. Pégard, L. Waller, and R. Muller, "A MEMS-based optical scanning system for precise, high-speed neural interfacing," IEEE Journal of Solid-State Circuits (JSSC), Jun. 2022.

Chapter 2

A Lightweight Algorithm for 3D Point-Cloud Holography

One of the bottlenecks in realizing a closed-loop, fast all-optical neural interface based on point cloud holography, is the computation of the hologram itself. CGH algorithms typically rely on algorithms that propagate a hologram to the target volume and apply an implicit cost function or extract an explicit cost function to evaluate, optimize and modify the hologram in an iterative fashion to converge onto the desired target volume intensity distribution. Recent development in neural network-based non-iterative algorithms aim to solve the excess computation time and improve upon the generated hologram quality, but such algorithms still require bulky hardware (particularly for the training) and need to constrain the model size by making concessions such as limited field-of-view and greatly reduced target depth plane counts. In this chapter, we present a point cloud CGH algorithm that relies purely on analytical calculation and stitching of phase masks to partition the available SBP of an SLM between the target points and provide a hologram that matches the performance of the most commonly used CGH algorithm at 2-6 orders of magnitude faster computation times.

2.1 Introduction

Several considerations frame the capabilities and limitations of SLMs in the context of 3D point cloud patterning. Under a fixed FoV, increasing the number of points targeted by an SLM frame decreases the available SBP apportioned to each target, doubly impacting target irradiance by simultaneously reducing target resolution and total optical power per target. Application-dependent requirements on target illumination therefore cap the allowable patterning throughput, such that target point count scales with SLM pixel count, i.e the total degrees of freedom available for modulation [8]. Hence, with expanding volumetric processing requirements across point cloud patterning applications, pixel count has emerged as a major bottleneck limiting targeting throughput in state-of-the-art SLMs, with diminishing returns impacting marginal improvements to SLM formats as a result of ever-increasing

power consumption, pixel crosstalk, device form factor, and cost [52, 53]. Importantly, even assuming an ideal SLM with unlimited SBP, the one-to-one correspondence formalized by the Fourier transform relationship that exists between complex amplitude profiles at the SLM and patterned planes does not extend to 3-dimensional volumes. Most 3D intensity distributions are indeed not optically realizable as they do not satisfy energy conservation and wave propagation principles and therefore require decomposition into sparse point clouds [54, 55]. Beamlets targeting distinct points also inevitably introduce undesirable illumination to non-target regions as they converge and diverge through the patterned volume, resulting in finite contrast. For a given SLM of finite pixel count and therefore finite SBP, increasing point cloud density in the patterned volume reduces contrast as irradiance in non-target regions gradually approaches irradiance in targeted spots [54]. Application-dependent contrast requirements therefore establish a critical scaling relationship between available pixel count and achievable target point count under 3D patterning. For instance, bioengineered opsins responsible for the modulation of neural activity under photostimulation typically require illumination contrasts in excess of 20 in order to avoid off-target excitation [54, 35].

Given such SBP constraints, time-multiplexed operation is often a critical requisite to true 3D point cloud patterning as it can decompose 3D intensity distributions into sets of sparse and separately realizable point clouds, in addition to providing speckle mitigation and polychromatic operation capabilities [34, 35, 36, 37]. Time multiplexing also aligns with developing performance regimes across current and future SLMs, where significant improvements to SLM speed are being achieved by eschewing slower phase modulation technologies (e.g. liquid crystal on silicon) in favor of faster micromirror-based or solid-state transduction mechanisms [41, 56, 48, 57]. Such higher-speed SLMs have already exceeded the kilohertz timescales needed to accommodate the millisecond integration windows of neuron opsins and voltage indicators in neuroscience [30, 58]. Under time multiplexing, contrast becomes doubly crucial as duty-cycled illumination across multiple frames erodes mean spot power relative to background irradiance, demanding higher contrast performance from each individual SLM frame [19, 20]. For example, in 3D nanofabrication, the power law-dependent efficiency of polymerization photoreactions in multiphoton lithography determines photon budgets, exposure times and irradiance thresholds for target and non-target regions, setting allowances on both contrast and the number of multiplexed frames [18, 1, 19, 20]. Accordingly, the performance of the CGH algorithm employed for phase mask computation is critically important to both maximizing time-averaged target contrast and achieving real-time operation for scenarios including closed-loop read/write interfacing with biological tissue or interactive AR/VR [7, 9, 11, 13].

Yet existing phase retrieval algorithms are unable to strike the necessary balance between a volumetric model that is suitable for the generation of 3D point clouds of sufficient quality and minimally-intensive computation and memory requirements that could be compatible with real-time deployment. The most accurate point cloud representations in implemented computation models construct holograms by superimposing the phase patterns that are associated with each target location from a look-up table [59, 60]. However, such approaches require exorbitant storage and data transfer speeds as the stored data for each point location

consists of a full SLM-sized 2D matrix, despite each target being apportioned a fraction of the SLM's total SBP. Efforts to reduce memory usage via quality reduction of the 3D representation map or additional real-time computation remain nonviable for real-time deployment [10, 12, 1, 60]. Alternatively, efforts to alleviate computation for each point location via a wavefront recording plane located near enough to the target volume that propagations from separate targets do not overlap mandate a small depth range, severely constraining the accessible volume for patterning [61, 62].

The memory and computation burden imposed by true volumetric point cloud representations have driven the emergence of layer-based methods that discretize the available depth range to a finite set [63, 64]. The most straightforward and widely used implementations of this strategy make use of the iterative Gerchberg-Saxton (GS) algorithm, which computationally propagates complex fields back and forth between SLM and target volume planes while enforcing amplitude constraints in order to converge on a hologram for a given point cloud distribution [55, 65]. Gradient descent-based iterative approaches may also be employed for improved hologram optimization via custom penalties at the cost of more computation [66, 67]. Yet despite the concession of a limited number of addressable depth planes, iterative approaches remain prohibitive to real-time computation as each propagation step requires a computationally expensive, full SLM format-size fast Fourier transform (FFT) operation. For an SLM of pixel format $F \times F$ targeting a volume discretized into M planes, the resulting computational complexity for each iteration is $O(MF^2\log(F))$, displaying scaling behavior that negates the benefits of ongoing improvements to spatiotemporal SLM bandwidth. While deep learning offers a promising avenue to fast CGH, employed memory-intensive convolutional neural networks (CNNs) require costly graphics processing unit (GPU) or accelerator resources and demand an onerous and context-specific training process that limits generalizability and easy redeployment across SLM formats and point-cloud requirements. More importantly, they drastically limit addressable depths to a handful of distinct planes [13, 68, 69, 70].

Fundamentally, current compute regimes can be attributed to the inefficient allocation of available degrees of freedom to target points in the CGH computation process. For a 3D point cloud consisting of T target points, each point separately adds expensive $F \times F$ matrix-wide computation despite effectively being allotted only F^2/T degrees of freedom. In order to achieve real-time compute regimes under true volumetric point cloud patterning, we propose an FFT-free efficiency-driven approach employing lightweight deterministic phase calculations that scale primarily with T for the optimal allocation of available degrees of freedom across targets. Accordingly, we present NIMBLE-PATCH, an algorithm with Non-Iterative, Multi-Block, Local Efficiency-driven Point Assignment and Targeting for Cloud-based Holography. In addition to maximizing both overall diffraction efficiency and FoV to make optimal use of pixel count and pitch constraints impacting current SLM offerings, NIMBLE-PATCH employs a patchwork hologram construction approach that is FFT-free and sampling-agnostic, allowing for truly arbitrary target positions and aliasingfree performance.

In order to rigorously evaluate algorithm performance, we developed a computational

simulation framework that is agnostic to optical system parameters and accounts for volume and resolution scaling across SLM format F, target count T, and time-multiplexed frame count N. This framework was employed for a systematic comparison between NIMBLE-PATCH and GS-based algorithms across F and T to identify performance and computation time trends. Compared against the least computationally burdensome implementation of GS involving minimal sampling of the SLM and target planes, NIMBLE-PATCH reaches double the contrast values of GS at SLM formats as low as 512×512 within compute times that are $> 10^4$ faster. In addition, an improved implementation of GS matching the contrast performance of NIMBLE-PATCH via higher sampling was found to be $> 10^5$ slower for formats as low as 512×512 . The obtained results were subsequently confirmed with experimental demonstrations involving the effective reformatting of a real SLM. Lastly, a time-averaging investigation demonstrated that NIMBLE-PATCH best mobilized the excess SLM refresh rate available for N-multiplexed operation as it achieved the best contrast while retaining a compute time advantage of several orders of magnitude relative to GS.

2.2 Description of NIMBLE-PATCH Algorithm and Evaluation Framework

We introduce the principles underlying the NIMBLE-PATCH algorithm with a treatment of 3D point steering holograms and their target location-dependent variations in regional diffraction efficiency. For a given prototypical CGH system with a Fourier-transforming lens of focal length f and at an optical wavelength λ , the phase shift φ required at (x, y) positions across the hologram plane to achieve lateral point steering to a location $(x', y') = (d_{x'}, d_{y'})$ at the rear focal plane is given by:

$$\varphi_{\text{lateral}}(x,y) = -2\pi \left(\frac{d_{x'}}{\lambda f}x + \frac{d_{y'}}{\lambda f}y\right)$$
(2.1)

Similarly, the hologram phase mask required to achieve axial steering to a depth $z' = d_{z'}$ is a spherical profile paraxially approximated as a paraboloid [56] as given by:

$$\varphi_{\text{axial}}(x,y) = \frac{2\pi}{\lambda} \left(\frac{f^2}{d_{z'}} - \frac{f^2}{d_{z'}} \sqrt{1 - \left(\frac{d_{z'}}{f^2}\right)^2 (x^2 + y^2)} \right) \approx \frac{\pi d_{z'}}{\lambda f^2} \left(x^2 + y^2 \right)$$
(2.2)

Convolving lateral deflection together with axial deflection for joint 3D point steering entails multiplying the complex axial and lateral fields together in the Fourier domain at the hologram plane, which corresponds to summing the phase profiles given in Eqs. (2.1) and (2.2):

$$\varphi_{3D}(x,y) = \frac{\pi}{\lambda f} \left(\frac{d_{z'}}{f} x^2 - 2d_{x'}x + \frac{d_{z'}}{f} y^2 - 2d_{y'}y \right)$$

$$= \frac{\pi d_{z'}}{\lambda f} \left(x - \frac{d_{x'}f}{d_{z'}} \right)^2 + \frac{\pi d_{z'}}{\lambda f} \left(y - \frac{d_{y'}f}{d_{z'}} \right)^2 - \frac{\pi}{\lambda d_{z'}} \left(d_{x'}^2 + d_{y'}^2 \right)^2 \qquad (2.3)$$

$$\equiv \frac{\pi d_{z'}}{\lambda f} \left(x - \frac{d_{x'}f}{d_{z'}} \right)^2 + \frac{\pi d_{z'}}{\lambda f} \left(y - \frac{d_{y'}f}{d_{z'}} \right)^2$$

Completing the square and dropping the piston phase offsets as shown in Eq. (2.1) demonstrates that 3D point-steering phase masks simply correspond to the parabolic profiles required for pure axial steering with a laterally shifted vertex location given by $(d_{x'}f/d_{z'}, d_{y'}f/d_{z'})$. Examples of such phase masks are illustrated in Fig. 2.1. As the target depth plane approaches the rear focal plane $d_{z'} = 0$, Eq (2.3) simplifies to the pure lateral steering expression in Eq. (2.1), resulting in a uniform phase gradient as seen in Fig. 2.1(b).

Under the spatially discretized phase mask produced by a real SLM of pixel pitch p and format $F \times F$, the regional diffraction efficiency η at a given location (x_0, y_0) on the SLM plane can be determined from the mean phase steps $\Delta \varphi_x(x_0, y_0)$ and $\Delta \varphi_y(x_0, y_0)$ to adjacent phase pixels at that location along the two orthogonal axes of the SLM as given by the following relationship [52]:

$$\eta(x_0, y_0) = \left(\frac{\sin\left(\frac{\Delta\varphi_x(x_0, y_0)}{2}\right)}{\frac{\Delta\varphi_x(x_0, y_0)}{2}}\right)^2 \left(\frac{\sin\left(\frac{\Delta\varphi_y(x_0, y_0)}{2}\right)}{\frac{\Delta\varphi_y(x_0, y_0)}{2}}\right)^2$$

$$= \left(\frac{\sin\left(\frac{p}{2}\frac{\partial\varphi}{\partial x}(x_0, y_0)\right)}{\frac{p}{2}\frac{\partial\varphi}{\partial x}(x_0, y_0)}\right)^2 \left(\frac{\sin\left(\frac{p}{2}\frac{\partial\varphi}{\partial y}(x_0, y_0)\right)}{\frac{p}{2}\frac{\partial\varphi}{\partial y}(x_0, y_0)}\right)^2$$
(2.4)

Under purely lateral steering, i.e. $d_{z'} = 0$, the uniform phase gradient results in the following simplified relationship:

$$\eta(x_0, y_0) = \left(\frac{\sin\left(\frac{\pi p d_{x'}}{\lambda f}\right)}{\frac{\pi p d_{x'}}{\lambda f}}\right)^2 \left(\frac{\sin\left(\frac{\pi p d_{y'}}{\lambda f}\right)}{\frac{\pi p d_{y'}}{\lambda f}}\right)^2 \tag{2.5}$$

Under joint axial and lateral steering, i.e. $d_{z'} \neq 0$, phase gradients along each axis scale linearly with distance to the parabolic vertex location $(d_{x'}f/d_{z'}, d_{y'}f/d_{z'})$ identified from Eq. (2.3), resulting in the following generalized expression for regional diffraction efficiency:

$$\eta(x_0, y_0) = \left(\frac{\sin\left(\frac{\pi p d_{z'}}{\lambda f^2} \left(x_0 - \frac{d_{x'}f}{d_{z'}}\right)\right)}{\frac{\pi p d_{z'}}{\lambda f^2} \left(x_0 - \frac{d_{x'}f}{d_{z'}}\right)}\right)^2 \left(\frac{\sin\left(\frac{\pi p d_{z'}}{\lambda f^2} \left(y_0 - \frac{d_{y'}f}{d_{z'}}\right)\right)}{\frac{\pi p d_{z'}}{\lambda f^2} \left(y_0 - \frac{d_{y'}f}{d_{z'}}\right)}\right)^2$$
(2.6)



Figure 2.1: Illustration of regional variations in phase mask efficiency under 3D point steering for a fixed lateral target position and (a) positive, (b) zero, and (c) negative target focus depths. Left-column plots correspond to SLM phase shift values along the x-axis for pixels at y = 0. Relative power contributions to target and non-target diffraction orders across SLM regions shown in center-column optical schematics and right-column efficiency plots (with zeroth order steering ranges denoted by gray zones).

The expression in Eq. (2.6) can serve to evaluate the relative contributions of different regions of the SLM to the targeted spot under 3D point steering. As illustrated in Fig. 2.1, the most efficient SLM region for a given target corresponds to its phase mask's parabolic vertex, whose location depends on the 3D position of the targeted spot. Additionally, once the phase step $\Delta \varphi(x_0, y_0)$ at a given SLM location away from the vertex exceeds π along either axis, the targeted spot falls outside of the achievable angular diffraction range of $arctan(\lambda/p)$ at that SLM location. Alternatively stated, that SLM region directs the bulk of its optical power to an off-target 3D point position considered to be within the region's zeroth diffraction order, thereby contributing only marginal power to the targeted 3D point position as a higher diffraction order. This behavior has implications on sampling requirements for FFTbased CGH computation approaches, including GS. The least computationally burdensome implementations of such approaches involving minimal sampling of the hologram plane (i.e., 1) computational pixel per SLM pixel) risk poor performance by failing to properly account for relative contributions to different diffraction orders as a result of aliasing [71]. This impact is especially prominent with increased axial steering away from z' = 0, which entails steeper parabolic phase gradients.

Given a 3D point cloud consisting of multiple target point positions with different associated peak-contribution regions across the SLM plane, NIMBLE-PATCH exploits the described deterministic relationships to partition the available SBP across points, allocating SLM pixels to each target location for maximum overall efficiency. The algorithm's procedure involves initially partitioning the SLM array into evenly-sized patches, i.e. pixel blocks, in accordance with FoV, resolution, and throughput requirements. The full 3D point cloud is subsequently decomposed into subsets of point clouds to be addressed by a number of frames set in accordance with time multiplexing capabilities, with every target point being assigned to a specific SLM patch on its respective frame. This allocation is accomplished by calculating regional diffraction efficiencies at each patch center for each target using Eqs. (2.5) and (2.6) in order to construct a cost matrix. The associated linear assignment problem between each distinct patch in every available SLM frame and each separate target spot is solved via the Jonker-Volgenant method [72]. Finally, phase masks are computed separately for each target and only across the corresponding SLM patch using the simple beam-steering and focus-tuning relationships in Eqs. (2.1) and (2.3), then stitched together to generate the full-frame phase profiles.

Since CGH computation algorithm performance is heavily dependent on SLM format Fand targeting throughput, given by total target count T across N time-multiplexed frames, NIMBLE-PATCH was implemented and evaluated against GS across sweeps involving all three parameters [73]. Two versions of GS were included in the comparison: one implementation denoted by GSx1 minimizes computation with a sampling scheme of 1x1 computational pixel per SLM pixel, and a second implementation denoted by GSx3 prioritizes performance at the expense of computation burden with a sampling scheme of 3x3 computational pixels per SLM pixel. In order to ensure that the comparison across algorithms is agnostic to context-dependent optical system parameters (including optical wavelength, lens focal length, and SLM pitch), sweeps across F, T, and N were normalized to the available SBP



Figure 2.2: SLM partitioning principle underlying NIMBLE-PATCH algorithm with SLM format and phase mask in the left column, volume dimensions and target point cloud in the center column, and target spot dimensions in the right column. Increasing the SLM format (first to second rows) at fixed SLM pitch and target count maintains lateral FoV and reduces axial FoV and spot size. Subsequently, increasing the target count by the same amount (second to third rows) restores the original FoV and spot size dimensions.

across the unit SLM patch targeting a distinct point. Specifically, lateral and axial FoVs were confined to the allowable volume determined by the patch format, and target spot sizes used for irradiance, contrast, accuracy, and efficiency calculations were set based on the patch's achievable lateral resolution. The partitioning principle of the NIMBLE-PATCH algorithm and its associated impact on target FoV and spot dimensions is illustrated in Fig. 2.2.

As evidenced by Eq. (2.1), lateral point steering is subject to a gradual efficiency roll-off that caps lateral FoV as given by $\text{FoV}_{x,y} = \lambda f/p$. The F-fold increase of this FoV relative to the Abbe diffraction limit for lateral spot width $w_{x,y} = (\lambda f)/(pF)$ captures the SBP available for lateral steering along one axis. An axial FoV can similarly be calculated by noting that lateral FoV is capped at $\Delta \varphi = \pi$ and solving for the depth at which mean $\Delta \varphi$ along a given SLM axis is equal to π : FoV_z = $(16f^2\lambda)/(F(\lambda^2 + 4p^2)) \approx (4f^2\lambda)/(Fp^2)$. The F/2-fold increase of the axial FoV relative to the Abbe diffraction limit for axial spot width $w_z = (8f^2\lambda)/(F^2p^2)$ can be attributed to the fact that focus tuning employs circularly symmetric phase masks with a radial range spanning F/2 pixels along either SLM axis.

Efficiency roll-off profiles along both a single lateral axis and the axial dimension are shown in Fig. 2.3, with axial steering experiencing compounded loss relative to lateral 1D



Figure 2.3: Efficiency roll-off profiles along the (a) X axis and (b) Z axis obtained from single point steering using NIMBLE-PATCH in the simulation framework. Lateral 1D steering aligns with theoretical prediction, and axial steering experiences compounded loss relative to lateral 1D steering as a result of the 2D radial phase modulation required for focus tuning.

steering as a result of the 2D radial phase modulation required for focus tuning. These efficiency roll-offs may place additional constraints on allowable FoV for applications that require minimal variation in spot power across point clouds [74]. For instance, optical interfacing in biological microscopy sets a minimum bound on spot power based on the required photoexcitation as well as a maximum bound based on tissue heating and photobleaching limits [8, 30]. Accordingly, lateral and axial FoV ratio parameters were incorporated into our implemented simulation framework in order to provide the option of constraining 3D point cloud volume to desired fractions of the full FoVs. We note that an axial range ratio of 0.25 can serve to ensure that on-axis targets (i.e. targets along z at (x', y') = (0, 0)) remain within the zeroth diffraction order across all regions of the SLM for FFT-based algorithms involving minimal sampling [55, 66]. However, this restriction on the SBP available for focus tuning does not extend to off-axis points $((x', y') \neq (0, 0))$, such that costlier SBP concessions would be needed to ensure zeroth order targeting across the full point cloud volume.

The simulation framework constructed for this work (implemented in Python 3.9) models SLM-driven volumetric point cloud patterning under a 2f optical configuration involving a Fourier-transforming lens via Fast-Fourier Transform (FFT) and Fourier-domain Fresnel propagations. For GS computation, an iteration count of 50 was used to ensure a sufficient level of convergence for the point cloud hologram [66, 75]. Each SLM pixel is represented by a programmable number of simulation pixels (set to 1x1 for GSx1 phase mask computation, 3x3 for GSx3 phase mask computation, and 5x5 for overall algorithm evaluation and metric calculation) to capture higher diffraction orders in the target volume. Additionally, the SLM phase mask is zero-padded (with a pad size equal to SLM size on all sides) to position diffraction-limited target spots with sub-diffraction-limit precision, to represent the spot shape as a disk for improved GS performance, and to minimize binning-related inaccuracies from FFT operations due to insufficient spatial granularity. To account for electronic drive non-idealities, each computed phase mask is discretized to a programmable number of bits

(set to 8 in all performed sweeps). Efficiency roll-off profiles obtained from the simulation framework both along a single lateral axis and the axial dimension shown in Fig. 2.3.

Axial range and depth plane count for random target generation was computed from each run's F, T, and N values. Target location generation was further constrained in the axial dimension to a discrete set of distinctly resolvable depth planes for the computational benefit of GS, as each additional depth plane linearly impacts GS computation while not changing the computational complexity of NIMBLE-PATCH. In practice, spot positioning may require sub-spot-size precision in applications including super-resolution microscopy, further increasing the computational burden required of GS [76]. In the lateral dimensions, target positions were generated randomly across any of the allowed depth planes, with placement granularity determined by the choice of zero padding. An additional target spacing constraint was introduced to prevent target spots from overlapping.

2.3 Single-frame Algorithm Performance Comparison

NIMBLE-PATCH was first compared against GSx1 and GSx3 under single-frame patterning (N=1) for randomly distributed 3D point clouds. The comparison was made across 5 SLM formats (F=32,64,128,256,512) and up to 5 total target counts (T=1,4,16,64,256 targets). Lateral and axial range ratios were set to 0.9 and 0.75, respectively, for the target volume. This reduction in FoV ensures that all generated spots have a theoretical efficiency 27.5%(Fig. 2.3). In order to accommodate the discretization limitation of the GS algorithms, allowable target point depths were constrained to a finite number of depth planes evenly spaced across the axial FoV in accordance with the available SBP across each patch. Maximum target count simulated for a given F is therefore reached once distinct, non-overlapping spots can no longer be placed across these target depth planes. For each (F, T) pair, at least 25 randomized distributions were simulated (>80 for F <512), and each randomly generated distribution was evaluated across all three algorithms. For each simulation, the target volume intensity distribution is denoted with I(x', y', z') and the generated volume intensity distribution is denoted with G(x', y', z'). I(x', y', z') is discretized in z, and is only defined for depth planes that include targets. Targets in I(x', y', z') are disk-shaped pixel regions with a value of 1 and a disk size corresponding to the Abbe diffraction limit, I(x', y', z') is 0 across all non-target regions. In order to quantify the results of the simulations and compare CGH algorithms, two main metrics were considered: contrast C and computation time. Contrast measures the ratio of irradiance in target regions to irradiance in non-target regions:

$$C = \frac{\frac{\sum G(x',y',z')I(x',y',z')}{\sum I(x',y',z')}}{\frac{\sum G(x',y',z')(1-I(x',y',z'))}{\sum (1-I(x',y',z'))}}$$
(2.7)

Computation time measures the time elapsed between the input of the volume information to the algorithm (point cloud target coordinates for NIMBLE-PATCH, I(x', y', z') for GSx1 and GSx3) and the output of the computed phase mask. Reported computation times

were obtained from phase mask calculations performed on a single core of the Intel Xeon E5-2670 v3 CPU and 384 GB of RAM.

In addition to contrast and computation time, two supplementary metrics were calculated for each distribution: accuracy α and efficiency η . Accuracy α corresponds to the crosscorrelation between the desired and generated intensity distributions and is defined as:

$$\alpha = \frac{\sum G(x', y', z')I(x', y', z')}{\sqrt{(\sum G(x', y', z')^2)(\sum I(x', y', z')^2)}}$$
(2.8)

Efficiency η corresponds to the ratio of the power in the targets to the total power in the volume, and is defined as:

$$\eta = \frac{\sum G(x', y', z')I(x', y', z')}{\sum G(x', y', z')}$$
(2.9)

The choice to present contrast as the main metric stems from its importance in pointcloud holography applications [76] and its well-behaved and predictable scaling behavior as a function of target count. As the target count quadruples, the SLM patch area dedicated to each target shrinks by a factor of 4, reducing the available optical power directed to the spot by the same factor. Furthermore, since the patch size reduction also corresponds to an increase in spot size by the same factor as depicted in Fig. 2.2, target irradiance drops by an additional factor of 4. This results in a theoretical decrease in target irradiance that is proportional to the square of the number of targets, i.e., a ~40 dB/decade decay in contrast plotted against the target count (ignoring the associated increase in the background irradiance and without taking algorithm-specific performance into account).

Fig. 2.4 (a) and (b) show aggregated contrast and computation time results, respectively, across all single-frame simulation runs. Contrast follows the general 40 dB/decade decay trend for all algorithms across all formats, with the added contribution of algorithm-specific performance deviations. For NIMBLE-PATCH, computation time initially drops with increasing T due to the speed-up of phase mask computations under increased SLM partitioning resulting in smaller patch sizes, decreasing the computation time of matrix declarations associated with the computation. Computation time then rises, converging between formats F, as the linear sum assignment solution, which depends only on T, emerges as the limiting factor. For GSx1 and GSx3, increasing T raises computation time on average as the number of distinct target depth planes that need to be accounted for grows together with point cloud density. However, once T exceeds a certain threshold for each F, a roll-off in computation time occurs as randomly generated target locations reliably cover all allowable target depths. Any subsequent increase in T causes a decrease in this cap of allowable depth planes, speeding up computation. Fig. 2.4(c) shows the ratios between contrast in NIMBLE-PATCH and contrast in either of the two implementations of GS. Up until the SBP limit for the SLM format is reached and the absolute values for contrast drop below ~ 10 , NIMBLE-PATCH achieves on average $\sim 1.6-1.8x$ higher contrast compared to GSx1, likely as a result of aliasing from minimal sampling in GSx1. GSx3, on the other hand, matches NIMBLE-PATCH


Figure 2.4: Sweep results comparing NIMBLE-PATCH to GS for single-frame simulations. (a) Contrast and (b) computation time as a function of target count T across SLM formats F. (c) Contrast ratios and (d) computation time ratios between NIMBLE-PATCH and the two GS algorithms (top: GSx1, bottom: GSx3). 95% confidence intervals for the mean value of the curves are shown as shaded regions for each plot.

patch performance as early as T > 1, further corroborating the sampling-related limitations of GSx1. Fig. 2.4(d) shows the ratio of computation time between NIMBLE-PATCH and the two implementations of GS. NIMBLE-PATCH was found to complete phase computation 1.5 – 5 orders of magnitude faster than GSx1 and 2 – 6 orders of magnitude faster than GSx3 for the considered SLM formats.

Fig. 2.5(a,b) shows single-frame scaling trends for these two metrics across SLM format and target count per frame, and Fig. 2.5(c,d) provides ratiometric comparison results between the three algorithms. NIMBLE-PATCH outperforms GSx1 in terms of accuracy α across most F and T cases, and GSx3 at low T regimes in all F cases. It is important to note that efficiency η is closely related to contrast C. For a given target distribution, efficiency η can be approximated to be contrast C scaled by the ratio of total target spot area to the total size of the non-target area across all depth planes:

$$\eta \approx \frac{\sum I(x', y', z')}{\sum (1 - I(x', y', z'))} C$$
(2.10)

This approximation holds for sparsely populated target point cloud volumes, which is generally applicable across the performed sweeps. Though the scaling factor depends on the targeted point cloud, it remains equal across algorithms for a given set of targets. This relationship explains why efficiency ratios, shown in Fig. 2.5(d), closely follow the contrast ratios shown in Fig. 2.4(c). The motivation behind presenting contrast C as the main metric

CHAPTER 2. A LIGHTWEIGHT ALGORITHM FOR 3D POINT-CLOUD HOLOGRAPHY



Figure 2.5: Sweep results comparing NIMBLE-PATCH to GS for single-frame simulations in terms of the two auxiliary metrics. (a) Accuracy and (b) efficiency as a function of target count N across SLM formats F. (c) Accuracy ratios and (d) efficiency ratios between NIMBLE-PATCH and the two GS algorithms (top: GSx1, bottom: GSx3). 95% confidence intervals for the mean value of the curves are shown as shaded regions for each plot.

stems from its implicit normalization across different target profiles as compared to efficiency η , and its closer relevance to application-level performance as compared to accuracy α .

2.4 Experimental Evaluation of Single-Frame Algorithm Performance

Experimental Setup and Results

Algorithm performance findings from simulated sweeps were verified in experiment under the imaging setup shown in Fig. 2.6(a) using a 532 nm diode-pumped solid-state laser source (Thorlabs CPS532). Phase masks computed by each algorithm across a set of target point clouds were applied using a 1272x1024 SLM (X13138-01 12.5µm-pitch Hamamatsu LCoS) and the resulting patterned volumes were acquired as z-stacks using a 12-bit camera (Black-Fly S BFS-U3-200S6C-C) mounted onto an automated z-stage (Zaber X-LSQ150A). In order to replicate the sweeps performed in simulation across SLM format F and target count T, the central 1024x1024 pixel region of the SLM was reformatted to smaller effective pixel counts via real pixel grouping. A binary 0 and π checkerboard phase pattern was applied at a pitch of 2 real pixels across the remaining regions of the rectangular-format SLM in order to diffract light away from the point cloud volumes patterned by the central reformatted square SLM area. The range of allowable values for F and T was jointly constrained by this

reformatting scheme, the largest axial FoV that can be accommodated by the stage, and the sampling limit set by the pixel pitch of the camera. In order to better accommodate (F, T) sweeps, a focal length of 100 mm was chosen for the Fourier-transforming lens, and the lateral and axial range ratios constraining targeted point cloud volumes were chosen to be 0.8 and 0.5, respectively. Three SLM formats (F=64x64, 128x128, 256x256) and three target counts (T=1, 4, 16) were investigated, with two point-cloud distributions evaluated across all three algorithms for each (F, T) pair, for a total of 54 unique phase masks patterning distinct volumes.

Representative images of computed phase masks and 2D maximum intensity projections of both simulated and acquired point cloud volumes are shown in Fig. 2.6(b-f). The obtained projections show close agreement between simulated and measured results. Experimental non-idealities include some depth distortion from target depth-dependent aberrations (Fig. 2.8), and a central spot formed at (x', y', z') = (0, 0, 0) from the unmodulated illumination and DC diffraction component resulting from the physical SLM's finite pixel fill factor, phase error, etc. GSx1 performance was found to be consistently worse than NIMBLE-PATCH, with a reduction in target power that is increasingly apparent at higher target counts and distant depth planes (Fig. 2.6(e-f)). In addition, the disappearance of certain target spots at smaller target counts (Fig. 2.6(b,d)) further confirms the susceptibility of the minimally sampled GSx1 algorithm to aliasing impacts. Qualitatively, GSx3 performance was found to be comparable to that of NIMBLE-PATCH. We note that both GS algorithms tend to produce targets that are more confined relative to the desired spot sizes as provided by I(x', y', z'), resulting in higher intensity peaks for the same spot power. This can be attributed to the fact that patch partitioning is not enforced in GS. GS pixel ensembles targeting separate point locations are thus interleaved across the SLM, leading to higher spot resolution, as revealed by an examination of generated phase masks in Fig. 2.6(e-f).

Algorithm performance was quantified through a processing pipeline that identifies peak intensity locations and assigns them to target positions based on proximity. Target lateral full widths at half maximum (FWHMs) generally show good agreement between measured, simulated, and expected spot sizes (see Fig. 2.7(a)), with some deviation in FWHM estimation emerging in the experiment from either proximity to the static central spot (e.g. at F=64, T=1) or a large FoV that is more susceptible to aberration impacts (e.g. at 256x256) formats where axial FoVs reach 70 mm). Spot size was also found to be consistently smaller with GS algorithms, confirming that effective single-target SLM patch regions in GS are larger than those in NIMBLE-PATCH from phase mask interleaving. The accuracy of the simulation framework is further evidenced by the fact that experimentally measured target irradiances recapitulate the same trends seen in simulation, as shown in Fig. 2.7(b). Importantly, we note that irradiances in NIMBLE-PATCH and GSx3 follow each other closely, indicating that both algorithms perform similarly in their ability to allocate power to each target location. In alignment with simulation sweep results and qualitative evaluations of experimental projections, markedly degraded irradiance performance is also observed with GSx1.



Figure 2.6: (a) Experimental test setup for point-cloud volume acquisitions under SLM reformatting. ND: neutral density filter, OBJ: objective (10x), SF: spatial filter (25 µm pinhole), L1: collimating lens (focal length: 180 mm), P1: polarizer, M1/M2: alignment mirrors, BS: beamsplitter, L2: Fourier-transforming lens (focal length: 100 mm). Representative images of phase masks computed by each algorithm along with simulated and experimental 2D maximum intensity projections for (b) F=64, T=4, (c) F=128, T=4, (d) F=256, T=4, (e) F=128, T=16, and (f) F=256, T=16. Projection images span the FoV associated with each (F,T) condition, projection orientations across image layouts are specified in (b). Identified and non-identified targets are marked by orange and red ellipses, respectively.



Figure 2.7: (a) Lateral FWHM spot sizes measured across simulated and experimental point cloud volumes evaluated at each format/target count (F, T) condition (error bars: standard deviation). The shaded gray region, spanning theoretical FWHMs corresponding to the side and diagonal apertures of the associated square SLM patch, is shown as a reference range due to spot sampling and pixel integration limitations with the simulation environment and camera pitch. (b) Target irradiances in arbitrary units across simulated and experimental point cloud volumes evaluated at each (F, T) condition (error bars: standard deviation).



Figure 2.8: Intended target depth plotted against recorded peak intensity depth from on-axis focus tuning calibration for experimental acquisitions. The measured deviation is accurately captured by third-order polynomial fit.

Experimental Data Processing and Quantification

For each evaluated phase mask at a given format/target count (F, T) condition, the camera exposure time was adjusted to avoid saturation, and an image z-stack was obtained by averaging two acquisition runs at a z-stage step size corresponding to 1/12th of the Abbe axial spot size. In order to correct for aberrations impacting focus tuning as well as misalignment between the stage-mounted camera plane and the optical axis in the acquired z-stacks, a single-point on-axis focus tuning calibration was performed together in order to record both the measured depth and optical axis position (in camera pixel coordinates) corresponding to various targeted depths. The recorded deviation between measured and intended depths, which was accurately accounted for with a third-order polynomial fit as shown in Fig. 2.8, can likely be attributed to target depth-dependent spherical aberrations from the beamsplitter and Fourier-transforming lens.

Using linear regression of the recorded optical axis coordinates against measured depths, the acquired z-stacks were computationally realigned by centering each image slice to the optical axis then cropping to the lateral FoV bounds of the evaluated (F, T) pair. Peak intensity locations along x', y', and z' were subsequently identified, and a linear sum assignment was employed to assign each targeted location to an identified peak using distance as a cost matrix. Full widths at half maximum (FWHMs) along x', y', and z' were also recorded for each peak as part of the peak identification process. A target point was considered non-identified (red ellipses in Fig. 2.6) if the assigned peak is located more than 3 FHWMs away along

each dimension as given by: $\text{FWHM}_{x,y} = 1.02w_{x,y}$ and $\text{FWHM}_z = 0.9w_z$. The percentage of successfully identified targets and resulting position errors between identified targets and their assigned peak locations are shown in Fig. 2.9(a) and Fig. 2.9(b), respectively. In order to compare target optical power across (F, T) conditions, recorded camera pixel values were scaled in accordance with acquisition exposure times, and irradiances were evaluated across target size disk regions set by patch dimensions and defined in I(x', y', z').

2.5 Time-multiplexing Performance Evaluation

In order to maximize the SBP mobilized toward each target and hence the number of distinct targets in the target volume, the target count per frame (and hence patch count in NIMBLE-PATCH) must be minimized. However, the number of targets that must be generated within a given timeframe may be dictated by application-specific requirements. If the SLM has excess refresh rate capable of switching between multiple holograms above the temporal bandwidth relevant to the application, time averaging across multiple holograms can be employed to simultaneously achieve the desired resolution and target count across the given FoV and timeframe. NIMBLE-PATCH can be time-multiplexed since the total set of T targets are optimally distributed across frame-specific patches as part of the assignment process (see decomposed NIMBLE-PATCH, Fig. 2.10(a)). The same efficiency-driven interframe target decomposition process can be applied to GS, with each individual phase mask being computed by GS instead of NIMBLE-PATCH (see decomposed GS, Fig. 2.10(b)). For each 3D point cloud, a single GS hologram addressing the complete set of targets within just one frame was also computed to serve as a benchmark. The decomposed NIMBLE-PATCH, decomposed GSx1, and single-frame GSx1 benchmark algorithms were evaluated for their relative abilities in making optimal use of time-multiplexing capabilities via performance sweeps across the available excess refresh rate. GSx3 was not considered in this analysis as it could not be viably investigated for time-multiplexed performance given its significant computational burden.

Fig. 2.11 shows the results of excess refresh rate (N) sweeps at two total target count values (T=64 and T=256). As N increases, the target count per frame (and therefore the patch count in each NIMBLE-PATCH hologram) decreases, which results in a decrease in spot size and an increase in addressable depth plane count. At this higher spatial degree-of-freedom regime, contrast generally increases due to shrinking spot size relative to the accessible FoV. Notably, decomposed NIMBLE-PATCH outperforms both decomposed GS and single-frame GS in higher N regimes, with a crossover point that depends on F (but occurs within N_i16 regime for all cases). As increasing N decreases patch count per frame, the complexity of the balanced assignment problem remains constant in NIMBLE-PATCH, resulting in a relatively flat computation time across the refresh rate space. Another observation from these sweeps is that GS does not benefit from excess refresh rate in the same way NIMBLE-PATCH does, as single-frame GS outperforms decomposed GS for most cases.

Fig. 2.11 also serves to evaluate general tradeoffs between SLM format and speed. Spot



Figure 2.9: (a) Percentage of successfully identified targets across both point cloud volumes evaluated at each format/target count (F, T) condition in simulation and experiment. (b) Position error between identified targets and their matched peak intensity locations across both point cloud volumes evaluated at each (F, T) condition in simulation and experiment. Error bars represent standard deviation and dashed line reference corresponds to the 3D diagonal span of the bounding box formed by theoretical FHWM spot dimensions.



Figure 2.10: Top:Diagram illustrating the decomposition principle employed in NIMBLE-PATCH to multiplex holograms on the SLM for the time-averaged formation of the overall ensemble of targets within the time window of interest. Targets 1-8 are optimally assigned between frames and patches to maximize overall efficiency. Bottom: Illustration of the same decomposition concept as applied to GS.



Figure 2.11: Sweep results comparing NIMBLE-PATCH to decomposed GS for excess refresh rate scaling evaluation, using single-frame GS as a benchmark. (a) Contrast and (b) computation time scaling as a function of available excess refresh rate N across SLM formats F, for two total target count scenarios: T=64 and T=256. 95% confidence intervals for the mean value of the curves are shown as shaded regions for each plot.

size and FoV can simultaneously be made equal across two different SLM formats F_1 and F_2 for a given T by choosing $N_1/N_2 = (F_2/F_1)^2$. For most cases, higher N and smaller F SLMs perform better, pointing to the benefit of utilizing a faster SLM in point-cloud targeting applications. However, as target-count-per-frame (T/N) approaches 1 (i.e., point-scanning regime) a drop in contrast relative to lower N and higher F devices is observed. This is due to the fact that partitioning is obviated when T/N = 1, eliminating the advantage of the optimized assignment process that places patches nearer to the parabolic vertex locations of their corresponding targets for improved efficiency.

2.6 Discussion and Conclusions

NIMBLE-PATCH features sub-100-ms computation times across nearly all evaluated singleframe cases. It therefore already matches the fastest existing GPU-based CGH algorithms, all without utilizing parallelization, without making concessions in volumetric point-cloud quality (e.g., severely constraining addressable depth plane count), and without the need for time-intensive and context-specific model pre-training [13, 68]. Importantly, NIMBLE-PATCH boasts a highly-parallelizable architecture where SLM pixel values are independent of one another and computed non-iteratively in a manner that is compatible with single instruction multiple data (SIMD) processing. These features allow for NIMBLE-PATCH to be scaled both to low-latency computation through the use of parallelized hardware implementations (GPU or FPGA-based) for real-time phase retrieval in closed-loop applications, and to hardware-light computation on resources like single-core CPUs or dedicated lightweight DSP pipelines implementing Eq (2.3) on a per-pixel basis. To demonstrate the scalability of NIMBLE-PATCH in even consumer-grade CPU-based environments, we tested SLM formats of F=1024,2048,4096 across target counts of 1-256, on an Apple M1 Pro CPU and 16 GB of RAM, recording compute times ranging between 100 ms and 1 s. These same SLM phase mask formats cannot be feasibly computed on CPU by CNN- or FFT-based algorithms like GS, all of which require GPU-based computation at larger formats.

Though the NIMBLE-PATCH implementation evaluated in the work was specifically designed to simultaneously exploit the full FoV allowed by the SLM pitch and maximize overall efficiency (and therefore contrast), an entire class of implementations can be envisioned using the same underlying principles. For instance, maximizing overall efficiency may lead to the deprioritization of some distant targets with inherently low efficiency, impacting spot power uniformity. Accordingly, alternative cost functions or heuristics may be employed instead of the efficiency-based linear sum assignment in order to prioritize the weakest performing spots by assigning them to their optimal patches first. Similarly, target-specific spot irradiance constraints may be imposed by calculating optimal distances to parabolic vertices for each target based on desired efficiency, then minimizing deviation from these distances instead of distance to the vertex itself during patch assignment. Control over relative spot power across point cloud volumes can further be aided by redundantly assigning a different number of patches across different frames to each target, which also offers flexibility on allowable

target count. Patch shapes may be altered to allow for some amount of interleaving as well, such that high-desirability patches can be shared across multiple targets and spot resolution can be increased (at the expense of FoV). Furthermore, NIMBLE-PATCH is capable of producing cloud volumes from a variety of primitives instead of diffraction-limited spots, including lines and polygons, since convolving a desired shape across a point cloud simply requires adding the required phase mask to each patch [12, 77]. The partitioning geometry decomposing the SLM into patches may additionally be altered to better accommodate such primitives (e.g. strip-shaped patches for line primitives). Lastly, time multiplexed operation may be mobilized to improve speckle instead of throughput via random phase offsets applied to each patch in order to average out interference between distinct targets [34, 36, 37]. The widening gap between NIMBLE-PATCH and GS computation times observed under increasing SLM pixel count in this work indicates that, at state-of-the-art SLM formats beyond 512×512 and under the same compute resources, NIMBLE-PATCH can be expected to operate >10,000x faster than lightweight iterative CGH algorithms and >100,000x faster than iterative CGH algorithms of matched performance [52, 41]. This speedup is obtained by employing a predetermined scheme to apportion the SLM's available SBP for modulation across targets, thereby circumventing the computational burden of full-frame computation for each additional target. However, we note that this considerable speed advantage is earned at the expense of patterning flexibility, as iteratively generated phase masks can make use of interleaving or shared pixels across targets. Such functionality may be more suitable for low-contrast volumes such as continuous 3D intensity distributions. Nevertheless, in the context of point-cloud patterning, the NIMBLE-PATCH approach uncovers an entirely new computation regime that offers both a path to real-time true-3D holography and scaling behavior that is compatible with emerging time multiplexing approaches enabled by ongoing SLM improvements [41, 56, 48, 57].

This chapter was a direct adaptation of the following article:

N. T. Ersaro^{*}, C. Yalcin^{*}, L. Murray, L. Kabuli, L. Waller, and R. Muller "Fast, noniterative algorithm for 3D point-cloud holography," arXiv preprint, May 2023.

Personal contribution: I built the Python-based simulation environment, ran the sweeps, parsed and plotted the simulation result data. I wrote parts of the original manuscript.

Acknowledgements: N. T. Ersaro prepared the theoretical background of NIMBLE-PATCH, performed its Python implementation, parsed the experimental results, and wrote parts of the original manuscript. L. Murray and L. Kabuli built the optical setup for experimental verification and ran the experiments. R. Muller and L. Waller reviewed the manuscript and provided advice on the technical aspects of the research. R. Muller was the principal investigator of this work.

Chapter 3

A MEMS-based Optical Scanning System for Precise, High-speed Neural Interfacing

3.1 Introduction

As discussed in Chapter 1, one of the approaches to build an optical neural interface to stimulate singular neurons across mm-scale fields of view encompassing tens of thousands of neurons is to build a scan-based system. In such systems, lateral and axial scanners form a raster scan of the entire volume by adjusting the focal point of the optical system to cover each potential target location at least once, delivering the light to selected neurons by precise timing of the laser. Raster-scan based systems have the benefit of utilizing resonant elements, minimizing hysteretic effects and temporal crosstalk, and potentially speeding up the volume-scan rate if every voxel needs to be visited once. However for most applications, targets of interest are known before the scan begins, and a random-access system would be preferable to a raster-scan based system to improve both the overall throughput, and the temporal resolution in light delivery. For the optical system to not be the bottleneck in throughput or temporal resolution, the random-access optical elements need to operate at kHz speeds, as GEVI and opsin response times reside within the 100 µs to 10 ms orders of magnitude, depending on the specific choice of the engineered molecules.

Galvanometric scanner mirrors are commonly used lateral scanners, and can achieve kHz speeds, allowing high-throughput random access operation. In contrast, state-of-the-art varifocal elements are electrically tunable lenses (ETLs) and have settling times that exceed 15 ms, severely bottlenecking the response time of the overall optical system [78]. In another commercially available technology, the liquid crystal (LC) lens, the fluidic settling behavior of the LC molecules limits the refresh rate to <500 Hz, especially for longer wavelength ranges (>800 nm) [40, 79]. Faster optical modulation techniques have also been employed in varifocal applications, but such approaches either lack the crucial capability of random-

access scanning or require impractical drivers preventing easy integration into random access all-optical interfaces. One such method is the tunable acoustic gradient index of refraction (TAG) lens, which uses standing acoustic waves in fluidic environments to modulate the local index of refraction, sweeping the focal point of the optical system across a given range [80]. While these devices operate at tens of kHz, their resonant operation prohibits dwell capability. Another method employs continuous deformable mirrors (CDM), which can achieve kHz refresh rates with dwell capability, but require drive voltages on the order of 100V or more to achieve meaningful actuation ranges [42]. This requirement complicates driver requirements, increases system size, and limits the number of independent elements in an array that can be feasibly driven. CDMs also suffer from coupled actuation between neighboring pixels, preventing utilization of phase wrapping in the applied hologram and causing non-idealities, thereby limiting focus tuning range [81]. Digital micromirror devices (DMDs) are a fast and compact alternative that perform binary amplitude modulation, which can produce configurable Fresnel zone plates for varifocal operation. However, these devices suffer from very poor optical efficiency, with <5% of the optical power input to the system making it to the focal point [50].

Another consideration in building such systems is the compactness. While the most existing methods to implement single-cell resolution optical neural interfaces use many optical elements (diffusers, lateral scanners, axial scanners, auxiliary elements for alignment of these) that are bulky and therefore are only fit to operate on animals with fixated skull under a microscope objective, an ideal optical neural interface would be portable to both enable freely-moving animal experiments, and allow for scaling to clinical applications for use in humans. This creates a push to realize the lateral and axial scanning in a single device, capable of placing spots of light arbitrarily in 3D. As discussed in Chapters 1 and 2, the SLM with piston-motion micromirror unit-elements can serve this function in an optically and computationally efficient manner when operated with the NIMBLE-PATCH algorithm. Realizing full-fledged SLM operation requires a high number of electrical interconnects as each mirror must potentially be placed at a different height to introduce the desired phase profile. In order to explore and characterize the capabilities of the piston-motion MEMS micromirror technology, we first developed a reduced degree-of-freedom SLM that can be configured as a spherical phase surface. Such an SLM can serve as the varifocal element in a scanning system while not requiring complicated driving and integration schemes like conventional SLMs. In this chapter, the specification extraction for, and the development of the drive electronics for this device will be discussed.

3.2 MEMS Mirror Actuation Requirements

The operating principles of the MEMS-based varifocal mirror are shown in Figure 3.1. We designed and fabricated the array using the MEMSCAP PolyMUMPs process with thickness modifications and custom Au lift-off post-processing for metallization. Each micromirror pixel consists of a fixed bottom electrode that, through parallel-plate capacitive transduc-

tion, actuates an electrically biased mirror body supported by two clamped-guided suspension beams. Pixel-level phase shifting is achieved as the travel path of incident light is increased by an amount that corresponds to twice the mirror actuation displacement, as depicted in Figure 3.1(e). The array is capable of introducing radially symmetric phase masks, patterning incident beams into spherical wavefronts, effectively tuning the focal point of the overall optical system.

To determine the relationship between finite actuation resolution and hologram quality, SLM performance was simulated across various array formats at a fixed pitch of 22.5 µm. In the simulation, a 4f optical system imaging a laser spot was considered with the SLM located in the Fourier plane and light intensity distribution calculated at the target volume through Fresnel propagation. A focal length of 9mm was used and observation planes were located inside a range of µ1.5mm from the focal plane. Three target light intensity distribution cases were considered: (1) steering a single spot in X, Y and Z for 3D point-scan optogenetic stimulation, (2) generation of a 3D point cloud for multi-target holographic optogenetics, and (3) generation of arbitrary mesh-based shapes for general purpose holography. Holograms corresponding to target intensity distributions were computed analytically for the singlepoint scanning case and using global Gerchberg-Saxton algorithm for the multi-point and



Figure 3.1: Structural diagram and operating principles of the micromirror-based varifocal element. (a) Wiring scheme of the annular array with 23,852 square-shaped mirrors arranged into 32 individually addressable concentric rings, capable of introducing radially symmetric phase patterns. (b) Example array configurations for tuning the focal point of an offset lens. (c) Close-up photo of the micromirror array. (d) SEM image of a single mirror. (e) Principle of operation of a piston-motion micromirror depicting translation of vertical displacement difference of mirrors to phase difference of reflecting light.



Figure 3.2: Diagram illustrating the effects of resolution in phase on the hologram quality visually and in terms of the two metrics α and η - (a) Examples of target T(x,y—z=-1.5mm) and generated G(x,y—z=-1.5mm) light intensity distribution simulations for single-point scanning, and point cloud and mesh-based approaches of hologram generation, with images shown for 2- and 8-bit actuation resolution cases. At lower resolutions, artifacts such as higher order diffraction modes (top row) or excessive speckle noise (middle and bottom rows) degrade hologram quality. (b) α and η versus resolution, normalized to an infinite phase-resolution SLM. Results show 6 bits of resolution in phase modulation is sufficient to generate highly accurate and efficient holograms for all approaches.

mesh-based cases. The resulting phase masks were then discretized and summed with random noise to account for finite actuation resolution. Target intensity pattern I(x,y,z) is specified as a binary amplitude pattern with pixel values of 0 or 1. Generated intensity pattern G(x,y,z) is computed through the simulation of light propagation through the 4f optical system with the SLM expressing discretized phase mask. To quantify the quality of the generated pattern, accuracy (α) and efficiency (η) metrics as described in Eqs. (2.8) and (2.9) were used.

 α and η were then normalized to the metric achieved by an SLM of the same array format, with infinite actuation resolution. Figure 3.2 shows target and generated images, and normalized α and η for various drive resolutions in three kinds of SLM applications. For single-point and point cloud cases, randomized targets were used across a thousand-sample Monte Carlo simulation environment. For mesh-based cases, simple shapes such as letters



Figure 3.3: Voltage vs. displacement curve for a simulated MEMS micromirror with dimensions provided in the table. Dashed lines represent process corners with $\pm 5\%$ thickness variation.

of the alphabet were considered across various depth planes. In all cases, the accuracy of the generated hologram encounters a small amount of degradation at 4 bits and saturates at 6 bits of resolution in phase modulation. Therefore, in this work we have implemented a mirror driver that can provide a 6-bit control in phase modulation. Figure 3.3 shows simulated voltage actuation curves for a sample mirror structure, quantized with 6-bits of actuation, alongside dashed lines representing process corners with 5% thickness variation of the structural layers. This displacement-actuation voltage relation is nonlinear with respect to the applied voltage for a given displacement approximated by the equation [42]:

These micromirror structures have a strong (cubic) dependence on the thickness (vertical size) of the suspension beams, while having a weaker (linear) dependence on the width (lateral size). This manifests as strong global variations and relatively weak local variations, as while lateral features are susceptible to lithographic random processes, layer thicknesses are not random processes between neighboring structures but vary wafer-to-wafer or slowly across the same wafer. A conventional solution to tackle the problem of variation is to utilize an array of discrete high-resolution linear DACs and perform calibration using look-



Figure 3.4: Simplified block diagram of the mirror driver ASIC.

up tables. Since $V(\Delta z)$ is nonlinear, a linear DAC wastes dynamic range in the region of the curve where the transduction gain is low, and hence a higher voltage LSB can be used. Furthermore, existing LCoS SLM systems span multiple PCBs including external, discrete DAC arrays to write pixel voltages, alongside peripheral digital circuits providing timing signals. This results in typical system sizes used in neuroscience applications [77] to be on the order of 8x8x6 cm3, and limits these experiments to non-portable, benchtop systems. A driver ASIC with an integrated voltage generation scheme stands to shrink the system size to aperture size of the optical device by consolidating the discrete components to an ASIC/MEMS pair, allowing for integration of SLMs into compact holography systems, such as optogenetic stimulation devices for moving animals.

3.3 Driver ASIC Implementation

To overcome both the global variations in the MEMS process, and to provide a linear digital code-to-displacement conversion, we have developed a driver ASIC that employs a reconfigurable nonlinear 6-bit DAC [82]. Electrical connection to MEMS devices can be established either through $5.4 \times 5.4 \,\mu\text{m}^2$ pad openings arranged in a 200×200 pixel array for fully independent SLM operation, or through 32 wire-bond pads for reduced degree-of-freedom MEMS arrays. To minimize power consumption while retaining the required actuation range for MEMS devices with >0.5 μ m lateral features, the ASIC was designed with 8 V drive ca-

pability. As shown in Figure 3.3, for linearly spaced 64 displacement levels, the voltage differences between adjacent codes range from 1.1 V in the lowest end to 12 mV in the highest end across process corners for a simulated MEMS device with 500 nm vertical displacement under 0-8 V drive. The drive circuit for such an actuator requires 11-bit accuracy in the higher actuation regime, while only requiring 4-bit accuracy in the lower end of the curve. This property was exploited by designing a reconfigurable nonlinear DAC that reuses its precision setting capacitors as sample & hold capacitors to save power and area compared to a linear DAC that spans the entire dynamic range. Figure 3.4 shows the simplified block diagram of the ASIC. The nonlinear DAC generates 64 voltages that correspond to linearly spaced mirror displacement levels. Mirror displacement data is transmitted via a 4 Gbps LVDS link consisting of four channels, operating at 1 Gbps/channel with 6b/8b encoding to ensure DC balance. This data is then scanned into a shift register chain to configure analog multiplexers and select the corresponding voltages to be written to each pixel's DRAM cell. Each unit pixel contains a pad opening to bond a MEMS mirror, and two capacitors that comprise two DRAM cells. 32 of these pixels are connected to output buffers to drive the internal voltages off-chip. The entire array has a refresh rate of 10 kHz, although it is possible to window only the 32 pixels driving the output buffers to achieve refresh rates up to 2 MHz.

The nonlinear 6-bit DAC is composed of two sections: a voltage reference to generate and retain the 64 analog voltage values that correspond to each level of vertical displacement for a given actuation curve, and a distributed analog multiplexer and buffer pair per row to



Figure 3.5: The schematic of the DAC, generation of nonlinearly spaced voltages that correspond to linearly spaced displacement levels of the mirrors, and measured results of cases that correspond to two possible MEMS actuation curves are shown on the right. The two cases correspond to variations on the process parameters of the micromirror structure discussed in the previous section with its nominal behavior shown in Figure 3.3

select and write the corresponding voltage to each pixel. Figure 3.5 shows the schematic of the reference voltage generator section, alongside timing diagram with the generation and retention of voltage levels for two possible nonlinear actuation curves. A capacitor bank containing 64 unit capacitors ($C_{\text{UNIT}}=2.2\text{pF}$), a current source for controlled discharge, and a reset switch are all connected to a common node. Initially, all capacitors are reset to $V_{\text{RESET}}=8\text{V}$, and then discharged through the current source ($I_{\text{BIAS}}=2\mu\text{A}$). Capacitors are sequentially disconnected from the common node to sample voltages that correspond to their respective codes through the timing of $\phi_{con,i}$ signals. Timing is controlled by a state machine and on-chip memory containing discharge times for each code (8 bits/code) that define how many periods of T_{CLK} (50ns) discharge should occur, to yield ΔT_i . The generated voltage for a given code *i* is

$$V_i = V_{i+1} - \frac{I_{\text{BIAS}} \times \Delta T_i}{C_{\text{TOT}}(i)}$$
(3.1)

where $C_{\text{TOT}}(i)$ is the total capacitance connected to the discharge node for code *i*. As capacitors are removed from the common node, discharge speeds up and precision of the generated voltage decreases. Programmability of this voltage generation scheme allows for cancellation of mirror nonlinearity as well as calibration for MEMS (e.g. beam thickness, residual stress) and CMOS (e.g. capacitance, reference current) process variations. Voltages are buffered with rail-to-rail class AB amplifiers and distributed to the rest of the ASIC, to serve as reference voltages in the DRAM write chain depicted in Figure 3.5. Due to the leakage of stored charge on the capacitors to the bulk of the switch devices, the nonlinear DAC is refreshed every 2.5 ms, keeping drift <0.5 LSB error in mirror position. With typical values of discharge current and discharge durations, refresh operation takes <200 µs. While the DAC refresh is a periodic event, discharge durations are calibrated once per MEMS device and programmed into the ASIC during startup.

The pixels for array-scale drive are laid out in a 200×200 grid at a pitch of 22.5 µm and with 5.4 μ m \times 5.4 μ m pad openings for per-pixel MEMS connection. Each pixel contains five switches and two MOM capacitors ($C_{d1,i}$ and $C_{d2,i}$, 250 fF each) that serve as analog DRAM elements. The flow of operation to update the drive voltages in the pixel array is shown in Figure 3.6. Digital select codes are transferred to the chip through the LVDS link and distributed to each row through a chain of shift registers. For each write operation, the MEMS capacitor is reset to VSS to prevent frame-to-frame hysteresis, the corresponding reference voltage is selected, the offset of the amplifier is cancelled through an auto-zero phase, and the buffered value is written to the corresponding pixel. The two DRAM capacitors in the pixel operate in a ping-pong fashion, alternating between storing value for the next frame and driving the MEMS pad. The capacitors switch roles with each new frame to provide global-shutter operation, minimizing downtime between subsequent frames and eliminating rolling shutter artifacts, which would prolong the effective settling time of the optical element. As the simulated value of the parallel plate capacitance of the mirror structure is <10fF, there is negligible charge sharing between the pixel capacitance and the actuator, which is accounted for by pre-distorting the reference voltages.



Figure 3.6: Principle of operation of the DRAM write chain with 4 phases of configuration shown. The two pixel capacitors are utilized in a ping-pong fashion, enabling global shutter operation to minimize down time between frames.



Figure 3.7: Chip micrograph with the inset showing MEMS pad openings, chip specifications and power breakdown.

3.4 Measurement Results

The IC was fabricated in TSMC's 40 nm HV CMOS technology node. The die micrograph and power consumption breakdown are shown in Figure 3.5. Measurements are divided into two sections: electrical measurements of the ASIC to verify the performance metrics of the nonlinear DAC and DRAM write chain, and optical measurements taken driving 32-channel MEMS varifocal mirror to demonstrate optical functionality and characterize precision and speed of electromechanical actuation of the ASIC-MEMS system.

Electrical Measurements

The nonlinear DAC was first characterized separately from the MEMS to verify that the electrical performance meets application specifications. Importantly, the ASIC should not cause more than 1 LSB error in displacement for any supported MEMS mirror actuation curve, including the drift caused by leakage from the DAC storage capacitors discussed in the previous section, which was budgeted 0.5 LSB, leaving another 0.5 LSB for the rest of the write chain. To determine the edge constraints, two extreme mirror actuation cases were considered: (1) a highly nonlinear voltage-displacement response such as the mirror model presented in Figure 3.3, and (2) a 0-8 V fully linear voltage-displacement response that is more pessimistic than any real actuation curve would be in the lower code regime. These two constraints are stringent on opposite ends of the actuation range. Figure 3.8 shows a comparison between the two sets of specifications, (1) indicated by magenta and (2) indicated by the green dashed lines, together with the measured post-calibration precision and maximum residual error of the nonlinear DAC for each code. Here, the precision is defined as the refresh-to-refresh standard deviation of the voltage corresponding to each code, and results from the noise of the DAC current source and amplifiers in the write chain.

The maximum residual error refers to the change that can be induced in the mean output voltage by tuning the discharge time of the given code by 1 bit, and is limited by the clock period, discharge current of the DAC and the code capacitance as described in Equation 3.1. This value represents how close a given code is guaranteed to approach an arbitrary voltage. The results show that the joint error in mirror displacement due to residual error and finite precision of the DAC is <1 LSB in mirror displacement for a wide range of possible mirror actuation profiles.



Figure 3.8: Measured precision and maximum residual error of the nonlinear DAC vs. DAC code. Two sets of constraints are also shown in dashed lines that correspond to the most stringent cases for different ends of the actuation curve. Maximum residual error of the DAC is defined as the change that is induced in mean DAC output for a given code when the code discharge duration is changed by 1 bit. Precision of the DAC is the standard deviation of a code output voltage measured refresh-to-refresh.

Optical Measurements

The 32-channel annular MEMS array was driven with the ASIC to form the varifocal system. A digital holographic microscope (DHM) was used to observe the behavior of individual mirrors inside the array. Since the MEMS array used in this work has a full-scale drive range of 32V, a -20V bias voltage was applied to the top electrode of each mirror to operate the device inside the high transduction gain region of the actuation curve.

Static measurements of two individual mirrors were performed to generate DNL and INL characteristics of the digital code-to-displacement conversion, and the results are shown in Figure 3.8(a-e). First, the mirror actuation curves were extracted using a discrete 14-bit linear DAC, and were fit on analytical curves as per equation (3). The ASIC was then programmed to implement the inverse nonlinearity of the mirror under study. To eliminate gain errors arising from the mismatch between applied reverse bias voltage and analytically fit curves, a gain calibration is performed by applying a scalar factor to all voltages in the actuation curve such that code 63 of the DAC corresponds to a 2π phase shift from code 0. For each digital input code, displacement value after full mechanical settling was recorded. The process was repeated for a mirror from a different MEMS die. Maximum DNL and INL values recorded across all codes and both mirrors were 0.21 LSB and 1.14 LSB, respectively. Main source of disparity between static behavior of the mirrors were determined to be beam thickness, residual stress, and resting gap height.

Dynamic measurements were made with the stroboscopic mode of the DHM and three



Figure 3.9: Static and dynamic measurements of the ASIC-MEMS system performed under a DHM. (a) Measured displacement vs. voltage behavior for two mirrors. (b) Measured transfer curve of the nonlinear DAC post-calibration for the two mirrors. (c) Measured displacement vs DAC code behavior for the two mirrors. (d-e) DNL and INL of displacement vs. DAC code. (f) Dynamic behavior of three mirrors while the ASIC was configured to switch between the two extreme ends of the actuation curve at 2 kHz.

mirrors on the same die were simultaneously observed while being driven between two displacement values. The results are shown in Figure 3.8(f) and the maximum 10-90% rise/fall times for these mirrors were measured to be 80 µs and 82 µs, respectively.

To demonstrate optical utility of the ASIC-MEMS system, a 4f imaging system was constructed to image a laser point, with the annular MEMS array located at the Fourier plane. A CMOS camera on an automated z-stage was used to capture images formed in the target volume for various configurations of the tunable lens. Figure 3.10 shows the diagram of the optical setup and images taken at 4 depths for 4 curvature configurations of the varifocal mirror. While deviations from aimed focus depths were observed due to aberrations and imperfect alignment of the optical system, these are deterministic effects that can be corrected by a one-time lookup calibration of aimed depths vs. observed focal plane depths. The volumetric efficiency of the system was quantified as the ratio of the energy located inside the spot full width at half maximum (FWHM) to the total energy located in the field of view and was found to be 38% at the focal plane of the lens. The spot FWHM was measured to be 10 µm in X and Y directions, and 900 µm in the axial direction with a full-scale continuous tuning range of ± 10 mm when used with a f=100 mm lens, spanning 22 fully resolvable depth planes at refresh rates greater than 12 kHz. Through the demagnification of the imaged spot, this device can address 10 µm-sized targets across an axial range of 220 µm.

3.5 Summary and Discussion

In this work, we've built a varifocal mirror system for high-speed, random-access 3D pointscanning systems for optogenetic stimulation. The system is comprised of an annular array of piston-motion MEMS mirrors wired into 32 concentric rings, and a driver ASIC. The ASIC features a reconfigurable nonlinear DAC that provides a linear code-to-displacement conversion by correcting the inherent nonlinearity of electrostatic actuation as well as global MEMS process variations. The system can address 22 distinct depth planes with refresh rates >12 kHz. Table 1 shows a comparison of this system with similar systems in the literature, with major challenges of realizing integrated, high-speed 3D point-scanning systems using these technologies highlighted in red. Our system's refresh rate exceeds the two most common varifocal elements (ETLs and LC lenses) by a factor of >36x, possesses random-access and dwelling capability lacking in resonant devices such as TAG lenses, and requires only 8V drive allowing scalability to large array formats. Compared to DMD-based approaches, this work offers 10x higher volumetric efficiency and 10x lower power consumption, using 33x fewer actuators.

An array of micromirrors with pixel-level independent actuation through the ASIC could unify lateral scanning and varifocal operation in a single chip-scale device, significantly miniaturizing 3D point scan systems. For example, a 10 kHz, 200x200 pixel SLM that can be supported by the ASIC in this work could target hundreds of neurons in a 500x500x500 µm3 volume of brain within 1 ms, a relevant timescale for neural signaling that corresponds to



Figure 3.10: (a) Optical measurement setup for the tunable lens system formed by the ASIC and 32-channel MEMS array. During the measurements, ASIC was programmed to implement the inverse nonlinearity of the mean actuation curve for the entire array. Effect of local mismatches are mitigated by the highly redundant nature of the radially symmetric phase masks being used [23]. (b) Z-stack measurements relative to background illumination for four target focus depth configurations. (c) Photographs of the optical measurement setup.

the duration of a single action potential. Such a high speed SLM can also be extended to applications outside of neuroscience, such as 3D holographic near-eye displays for AR/VR systems by overcoming two attributes that are limited by the slow refresh rates of the LCoS SLMS used in current systems. A higher refresh rate allow time multiplexing between three color domains to enable full-color holographic displays using a single SLM. Simultaneously, the time averaging capability enabled by the excess frame rate can be utilized through ex-

isting speckle noise reduction techniques to improve hologram accuracy and overall image quality.

	[9]	[11]	[12]	[13]	[15]	This work
Actuator Type Settling Time [*] Dwelling Capability Required Driver Voltage Volumetric Efficiency ^{**} Number of Actuators Number of Independent Channels	ETL > 15 ms Yes < 20V N/A†	LC Lens > 3 ms Yes 10 - 100V >90%	TAG Lens $< 15\mu s$ No < 20V N/A†	CDM < $100\mu s$ Yes 250V - 1k-3k 1k-3k 1k-3k	DMD 45µs Yes 1V 3.7% 786k 786k	Piston Mirror 82 μs Yes 8V 38% 24k 32
IC Power Consumption Required Datarate	_	_	_	_	2-4.4 W 25.6 Gbps	308 mW 3 Mbps

	Table 3.1 :	Varifocal	Element	Comparison	Table
--	---------------	-----------	---------	------------	-------

* 10% to 90% settling time

** Measured at the focal plane

† Not a diffractive device

This chapter was a direct adaptation of the following article:

C. Yalcin, N. T. Ersaro, M. M. Ghanbari, G. Bocchetti, S. F. Alamouti, N. Antipa, D. Lopez, N. C. Pégard, L. Waller, and R. Muller, "A MEMS-based optical scanning system for precise, high-speed neural interfacing," IEEE Journal of Solid-State Circuits (JSSC), Jun. 2022.

Personal contribution: I designed, implemented, and tested the ASIC, contributed to specifications and testing of the MEMS, and performed the optical measurements of MEMS-ASIC joint system. I wrote the original draft of the manuscript.

Acknowledgements: N. T. Ersaro designed the MEMS array and helped with the optical setup and DHM measurements. N. T. Ersaro and N. Antipa constructed the optical system for measurements. M. M. Ghanbari helped implement the ASIC, and provided drawings for the figures. G. Bocchetti and S.F. Alamouti helped implement the ASIC. R. Muller, D. Lopez, N. C. Pégard, L. Waller reviewed the manuscript and provided advice on the technical aspects of the research. R. Muller was the principal investigator of this work.

Chapter 4

A 32×32, 29-V Driver ASIC for High-Parasitic Micromirror Arrays

4.1 Introduction

The varifocal element developed in the work presented in Chapter 3 allowed the characterization of piston-motion MEMS micromirror structures. The issues identified in the firstgeneration devices were: (1) double-clamped structure led to tilting under electrostatic actuation due to an unconstrained degree of freedom, (2) pixel-to-pixel local variations were higher than expected, (3) negative biasing of the top electrodes of the array exacerbated the local variations, and (4) 22.5 µm pixel pitch of the ASIC was infeasible for the mirror structures, both in terms of the mechanical behavior and accessible chip-to-chip integration options, Addressing these problems required the development of a second generation of micromirror devices and the driver ASIC. In this chapter, we present the structure of the second-generation MEMS mirrors and the design, implementation, and verification of a high-voltage micromirror driver ASIC tailored for these devices and their high-parasitic integration scheme.

4.2 Second-Generation Mirror Drive Requirements

The specifications imposed by the second-generation MEMS mirrors upon the driver ASIC are discussed in two categories: structure-driven requirements that arise from the behavior of a singular mirror structure, and integration scheme-driven requirements that arise from the parasitics of the interposer routing and pixel co-wiring.

Structure-driven Requirements

The second-generation structure's stack up and actuation behavior are shown in Figure 4.1. The mirror comprises doped polysilicon structural elements, isolating nitride layers, and a

gold finish for high reflectivity. Nitride 0 forms an isolating barrier between the substrate and the mirror structure. Poly X allows for high-density fanout of mirror structures without sacrificing fill factor or electromechanical performance to the routing. Nitride 1 forms an isolating barrier between Poly X and Poly 0. Poly 0 forms the bottom electrode and provides shielding between neighboring electrodes. Poly 1 forms the top electrode and the four support beams of the piston structure. The gap between Poly 0 and Poly 1 is adjusted to 2 µm to allow for ≤ 520 nm vertical displacement necessary for 2π modulation of wavelengths below 1040 nm without running into the pull-in phenomenon while performing voltage-mode drive. Poly 2 forms the main body of the mirror structure and provides structural robustness against curling due to the stress mismatch between the polysilicon and the reflective gold layer forming the mirror. Perforation holes are placed throughout the structure, allowing for the reduction of squeeze film damping.



Figure 4.1: The structure, static response, and dynamic response of the second-generation mirror structures - (a) Second-generation mirror structure with labeled layers and thicknesses. Two Poly 1 structure variants are shown: the standard structure and the beam-extended low-stiffness version. (b) Static actuation behavior of the two variations with process mean and corners. (c) Dynamic response of a process-extreme 38% beam-extended variant to a step voltage. 5% settling region is shown in red dashed lines.

As shown in Figure 4.1(a), Poly 1 support beams can be extended into the mirror structure to reduce the spring constant at the cost of reduced top electrode area and perforation hole count. Two variants of the unit mirror have been designed and taped out: a standard beam length structure labeled as the 0% extension variant, and a low-spring-constant structure labeled as the 38% extension variant to indicate the additional beam length compared to the baseline structure. The two variants were designed to offer a choice of trade-off between drive voltage and step-response settling times, the 0% extension variant requires higher drive voltages but settles faster, while the 38% extension variant can be driven with a lower voltage swing, at the cost of a slower settling. However, as will be discussed below, it is possible to speed up the settling of these structures by utilizing a multi-step drive instead of a single-step drive.

The major changes to the mirror structure from the first generation can be summarized as (1) the addition of two more support arms to prevent tilting and to provide additional robustness to process variations such as mask misalignment, (2) an increase in mirror size from 48 µm to 70 µm in order to preserve similar actuation voltage profiles to first-generation devices (<30 V), (3) extension of support beams into the top electrode body to lower the required actuation voltage range even further, (4) introduction of perforation holes to decrease the additional damping due to larger mirror body causing an increased settling time, (5) a buried polysilicon layer to allow electrical fanout of individual mirrors for single-pixel addressability without sacrificing fill factor. These changes have a significant impact on both the static and dynamic responses of the mirror. From the static aspect, the ASIC needs to be able to provide a voltage resolution to achieve 6-bit linearity in vertical displacement of structures with widely varying transduction characteristics. From the dynamic aspect, the increased mirror body size and especially the reduced spring stiffness of the 38% structures result in settling times as long as 750 µs, despite the perforations for damping reduction.

Figure 4.2 shows the static transduction curve and adjacent-code step sizes for the 6bit actuation of the process-mean 0% and 38% extension variants. The second-generation driver ASIC needs to support a >27.5 V full-scale output voltage range for the 0% extension structure, while still maintaining <33 mV voltage resolution to generate small enough steps for 6-bit LSB in displacement for the 38% extension structure. This corresponds to >10 bits of dynamic range in the output voltage generation pipeline of the ASIC.

The slow step-responses of the structures can be sped up using the two-step drive technique [83]. In this technique, an initial step voltage (V_{step1}) is applied for a short duration (T_{step1}) until the dynamics of the mirror's mass-damper-spring system allow for a shorter settling time when switched to the final drive voltage (V_{step2}) . For single-pole or overdamped systems, this scheme is relatively simple as any amount of overdrive $(V_{step1} > V_{step2})$ helps the system settle faster, so long as the target is not overshot beyond the settling tolerance window. Figure 4.3 shows the voltage profile of the two-step drive technique, as well as the settling behavior of the 38% extended mirror structure versus varying V_{step1} and T_{step1} parameters. In the second-generation ASIC, this scheme simply corresponds to loading a V_{step1} frame before switching to a V_{step2} frame for the entire array, with every pixel sharing the same T_{step1} . The array then maintains the V_{step2} mask during the laser exposure.



Figure 4.2: Simulated voltage vs. displacement curve for the two variants of the second-generation micromirror structures, with annotated maximum voltage requirements and smallest step sizes.



Figure 4.3: Simulated two-step drive profiles of the 38% extended structures - (a) Generalized voltage profile of the two-step drive scheme with the three parameters shown. (b) Dynamic response of a process extreme 38% extended structure for different V_{step1} values for $T_{step1}=100$ µs and $V_{step2}=19$ V. (c) Dynamic response of a process extreme 38% extended structure for different T_{step1} values for $T_{step1}=22$ V and $V_{step2}=19$ V.

Integration scheme-driven Requirements

Scaling the singular mirror structure to an array of individually addressable phase-shifting elements requires dedicated electrical connections between the bottom plate of each mirror and its driver electronics. While the first-generation devices demonstrated co-wired array capability through bond wires, this approach is limited to 10^{1} - 10^{2} orders of magnitude due to bond wire density limitations and system size considerations. Scaling to 10^{3} mirrors and beyond requires high-density interposer-based (2.5-D) or through-silicon-via-based (TSV-based, 3-D) integration schemes.

In this work, we have utilized the MEMS die as the interposer board and the polysilicon and gold layers of the process as interconnect routing. Figure 4.4 shows two variants of MEMS dies: the improved annular mirror array structure with 64 independently addressable rings using the second-generation mirror structure and a 32×32 mirror array capable of true SLM operation. In this scheme, the driver electronics ASIC discussed later in this Chapter is flip-chip bonded onto the MEMS die. In the case of the annular array shown in Figure 4.4(a), the electrical trace of each ring is routed from the ASIC to the periphery of the annular array through a stack of each polysilicon layer and the gold layer to maintain low resistance. To preserve the fill factor of mirrors, the traces are then reduced to only PolyX when propagating this signal within the ring. This routing scheme results in a total series parasitic resistance of $R_{\rm trace}=25$ k Ω for the innermost ring and 92 k Ω for the outermost ring, and total parasitic capacitance of $C_{\rm trace}=15$ pF for the innermost ring and 1 nF for the outermost ring.

In the case of the 32×32 SLM integration scheme shown in Figure 4.4(b), a minimum design rule (2 µm trace width and spacing) interconnect routing was designed. These traces utilize only Poly X throughout the entire path to prevent conductive debris from causing electrical shorts through exposed layers over prolonged use. The combination of poor conductivity of Poly X and reduced trace width results in total series parasitic resistance values of between $R_{\text{trace}}=40 \text{ k}\Omega$ to 400 k Ω , and total parasitic capacitance values of between $C_{\text{trace}}=1 \text{ pF}$ to 25 pF.

These two integration schemes impose a wide range of load profiles under which the ASIC must provide a high-speed drive of electrical signals. Particularly, each trace must be charged and discharged with a certain amount of maximum current depending on the total parasitic capacitance and the desired refresh rate. This requires the ASIC to support all ends of the parasitic load design space, which results in inefficiency in the design of the driver pixel if each individual pixel is specified to handle the entire range. Instead, the ASIC was designed with the observation that under the highest capacitance (and hence, the highest maximum current) specifications, the array possesses fewer degrees of freedom (i.e., 64 rings instead of 1024 singular traces), and even each trace is not identical in its parasitic load to one another. The proposed solution is to build an ASIC that allows the capability to combine the drive strengths of multiple pixels by electrically connecting them on the MEMS die depending on the drive requirements of each particular trace.



Figure 4.4: MEMS die layouts of two micromirror array variants: (a) An annular array comprising $\sim 50,000$ micromirrors electrically grouped to 64 concentric rings for spherical/parabolic phase profiles. (b) A square-shaped array comprising 1024 individually controlled micromirrors arranged in a 32×32 format, for low-target-count point-scanning holography.

Another constraint that the wide range of load parasitics impose on the ASIC is that the low-pass filtering effect of $R_{\text{trace}}C_{\text{trace}}$ ends up being a significant portion of the total allotted settling time for the whole system for 10 kHz operation. In fact, the higher end of the parasitics for the annular array itself causes an electrical low-pass filtering effect with a time constant of ~100 µs preceding the mechanical response of the mirror. To alleviate this problem, a two-step drive method can be employed for the electrical settling, similar to the approach taken to mitigate the mechanical settling. However, this implies an increased refresh rate for the ASIC, as the 100 µs settling needs to encompass both the electrical and mechanical settling after the switch from one phase mask to the next is initiated. Therefore, the ASIC was designed to support up to 20 kHz refresh capability for high-parasitic-load use cases. This way, the combined electromechanical transfer function of a trace can be treated with a 50 µs pulse width two-step drive within each 100 µs period.

4.3 Design and Implementation of the ASIC

We have developed the second-generation ASIC with the goal of providing 0-29 V output swing, >10-bit linearity, and 20 kHz global-shutter refresh rate drive to up to 1024 traces. The block diagram of the ASIC is shown in Figure 4.5. The ASIC comprises a 32×32 pixel array, 8 12-bit resistive ladder DACs (R-DACs), a digital controller for configuration and timing signal generation, and a bandgap reference-based bias generation block. 12-bit resistive ladder linear R-DACs were chosen to replace the 6-bit nonlinear DAC in the firstgeneration ASIC for the sake of flexibility in driving different structures of actuators and to have the capability to address local variations. Pixels are spaced with 150 µm pitch, each pixel having a pad opening for flip-chip integration with an individually programmable output between 0-29 V. The pixel array is segmented into 8 panels, each comprising 4 rows and 32 columns of 150-µm pitch pixels. Pixels within a panel share a single R-DAC, which writes the 0-3.2 V analog data representing the output value desired from each pixel. The digital data for the DACs are transmitted through a 6-bit parallel CMOS interface operated with a 50 MHz clock. Configuration data for the chip is transmitted through a 4-wire SPI, similar to the first-generation ASIC.

The circuit diagram of the pixel is shown in Figure 4.6(a). The second-generation pixel was specifically designed to simultaneously be no-load stable and settle within 10 µs with load capacitances of up to 25 pF without significant changes to the DC characteristics of the output. Each pixel comprises two major blocks: an analog DRAM to store the desired output information, and a level-shifting driver to translate the 0-3.2 V input to 0-29 V output. The level-shifting driver comprises a Class-D output stage formed by M0 and M1 LDMOS devices and series resistances, with the output voltage regulated through a feedback loop keeping the feedback signal V_{fb} between the two analog inputs of the circuit, V_{th,p} and V_{th,n} (loaded such that V_{th,p} < V_{th,n}). V_{fb} is generated from the output through a resistive and capacitive voltage divider with a gain of $\frac{1}{9}$ to map the 0-29 V output V_{out} to 0-3.2 V. The comparators X0 and X1 enable the pull-down transistor M0 or the pull-up transistor

CHAPTER 4. A 32×32, 29-V DRIVER ASIC FOR HIGH-PARASITIC MICROMIRROR ARRAYS 58



Figure 4.5: Simplified block diagram of the second-generation ASIC.

M1 to lower or raise the output voltage if $V_{\rm fb}$ is above $V_{\rm th,n}$ or below $V_{\rm th,p}$, respectively. The resistive feedback is necessary to provide DC operation when maintaining the same voltage at the output for prolonged periods, while the capacitive feedback provides a highfrequency path for the signal, preventing output overshoot during transitions. The resistive network continually discharges the output however, and at the steady state for a given pair of input voltages, $V_{\rm fb}$ eventually converges to a sawtooth waveform around $V_{\rm th,p}$ with M1 compensating the lost charge when X1 detects that $V_{\rm fb}$ fell below $V_{\rm th,p}$. The operation of the circuit is shown in Figure 4.6(b-c) for two load trace parasitics, which are modeled as a distributed RC network with 4 π -model segments. $V_{\rm mirror}$ is the voltage across the mirror capacitance after the low-pass filtering from the load trace.



Figure 4.6: (a) Schematic diagram of the pixel circuitry. (b-c) Simulation waveforms of the pixel output V_{out} , the mirror voltage after the load trace V_{mirror} , level-shifting driver input voltages $V_{th,p}$ and $V_{th,n}$, and the feedback voltage V_{fb} , for two extremes of load profiles for a single pixel. Glitches on $V_{th,p}$ and $V_{th,n}$ are due to the kickback from the StrongARM comparators, last <2 ns and do not have an impact on the operation of the circuit.
The pixel has two features that are enabled during transitions to speed up the operation. The high-current mode (HCM) feature utilizes M2 and M3 to lower the pull-up and pull-down resistances to increase the maximum current that can be delivered to the load, at the cost of stability especially for low-capacitance or high-series-resistance loads. In such a scenario, M0 and M1 can enter a fight state as V_{fb} can swing from below $V_{th,p}$ to above $V_{th,n}$ in a single comparator cycle (and vice versa) at low comparator clock frequencies. The fast-feedback (FFB) feature lowers the low-frequency impedance of the feedback network through M4 and M5 to alleviate gross mismatches between the resistive and capacitive voltage dividers. HCM and FFB are both globally configured modes and are enabled for the first 10 µs of a given frame by default.

The amplitude and frequency of the "ripple" at the output due to the sawtooth-like operation depend on the target V_{out} . A high V_{out} increases the discharge current through the feedback resistor, speeding up the decay of V_{out} towards $V_{th,p}$. A high V_{out} also decreases the pull-up current, and hence the charge delivered to the output during the cycle that M1 is on, reducing the overcharge above $V_{th,p}$. The combination of these effects results in a higher-frequency, lower-amplitude ripple at high V_{out} values. This is desirable, as higher V_{out} regions of the transduction curve shown in Figure 4.2 are more sensitive to non-idealities and voltage inaccuracies. However, even at relatively low V_{out} values (such as the scenario shown in Figure 4.6(b-c) with $V_{out} = \sim 6.2 \text{ V}$), the maximum ripple period is on the order of 1 µs, which is heavily filtered by the slow mechanical response of the micromirror.

X0 and X1 are implemented as StrongARM comparators preceded by PMOS-input, diode-loaded differential preamplifiers to alleviate the kickback effect and to provide a constant input common mode across the full range of $V_{th,p}$ and $V_{th,n}$ signals. The 0-5V to 24-29V logic level shifter is implemented as a modified DMOS level shifter, referred to as the fast-operation topology in [84]. Feedback resistors are built from high-resistivity-implant unsalicided polysilicon resistors. One issue with polysilicon resistors in high-voltage settings is the limited polysilicon-to-bulk voltage tolerance of the oxide. This issue was addressed by constructing the feedback resistor in segments situated above dedicated high-voltage Nwells biased by the midpoint of the resistor. This technique occupies a much smaller area compared to diffusion or high-voltage N-well type resistors for the same resistance.

The analog DRAM portion of the ASIC is very similar to the DRAM structure described in Chapter 3. There are two DRAM cells (each with two capacitors, C_{d0} and C_{d1}) per pixel in this architecture, one for $V_{th,p}$ and one for $V_{th,n}$. As these voltages determine the output voltage, global shutter operation is provided by utilizing C_{d0} and C_{d1} capacitors in a pingpong operation, identical to the operation described in Figure 3.6. The major difference in the DRAM write chain from the first-generation ASIC is that each pixel requires two analog voltage writes per frame instead of one, which are written sequentially to each pixel. While this doubles the analog bandwidth of the write chain, it has no implication on the data throughput into the chip, as the offset between $V_{th,p}$ and $V_{th,n}$ is digitally programmed into the ASIC.

The R-DACs are implemented as segmented resistive ladder DACs with inverted MSB and LSB ladders [85] operated 0-0.8 V, followed by a programmable-gain non-inverting resistive-

feedback amplifier with a nominal gain of 4, mapping the DAC range to 0-3.2 V. For this architecture, the size of the MSB step is desired to be as small as possible to remove the residual distortion from the LSB switches seeing a different V_{GS} when the code changes [85]. A 7-bit MSB/5-bit LSB split was chosen as a reasonable area, power and complexity trade-off. Each DAC writes the $V_{th,p}$ and $V_{th,n}$ values to all 128 pixels in its panel within the frame time of 50 µs, operating at >5 MS/s at 10-bit settling.

4.4 Measurement Results

The IC was fabricated in a TSMC 180 nm BCD technology node. The die micrograph, chip specifications, and the power breakdown of the ASIC are shown in Figure 4.7.

Dynamic measurements of the ASIC were made by configuring the output to provide a square wave between two V_{out} values (~6.25 V and ~24.75 V), chosen to compare with the simulation results shown in Figure 4.6. The output of a pixel with only the PCB trace routing estimated to be around 6 pF as the load C_{trace} is shown in Figure 4.8. The results show the fast settling behavior and the steady state sawtooth waveform.

As discussed in Section 2 of this chapter, an efficient way to improve the load drive capability of the ASIC for reduced-degree-of-freedom SLMs that combine hundreds of mirrors in singular traces for co-wired operation is to combine the drive strength of multiple pixels. Figure 4.9 shows the measurement results with the pixel output driving a capacitive load of 100 pF. Under standard drive settings, a single pixel is not capable of driving this load within 50 µs, and an incomplete settling is observed. By simply shorting the output of 4 pixels, it is possible to achieve fast settling for high capacitance loads. The dead zone formed by the gap between $V_{th,p}$ and $V_{th,n}$ forms a protection against fight between pixels due to local mismatches introducing offsets in comparators and the feedback factor.



Figure 4.7: A microphotograph of the second-generation ASIC, the summary of the chip specifications, and the power breakdown of major power consumers on the ASIC.



 $C_{trace} \sim 6 \text{ pF}, R_{trace} \sim 0 \Omega$

Figure 4.8: Measurement results from one of the ASIC pixels driving a square wave between two V_{out} values (~6.25 V and ~24.75 V), showing good agreement between simulations and measurements in terms of ripple amplitude and overall circuit behavior.

Static measurements of the pixel outputs were done by sweeping the input code full-scale and averaging 100 samples per output code, with the results shown in Figure 4.10(a). Fullscale INL and DNL behaviors show reduced linearity for lower voltages, which is by design due to the PMOS output stage of the R-DACs as lower voltages do not have stringent drive requirements as discussed in Figure 4.2 as well as Chapter 3. High linearity is desired particularly at the highest end of a 16-V sub-range of the transduction curve, as this corresponds to the V_{step2} of the 38% extended structure, which has the most stringent voltage resolution requirements. Figure 4.10(b) shows such a sub-range of the transduction curve. The required >10-bit resolution in voltage steps corresponds to <2 LSB DNL for the 12-bit drive chain, which is achieved across the sub-range. The frame-to-frame standard deviation of the mean output voltage is measured across 1000 frames as <12 mV for all codes across this range, corresponding to a <0.3 LSB noise in displacement in the highest-transduction-gain-section of the actuation curve of the 38% extended structure.

The second-generation micromirror structures were in fabrication during the writing of



Figure 4.9: Measurement results for a C_{load} of 100 pF with (top) 1 pixel is used to drive, showing incomplete settling, and (bottom) 4 pixel outputs are shorted to combine their drive strengths, showing complete settling. Close-ups of the drive profile are also shown, displaying the lower ripple due to increased filtering of the capacitance.



Figure 4.10: Measured transduction, INL, and DNL curves of a pixel output versus digital input code for (top) a 26-V full-scale output configuration, and (bottom) a 16.4-V sub-range in this configuration with a high-linearity to generate V_{step2} for the 38% extension structure, showing <2 LSB DNL for all codes but one sparkle code, corresponding to >10 bits in linearity.

this dissertation. To demonstrate the co-operability of the ASIC and the new MEMS structures, the output of an ASIC pixel was recorded via a Keysight CX3322A Current Waveform Analyzer and applied as the actuation voltage to a micromirror model in MATLAB. The voltage and the resulting actuation profiles are shown in Figure 4.11. As expected, single-step drive results in incomplete settling while two-step drive is able to achieve <100 µs settling for the low-stiffness structures.



Figure 4.11: Response of a MATLAB-based model of the 38% extended structure to the measured output of the pixel to (a) single-step drive and (b) two-step drive voltage profiles. 6-bit settling limits for the chosen displacement values are shown in red.

4.5 Summary and Discussion

We present a second-generation micromirror device design, and the design, implementation, and verification of a driver ASIC tailored for the high-parasitic, high-voltage-swing drive characteristics of these devices. The new micromirror structure aims to improve upon the first-generation devices through increased fill factor, greatly reduced tilting behavior, and individual pixel addressability. Electrical drive requirements of two variants, 0% extended and 38% extended, and their implication on ASIC specifications, are presented. The integration scheme of the CMOS and MEMS dies is considered to further extract ASIC specifications.

The second-generation ASIC is designed to support 0-29V, 20 kHz drive of 1024 traces, each having up to 25 pF of parasitic capacitance. The output of each pixel is controlled via 12-bit resistive-ladder DACs with >10 bits of linearity in the sub-range of interest for >6bit actuation in vertical displacement of the mirrors. Electrical measurements of the ASIC demonstrate both the close matching of the individual pixels to the simulated behavior and the capability to combine the drive strengths of multiple pixels by simply shorting the

outputs for reduced-degree-of-freedom SLMs with hundreds of mirrors co-wired into singular electrical traces.

The devices developed through this work aim to build a development platform that demonstrates the scalability of piston-motion micromirror-based phase modulation to thousands of individually actuatable elements without the need for concessions such as binary (in the case of DMDs) or highly-nonlinear actuation (in the case of PLMs) that greatly impact optical performance. The array formats designed in this work can be applied to point cloud holography across fields of views of hundreds of µm with sub-10-µm spot sizes, allowing for random-access optogenetic photostimulation or fluorescence imaging of neurons at >20 kHz refresh rates. Such capability, which cannot be achieved in existing systems, may pave the way to closed-loop all-optical neural interfaces capable of interrogating and stimulating thousands of neurons in sub-action potential timescales.

Personal contribution: I designed, implemented, and characterized the secondgeneration ASIC, and contributed to the specification and structural design of the secondgeneration MEMS structure.

Acknowledgements: Nathan Tessema Ersaro designed the structure, stack-up and the physical layout of the MEMS array. Liz Murray, Aviral Pandey and Ashwin Rammohan helped with the schematic design, layout design, and design verification of the ASIC. Liz Murray helped with the bring-up and characterization of the ASIC. Rikky Muller was the principal investigator of this work.

Chapter 5 Conclusion

In this work, we presented a set of approaches to solve the two bottlenecks limiting the temporal capabilities of 3-D point cloud holography systems: computation-heavy CGH algorithms and slow-settling SLM unit elements. We present a lightweight, non-iterative CGH algorithm NIMBLE-PATCH, specifically tailored for 3-D point cloud holography, and two generations of MEMS-based SLM devices that improve upon the settling times of the industry standard LCoS SLMs by a factor of >30x. We believe this work can greatly extend the temporal capabilities of holographic 3-D point cloud holography systems and specifically in the context of all-optical neural interfaces enable closed-loop modulation of neurons in sub-action potential timescales, a long-standing holy grail of neuroscience.

NIMBLE-PATCH divides the available SBP of an SLM between targets, optimally assigns regions on the SLM to achieve maximum volumetric efficiency, and allows for non-iterative computation of the regional phase mask. Resulting holograms are minimally diffracted onto higher diffraction orders without requiring bulky computation, matching the quality of industry-standard algorithms, at 10²-10⁶ times faster computation times. NIMBLE-PATCH makes no concessions on depth plane counts or spot center placement resolution, both of which are useful practical features in applications such as superresolution microscopy. Furthermore, NIMBLE-PATCH computation requires no forward propagation of the hologram, neural network training or even storing the full hologram, allowing for scalable computability using commercial-grade CPUs as demonstrated with array formats as high as 16 Megapixels.

The MEMS micromirror array-based devices developed in this work aim to extract the electrical drive requirements of piston-motion micromirror-based SLMs, demonstrate the utility of these devices in optical settings, and provide a pathway to building a full-fledged SLM allowing for individual element control in array sizes up to thousands of micromirrors. The first-generation devices were built with the goal of characterizing the micromirror structures and achieving a reduced-degree-of-freedom SLM operation that allows for high-speed, dwell-capable axial steering of focused points. To this end, we developed an annular array comprising >23000 micromirrors electrically grouped into 32 concentric rings that can introduce radially symmetric phase profiles such as spherical or parabolic curvatures for varifocal operation. We co-developed a driver ASIC that aims to shrink bulky SLM PCB electronics

by integrating the DAC into the driver. Through a novel DAC architecture, we were able to build a reconfigurable voltage DAC that can be programmed to implement the inverse non-linearity of the mirror drive and account for global process variations of MEMS fabrication, resulting in an area- and power-efficient method to provide the mirror voltages. The device formed by the board-level integration of the ASIC and MEMS dies was used as a varifocal element and achieved a distinctly resolvable depth plane count of >22 at refresh rates >12 kHz.

The second-generation structure of the piston-motion SLMs was developed after identifying the limitations of the first-generation structure. The changes in both the unit element structure and the envisioned 2.5-D integration scheme required the development of a new ASIC capable of providing fast drive to high-parasitic traces. The developed ASIC provides a closed-loop push-pull-regulated output at each pixel and allows for the drive strengths of multiple pixels to be combined without inter-pixel fight. While the second-generation micromirrors were still under fabrication at the time of the writing of this dissertation, recorded waveforms from the ASIC were used to actuate a simulated model of the second-generation mirror structure, achieving <100 μ s settling times with 6-bit resolution in displacement via two-step drive.

The most important effort in the continuation of this work on piston-motion micromirrorbased high-speed SLMs is the scaling of actuators, both in terms of count and size. While the second-generation structure resolves some issues regarding tilting and uniformity, its large pixel pitches result in impractical array sizes when tiled to Megapixel formats. Furthermore, the presented integration scheme with buried polysilicon interconnects is extremely cumbersome and power-inefficient to scale beyond arrays with tens of thousands of elements, even with multiple layers of buried routing. The latter problem can be solved with existing 3-D integration technologies such as through-silicon-vias (TSVs), which also address the highparasitic nature of 2.5-D integration. However, small pixel pitches required to co-scale the array formats within practical aperture sizes require mirror and TSV pitches down to 10 μ m. Reduced mirror pitch brings structural issues such as exaggerated non-uniformity and excessive drive voltage requirements due to the high stiffness of short support arms. Driver electronics capable of exploiting the reduced parasitics for closed-loop drive of the micromirrors through capacitive sensing may resolve the uniformity issues, while modifications to the mirror structure such as vertical comb drive may alleviate excessive voltage requirements while easing requirements on capacitive sensing. Finally, as the ever-increasing speed and count of unit elements put a burden on the links between the DSP elements and the SLM, a hardware implementation of NIMBLE-PATCH can greatly reduce the required data throughput to the ASIC while allowing for efficient (and even locally in-pixel) computation of the phase mask without loss in optical quality.

Bibliography

- Erdem Sahin, Elena Stoykova, Jani Mäkinen, and Atanas Gotchev.
 "Computer-Generated Holograms for 3D Imaging: A Survey". In: ACM Computing Surveys 53.2 (2021), 32:1–32:35.
- [2] Bahram Javidi, Artur Carnicer, Arun Anand, George Barbastathis, Wen Chen, Pietro Ferraro, J. W. Goodman, Ryoichi Horisaki, Kedar Khare, Malgorzata Kujawinska, Rainer A. Leitgeb, Pierre Marquet, Takanori Nomura, Aydogan Ozcan, YongKeun Park, Giancarlo Pedrini, Pascal Picart, Joseph Rosen, Genaro Saavedra, Natan T. Shaked, Adrian Stern, Enrique Tajahuerce, Lei Tian, Gordon Wetzstein, and Masahiro Yamaguchi. "Roadmap on digital holography". In: *Optics Express* 29.22 (2021), pp. 35078–35118.
- [3] Wolfgang Osten, Ahmad Faridian, Peng Gao, Klaus Körner, Dinesh Naik, Giancarlo Pedrini, Alok Kumar Singh, Mitsuo Takeda, and Marc Wilke. "Recent advances in digital holography". In: *Applied Optics* 53.27 (2014), G44–G63.
- [4] Ulf Schnars, Claas Falldorf, John Watson, and Werner Jüptner. *Digital Holography* and Wavefront Sensing. Springer, 2015.
- [5] David C. Brotherton-Ratcliffe, Stanislovas J. Zacharovas, Ramunas J. Bakanas, Julius Pileckas, Andrej Nikolskij, and Jevgenij Kuchin. "Digital holographic printing using pulsed RGB lasers". In: *Optical Engineering* 50.9 (2011).
- [6] James H Marshel, Yoon Seok Kim, Timothy A Machado, Sean Quirin, Brandon Benson, Jonathan Kadmon, Cephra Raja, Adelaida Chibukhchyan, Charu Ramakrishnan, Masatoshi Inoue, Janelle C Shane, Douglas J McKnight, Susumu Yoshizawa, Hideaki E Kato, Surya Ganguli, and Karl Deisseroth. "Cortical layer-specific critical dynamics triggering perception". In: *Science* 365.6453 (Aug. 2019), eaaw5202. DOI: 10.1126/science.aaw5202. URL: http://science.sciencemag.org/content/365/6453/eaaw5202.abstract.
- [7] Valentina Emiliani, Adam E Cohen, Karl Deisseroth, and Michael Häusser.
 "All-Optical Interrogation of Neural Circuits." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35.41 (Oct. 2015), pp. 13917-26. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.2916-15.2015. URL: http://www.ncbi.nlm.nih.gov/pubmed/26468193%20http: //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4604230.

- [8] Nicolas C. Pégard, Alan R. Mardinly, Ian Antón Oldenburg, Savitha Sridharan, Laura Waller, and Hillel Adesnik. "Three-dimensional scanless holographic optogenetics with temporal focusing (3D-SHOT)". In: *Nature Communications* 8.1 (Dec. 2017), p. 1228. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01031-3. URL: http://www.nature.com/articles/s41467-017-01031-3.
- [9] Nathan Tessema Ersaro, Cem Yalcin, and Rikky Muller. "The future of brain-machine interfaces is optical". In: *Nature Electronics* 6 (2023), pp. 96–98.
- [10] A. Maimone, A. Georgiou, and J. S. Kollin. "Holographic Near-Eye Displays for Virtual and Augmented Reality". In: *ACM Transactions on Graphics* 36.4 (2017).
- [11] Chenliang Chang, Kiseung Bang, Gordon Wetzstein, Byoungho Lee, and Liang Gao. "Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective". In: Optica 7.11 (2020), pp. 1563-1578. DOI: 10.1364/OPTICA.406004. URL: https://opg.optica.org/optica/abstract.cfm?URI=optica-7-11-1563.
- [12] Dapu Pi, Juan Liu, and Yongtian Wang. "Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display". In: Light: Science & Applications 11.1 (2022), p. 231. ISSN: 2047-7538. DOI: 10.1038/s41377-022-00916-3. URL: https://doi.org/10.1038/s41377-022-00916-3.
- [13] Liang Shi, Beichen Li, Changil Kim, Petr Kellnhofer, and Wojciech Matusik.
 "Towards real-time photorealistic 3D holography with deep neural networks". In: Nature 591.7849 (2021), pp. 234–239. ISSN: 1476-4687. DOI: 10.1038/s41586-020-03152-0. URL: https://doi.org/10.1038/s41586-020-03152-0.
- [14] Aravind Jakkinapalli, Balaji Baskar, and Sy-Bor Wen. "Femtosecond 3D photolithography through a digital micromirror device and a microlens array". In: *Applied Optics* 61.16 (2022), pp. 4891–4899. DOI: 10.1364/AO.457847. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-61-16-4891.
- [15] Atsushi Taguchi, Atsushi Nakayama, Ryosuke Oketani, Satoshi Kawata, and Katsumasa Fujita. "Multiphoton-Excited Deep-Ultraviolet Photolithography for 3D Nanofabrication". In: ACS Applied Nano Materials 3.11 (Nov. 2020), pp. 11434–11441. DOI: 10.1021/acsanm.0c02519. URL: https://doi.org/10.1021/acsanm.0c02519.
- [16] V Harinarayana and Y C Shin. "Two-photon lithography for three-dimensional fabrication in micro/nanoscale regime: A comprehensive review". In: Optics & Laser Technology 142 (2021), p. 107180. ISSN: 0030-3992. DOI: https://doi.org/10.1016/j.optlastec.2021.107180. URL: https://www.sciencedirect.com/science/article/pii/S0030399221002681.

- Yong-Sin Syu, Yu-Bin Huang, Ming-Ze Jiang, Chun-Ying Wu, and Yung-Chun Lee.
 "Maskless lithography for large area patterning of three-dimensional microstructures with application on a light guiding plate". In: *Optics Express* 31.8 (2023), pp. 12232-12248. DOI: 10.1364/0E.482160. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-8-12232.
- Yu Zhao, Ki-Chul Kwon, Munkh-Uchral Erdenebat, Seok-Hee Jeon, Mei-Lan Piao, and Nam Kim. "Implementation of full-color holographic system using non-uniformly sampled 2D images and compressed point cloud gridding". In: Optics Express 27.21 (2019), pp. 29746-29758. DOI: 10.1364/0E.27.029746. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-21-29746.
- [19] Yutaka Mori, Takahiko Fukuoka, and Takanori Nomura. "Speckle reduction in holographic projection by random pixel separation with time multiplexing". In: *Applied Optics* 53.35 (2014), pp. 8182–8188. DOI: 10.1364/A0.53.008182. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-53-35-8182.
- [20] Tianbo Liu, Milind Rajadhyaksha, and David L Dickensheets. "MEMS-in-the-lens architecture for a miniature high-NA laser scanning microscope". In: *Light: Science* & Applications 8.1 (2019), pp. 1–11. ISSN: 2047-7538.
- James A. Frank, Marc-Joseph Antonini, and Polina Anikeeva. "Next-generation interfaces for studying neural function". In: *Nature Biotechnology* 37 (2019), pp. 1013–1023. DOI: https://doi.org/10.1038/s41587-019-0198-8.
- [22] Jan M Rabaey Rikky Muller Simone Gambini. "A 0.013 mm², 5 μW, DC-Coupled Neural Signal Acquisition IC With 0.5 V Supply". In: Journal of Solid-State Circuits 47.1 (2012), pp. 232–243. DOI: https://doi.org/10.1109/JSSC.2011.2163552.
- [23] Andy Zhou, Samantha R. Santacruz, Benjamin C. Johnson, George Alexandrov, Ali Moin, Fred L. Burghardt, Jan M. Rabaey, Jose M. Carmena, and Rikky Muller. "A wireless and artefact-free 128-channel neuromodulation device for closed-loop stimulation and recording in non-human primates". In: *Nature Biomedical Engineering* 3 (2019), pp. 15–26. DOI: https://doi.org/10.1038/s41551-018-0323-x.
- [24] Rikky Muller, Hanh-Phuc Le, Wen Li, Peter Ledochowitsch, Simone Gambini, Toni Bjorninen, Aaron Koralek, Jose M Carmena, Michel M Maharbiz, Elad Alon, et al. "A minimally invasive 64-channel wireless μECoG implant". In: *IEEE Journal* of Solid-State Circuits 50.1 (2015), pp. 344–359.
- [25] Daniel R. Merrill, Marom Bikson, and John G.R. Jefferys. "Electrical stimulation of excitable tissue: design of efficacious and safe protocols". In: Journal of Neuroscience Methods 141.2 (Feb. 2005), pp. 171-198. ISSN: 0165-0270. DOI: 10.1016/J.JNEUMETH.2004.10.020. URL: https://www.sciencedirect.com/science/article/pii/S0165027004003826.

- [26] Ali Ersen, Stella Elkabes, David S Freedman, and Mesut Sahin. "Chronic tissue response to unterhered microelectrode implants in the rat brain and spinal cord". In: *Journal of neural engineering* 12.1 (2015).
- [27] David Kleinfeld, Lan Luan, Partha P. Mitra, Jacob T. Robinson, Rahul Sarpeshkar, Kenneth Shepard, Chong Xie, and Timothy D. Harris. "Can One Concurrently Record Electrical Spikes from Every Neuron in a Mammalian Brain?" In: *Neuron* (Sept. 2019). ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2019.08.011. URL: https: //www.sciencedirect.com/science/article/abs/pii/S0896627319306956.
- [28] Mei Hong Zhu, Jinyoung Jang, Milena M. Milosevic, and Srdjan D. Antic. "Population imaging discrepancies between a genetically-encoded calcium indicator (GECI) versus a genetically-encoded voltage indicator (GEVI)". In: Scientific Reports 11.5295 (2021).
- [29] Yiyang Gong, Cheng Huang, Jin Zhong Li, Benjamin F Grewe, Yanping Zhang, Stephan Eismann, and Mark J Schnitzer. "High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor." In: Science (New York, N.Y.) 350.6266 (Dec. 2015), pp. 1361-6. ISSN: 1095-9203. DOI: 10.1126/science.aab0810. URL: http://www.ncbi.nlm.nih.gov/pubmed/26586188%20http: //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4904846.
- [30] Emiliano Ronzitti, Rossella Conti, Valeria Zampini, Dimitrii Tanese, Amanda J Foust, Nathan Klapoetke, Edward S Boyden, Eirini Papagiakoumou, and Valentina Emiliani. "Submillisecond Optogenetic Control of Neuronal Firing with Two-Photon Holographic Photoactivation of Chronos." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 37.44 (Nov. 2017), pp. 10679–10689. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1246-17.2017. URL: http://www.ncbi.nlm.nih.gov/pubmed/28972125%20http: //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5666587.
- [31] Nicolas C. Pégard, Hsiou-Yuan Liu, Nick Antipa, Maximillian Gerlock, Hillel Adesnik, and Laura Waller. "Compressive light-field microscopy for 3D neural activity recording". In: Optica 3.5 (May 2016), p. 517. ISSN: 2334-2536. DOI: 10.1364/OPTICA.3.000517. URL: https://www.osapublishing.org/abstract.cfm?URI=optica-3-5-517.
- [32] Mary Ann Go, Max Mueller, Michael Lawrence Castañares, Veronica Egger, and Vincent R Daria. "A compact holographic projector module for high-resolution 3D multi-site two-photon photostimulation". In: *PloS one* 14.1 (2019).
- [33] H. Mikami, J. Harmon, H. Kobayashi, S. Hamad, Y. Wang, O. Iwata, K. Suzuki, T. Ito, Y. Aisaka, N. Kutsuna, K. Nagasawa, H. Watarai, Y. Ozeki, and K. Goda. "Ultrafast confocal fluorescence microscopy beyond the fluorescence lifetime limit". In: Optica 5.2 (2018).

- [34] Samuel J Yang, William E Allen, Isaac Kauvar, Aaron S Andalman, Noah P Young, Christina K Kim, James H Marshel, Gordon Wetzstein, and Karl Deisseroth.
 "Extended field-of-view and increased-signal 3D holographic illumination with time-division multiplexing". In: *Optics Express* 23.25 (2015), pp. 32573–32581. DOI: 10.1364/0E.23.032573. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-23-25-32573.
- [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: ACM SIGGRAPH 2022 Conference Proceedings. SIGGRAPH '22. New York, NY, USA: Association for Computing Machinery, 2022. ISBN: 9781450393379. DOI: 10.1145/3528233.3530734. URL: https://doi.org/10.1145/3528233.3530734.
- [36] V R Curtis, N W Caira, J Xu, A G Sata, and N C Pégard. "DCGH: Dynamic Computer Generated Holography for Speckle-Free, High Fidelity 3D Displays". In: 2021 IEEE Virtual Reality and 3D User Interfaces (VR). 2021, pp. 1–9. ISBN: 2642-5254 VO -. DOI: 10.1109/VR50410.2021.00097.
- [37] Byounghyo Lee, Dongyeon Kim, Seungjae Lee, Chun Chen, and Byoungho Lee.
 "High-contrast, speckle-free, true 3D holography via binary CGH optimization". In: Scientific Reports 12.1 (2022), p. 2811. ISSN: 2045-2322. DOI: 10.1038/s41598-022-06405-2. URL: https://doi.org/10.1038/s41598-022-06405-2.
- [38] Joseph W. Goodman. Introduction to Fourier Optics. McGraw-Hill, 1996.
- [39] Peter Wai Ming Tsang. Computer-Generated Phase-Only Holograms for 3D Displays: A Matlab Approach. Cambridge University Press, 2021. DOI: 10.1017/9781316999172.
- [40] Zichen Zhang, Zheng You, and Daping Chu. "Fundamentals of phase-only liquid crystal on silicon (LCOS) devices". In: *Light: Science & Applications* 3.10 (Oct. 2014), e213-e213. ISSN: 2047-7538. DOI: 10.1038/lsa.2014.94. URL: http://www.nature.com/articles/lsa201494.
- [41] Terry A Bartlett, William C McDonald, James N Hall, Patrick I Oden, Dennis Doane, Remington S Ketchum, and Taylor Byrum. "Recent advances in the development of the Texas Instruments phase-only microelectromechanical systems (MEMS) spatial light modulator". In: *Emerging Digital Micromirror Device Based Systems and Applications XIII* 11698 (2021), pp. 103–116.
- [42] Peter J Ryan, Steven A Cornelissen, Charlie V Lam, and Paul A Bierden. "Performance analysis of two high actuator count MEMS deformable mirrors". In: *MEMS Adaptive Optics VII*. Vol. 8617. International Society for Optics and Photonics, 2013, p. 861705.

- [43] Youmin Wang, Guangya Zhou, Xiaosheng Zhang, Kyungmok Kwon, Pierre-A Blanche, Nicholas Triesault, Kyoung-sik Yu, and Ming C Wu. "2D broadband beamsteering with large-scale MEMS optical phased array". In: Optica 6.5 (2019), pp. 557–562. ISSN: 2334-2536.
- [44] Robert Corrigan et al. "Silicon Light Machines—Grating Light Valve Technology Brief, Jun. 2001 ver C". In: Sunnyvale, California (), pp. 1–8.
- [45] Nicholas C. Harris, Yangjin Ma, Jacob Mower, Tom Baehr-Jones, Dirk Englund, Michael Hochberg, and Christophe Galland. "Efficient, compact and low loss thermo-optic phase shifter in silicon". In: Opt. Express 22.9 (May 2014), pp. 10487–10493. DOI: 10.1364/0E.22.010487. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-22-9-10487.
- [46] Yuriko Maegami, Guangwei Cong, Morifumi Ohno, Makoto Okano, Kazuto Itoh, Nobuhiko Nishiyama, Shigehisa Arai, and Koji Yamada. "High-efficiency strip-loaded waveguide based silicon Mach-Zehnder modulator with vertical p-n junction phase shifter". In: *Opt. Express* 25.25 (Dec. 2017), pp. 31407–31416. DOI: 10.1364/0E.25.031407. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-25-25-31407.
- [47] Jacqueline Geler-Kremer, Felix Eltes, Pascal Stark, David Stark, Daniele Caimi, Heinz Siegwart, Bert Jan Offrein, Jean Fompeyrine, and Stefan Abel. "A ferroelectric multilevel non-volatile photonic phase shifter". In: *Nature Photonics* 16.7 (2022), pp. 491–497. DOI: 10.1038/s41566-022-01003-0. URL: https://doi.org/10.1038/s41566-022-01003-0.
- [48] Christopher L Panuski, Ian Christen, Momchil Minkov, Cole J Brabec, Sivan Trajtenberg-Mills, Alexander D Griffiths, Jonathan J D McKendry, Gerald L Leake, Daniel J Coleman, and Cung Tran. "A full degree-of-freedom spatiotemporal light modulator". In: *Nature Photonics* 16.12 (2022), pp. 834–842. ISSN: 1749-4893.
- [49] G. P. Watson, V. Aksyuk, M. E. Simon, D. M. Tennant, R. A. Cirelli, W. M. Mansfield, F. Pardo, D. O. Lopez, C. A. Bolle, A. R. Papazian, N. Basavanhally, J. Lee, R. Fullowan, F. Klemens, J. Miner, A. Kornblit, T. Sorsch, L. Fetter, M. Peabody, J. E. Bower, J. S. Weiner, and Y. L. Low. "Spatial light modulator for maskless optical projection lithography". In: Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena 24.6 (Nov. 2006), pp. 2852–2856. ISSN: 1071-1023. DOI: 10.1116/1.2387156. URL: http: //scitation.aip.org/content/avs/journal/jvstb/24/6/10.1116/1.2387156.
- [50] David M. Benton. "Multiple beam steering using dynamic zone plates on a micromirror array". In: *Optical Engineering* 57.07 (July 2018), p. 1. ISSN: 0091-3286. DOI: 10.1117/1.0E.57.7.073109. URL:

https://www.spiedigitallibrary.org/journals/optical-engineering/volume-57/issue-07/073109/Multiple-beam-steering-using-dynamic-zone-plateson-a-micromirror/10.1117/1.0E.57.7.073109.full.

- [51] Richard Osgood jr and Xiang Meng. Principles of Photonic Integrated Circuits. Springer, 2022. DOI: https://doi.org/10.1007/978-3-030-65193-0.
- [52] Janelle C. Shane, Douglas J. McKnight, Adrian Hill, Kevin Taberski, and Steve Serati. "Designing a new spatial light modulator for holographic photostimulation". In: Optical Trapping and Optical Micromanipulation XVI. Ed. by Kishan Dholakia and Gabriel C. Spalding. Vol. 11083. SPIE, Sept. 2019, p. 3. ISBN: 9781510628595. DOI: 10.1117/12.2528558. URL: https://www.spiedigitallibrary.org/conference-proceedings-ofspie/11083/2528558/Designing-a-new-spatial-light-modulator-forholographic-photostimulation/10.1117/12.2528558.full.
- [53] Grace Kuo, Laura Waller, Ren Ng, and Andrew Maimone. "High Resolution éTendue Expansion for Holographic Displays". In: ACM Trans. Graph. 39.4 (Aug. 2020). ISSN: 0730-0301. DOI: 10.1145/3386569.3392414. URL: https://doi.org/10.1145/3386569.3392414.
- [54] M Hossein Eybposh, Vincent R Curtis, Jose Rodríguez-Romaguera, and Nicolas C Pégard. "Advances in computer-generated holography for targeted neuronal modulation". In: *Neurophotonics* 9.4 (2022), p. 41409. ISSN: 2329-423X.
- Rafael Piestun, Boris Spektor, and Joseph Shamir. "Wave fields in three dimensions: analysis and synthesis". In: Journal of the Optical Society of America A 13.9 (1996), pp. 1837–1848. DOI: 10.1364/JOSAA.13.001837. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-13-9-1837.
- [56] Nathan Tessema Ersumo, Cem Yalcin, Nick Antipa, Nicolas Pégard, Laura Waller, Daniel Lopez, and Rikky Muller. "A micromirror array with annular partitioning for high-speed random-access axial focusing". In: *Light: Science and Applications* 9.1 (2020). ISSN: 20477538. DOI: 10.1038/s41377-020-00420-6.
- [57] Junghyun Park, Byung Gil Jeong, Sun Il Kim, Duhyun Lee, Jungwoo Kim, Changgyun Shin, Chang Bum Lee, Tatsuhiro Otsuka, Jisoo Kyoung, Sangwook Kim, Ki-Yeon Yang, Yong-Young Park, Jisan Lee, Inoh Hwang, Jaeduck Jang, Seok Ho Song, Mark L Brongersma, Kyoungho Ha, Sung-Woo Hwang, Hyuck Choo, and Byoung Lyong Choi. "All-solid-state spatial light modulator with independent phase and amplitude control for three-dimensional LiDAR applications". In: *Nature Nanotechnology* 16.1 (2021), pp. 69–76. ISSN: 1748-3395. DOI: 10.1038/s41565-020-00787-y. URL: https://doi.org/10.1038/s41565-020-00787-y.

- [58] Vincent Villette, Mariya Chavarha, Ivan K Dimov, Jonathan Bradley, Lagnajeet Pradhan, Benjamin Mathieu, Stephen W Evans, Simon Chamberland, Dongqing Shi, and Renzhi Yang. "Ultrafast two-photon imaging of a high-gain voltage indicator in awake behaving mice". In: *Cell* 179.7 (2019), pp. 1590–1608. ISSN: 0092-8674.
- [59] Seung-Cheol Kim and Eun-Soo Kim. "Effective generation of digital holograms of three-dimensional objects using a novel look-up table method". In: Applied Optics 47.19 (2008), pp. D55–D62. DOI: 10.1364/A0.47.000D55. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-47-19-D55.
- [60] Dapu Pi, Juan Liu, Ruidan Kang, Zhiqi Zhang, and Yu Han. "Reducing the memory usage of computer-generated hologram calculation using accurate high-compressed look-up-table method in color 3D holographic display". In: Optics Express 27.20 (2019), pp. 28410-28422. DOI: 10.1364/OE.27.028410. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-20-28410.
- [61] Tomoyoshi Shimobaba, Nobuyuki Masuda, and Tomoyoshi Ito. "Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane". In: Optics Letters 34.20 (2009), pp. 3133–3135. DOI: 10.1364/OL.34.003133. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-34-20-3133.
- [62] Naotaka Hasegawa, Tomoyoshi Shimobaba, Takashi Kakue, and Tomoyoshi Ito.
 "Acceleration of hologram generation by optimizing the arrangement of wavefront recording planes". In: *Applied Optics* 56.1 (2017), A97-A103. DOI: 10.1364/A0.56.000A97. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-56-1-A97.
- [63] Panpan Yu, Yifan Liu, Ziqiang Wang, Jinyang Liang, Xingsi Liu, Yinmei Li, Chengwei Qiu, and Lei Gong. "Ultrahigh-density 3D holographic projection by scattering-assisted dynamic holography". In: Optica 10.4 (2023), pp. 481-490. DOI: 10.1364/OPTICA.483057. URL: https://opg.optica.org/optica/abstract.cfm?URI=optica-10-4-481.
- [64] Ghaith Makey, Özgün Yavuz, Denizhan K Kesim, Ahmet Turnalı, Parviz Elahi, Serim Ilday, Onur Tokel, and F Ömer Ilday. "Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors". In: *Nature Photonics* 13.4 (2019), pp. 251–256. ISSN: 1749-4893. DOI: 10.1038/s41566-019-0393-7. URL: https://doi.org/10.1038/s41566-019-0393-7.
- [65] Ralph W Gerchberg. "A practical algorithm for the determination of plane from image and diffraction pictures". In: *Optik* 35.2 (1972), pp. 237–246.

- [66] Jingzhao Zhang, Nicolas Pégard, Jingshan Zhong, Hillel Adesnik, and Laura Waller.
 "3D computer-generated holography by non-convex optimization". In: Optica 4.10 (Oct. 2017), p. 1306. ISSN: 2334-2536. DOI: 10.1364/OPTICA.4.001306. URL: https://www.osapublishing.org/abstract.cfm?URI=optica-4-10-1306.
- [67] Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Felix Heide, and Henry Fuchs.
 "Computing high quality phase-only holograms for holographic displays". In: *Proc.SPIE*. Vol. 11310. Feb. 2020, p. 1131006. DOI: 10.1117/12.2547647. URL: https://doi.org/10.1117/12.2547647.
- [68] M Hossein Eybposh, Nicholas W Caira, Mathew Atisa, Praneeth Chakravarthula, and Nicolas C Pégard. "DeepCGH: 3D computer-generated holography using deep learning". In: Optics Express 28.18 (2020), pp. 26636-26650. DOI: 10.1364/0E.399624. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-28-18-26636.
- [69] 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". In: ACM Trans. Graph. 40.6 (Dec. 2021). ISSN: 0730-0301. DOI: 10.1145/3478513.3480542. URL: https://doi.org/10.1145/3478513.3480542.
- [70] Youchao Wang, Daoming Dong, Peter J Christopher, Andrew Kadis, Ralf Mouthaan, Fan Yang, and Timothy D Wilkinson. "Hardware implementations of computer-generated holography: a review". In: *Optical Engineering* 59.10 (Feb. 2020), p. 102413. DOI: 10.1117/1.0E.59.10.102413. URL: https://doi.org/10.1117/1.0E.59.10.102413.
- [71] Wenhui Zhang, Hao Zhang, and Guofan Jin. "Frequency sampling strategy for numerical diffraction calculations". In: *Optics Express* 28.26 (2020), pp. 39916–39932.
 DOI: 10.1364/0E.413636. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-28-26-39916.
- [72] D F Crouse. "On implementing 2D rectangular assignment algorithms". In: *IEEE Transactions on Aerospace and Electronic Systems* 52.4 (2016), pp. 1679–1696. ISSN: 1557-9603 VO 52. DOI: 10.1109/TAES.2016.140952.
- [73] Nathan Tessema Ersaro and Cem Yalcin. Python implementation and evaluation of NIMBLE-PATCH. 2023. URL: https://github.com/MullerGroup/NIMBLE-PATCH-demo.
- [74] Mostafa Agour, Claas Falldorf, and Christoph Von Kopylow. "Digital pre-filtering approach to improve optically reconstructed wavefields in opto-electronic holography". In: *Journal of Optics* 12.5 (2010), p. 55401. ISSN: 2040-8986.

- [75] Yang Wu, Jun Wang, Chun Chen, Chan-Juan Liu, Feng-Ming Jin, and Ni Chen.
 "Adaptive weighted Gerchberg-Saxton algorithm for generation of phase-only hologram with artifacts suppression". In: *Optics Express* 29.2 (2021), pp. 1412–1427. DOI: 10.1364/0E.413723. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-2-1412.
- [76] Travis J Gould, Emil B Kromann, Daniel Burke, Martin J Booth, and Joerg Bewersdorf. "Auto-aligning stimulated emission depletion microscope using adaptive optics". In: Optics Letters 38.11 (2013), pp. 1860–1862. DOI: 10.1364/0L.38.001860. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-38-11-1860.
- Ya-Ping Zhang, Fan Wang, Ting-Chung Poon, Shuang Fan, and Wei Xu. "Fast generation of full analytical polygon-based computer-generated holograms". In: *Optics Express* 26.15 (2018), pp. 19206–19224. DOI: 10.1364/OE.26.019206. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-26-15-19206.
- [78] Daisuke Iwai, Hidetoshi Izawa, Kenji Kashima, Tatsuyuki Ueda, and Kosuke Sato. "Speeded-Up focus control of electrically tunable Lens by Sparse optimization". In: *Scientific reports* 9.1 (2019), pp. 1–6. ISSN: 2045-2322.
- [79] Hung-Chun Lin, Ming-Syuan Chen, and Yi-Hsin Lin. "A Review of Electrically Tunable Focusing Liquid Crystal Lenses". In: *Transactions on Electrical and Electronic Materials* 12.6 (Dec. 2011), pp. 234-240. ISSN: 1229-7607. DOI: 10.4313/TEEM.2011.12.6.234. URL: http://koreascience.or.kr/journal/view. jsp?kj=E1TEA0%7B%5C&%7Dpy=2011%7B%5C&%7Dvnc=v12n6%7B%5C&%7Dsp=234.
- [80] S. Kang, E. Dotsenko, D. Amrhein, C. Theriault, and C. B. Arnold. "Ultra-high-speed variable focus optics for novel applications in advanced imaging". In: *Photonic Instrumentation Engineering V.* Ed. by Yakov G. Soskind. Vol. 10539. International Society for Optics and Photonics. SPIE, 2018, p. 1053902. DOI: 10.1117/12.2294487. URL: https://doi.org/10.1117/12.2294487.
- [81] Lijun Zhu, Pang-Chen Sun, Dirk-Uwe Bartsch, William R. Freeman, and Yeshaiahu Fainman. "Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation". In: *Appl. Opt.* 38.1 (Jan. 1999), pp. 168–176. DOI: 10.1364/A0.38.000168. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-38-1-168.
- [82] Cem Yalcin, Nathan Tessema Ersumo, George Bocchetti, Mohammad Meraj Ghanbari, Nick Antipa, Sina Faraji Alamouti, Laura Waller, Daniel Lopez, and Rikky Muller. "19.3 A MEMS-Based Dynamic Light Focusing System for Single-Cell Precision in optogenetics". In: 2021 IEEE International Solid-State Circuits Conference (ISSCC). Vol. 64. 2021, pp. 290–292. ISBN: 2376-8606 VO -64. DOI: 10.1109/ISSCC42613.2021.9365792.

- [83] Matthias Imboden, Jackson Chang, Corey Pollock, Evan Lowell, Mehmet Akbulut, Jessica Morrison, Thomas Stark, Thomas G. Bifano, and David J. Bishop.
 "High-Speed Control of Electromechanical Transduction: Advanced Drive Techniques for Optimized Step-and-Settle Response of MEMS Micromirrors". In: *IEEE Control Systems* 36.5 (Oct. 2016), pp. 48–76. ISSN: 1066-033X. DOI: 10.1109/MCS.2016.2584338. URL: http://ieeexplore.ieee.org/document/7569123/.
- [84] Yashodhan Moghe, Torsten Lehmann, and Tim Piessens. "Nanosecond Delay Floating High Voltage Level Shifters in a 0.35 μm HV-CMOS Technology". In: *IEEE Journal of Solid-State Circuits* 46.2 (2011), pp. 485–497. DOI: 10.1109/JSSC.2010.2091322.
- [85] Y. Perelman and R. Ginosar. "A low-power inverted ladder D/a converter". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 53.6 (2006), pp. 497–501.
 DOI: 10.1109/TCSII.2006.875313.