

COMPUTATIONAL IMAGING ACROSS SIMULTANEOUS TIMESCALES

by

Mian Wei

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Computer Science
University of Toronto

© Copyright 2025 by Mian Wei

Mian Wei
Doctor of Philosophy

Department of Computer Science
University of Toronto
2025

Abstract

Modern digital cameras improve at a rapid pace, yet they remain limited by fixed exposure times and producing one image per exposure. While sufficient for scenes that vary at a relatively uniform pace, this constraint hinders the ability to capture dynamic scenes with events that unfold at vastly different rates. This thesis addresses this limitation by developing frameworks for emerging computational cameras that enable imaging across diverse timescales simultaneously.

We focus on two types of computational cameras—coded two-bucket (C2B) cameras, a new type of coded-exposure-imaging (CEI) camera, and single photon avalanche diode (SPAD)-based cameras, which are sensitive to individual photons and capable of timestamping their arrival.

First, we present a general framework for controlling C2B cameras. A C2B camera uses pixel-specific binary codes to direct incident light into one of two "buckets" for each pixel. The active buckets switch rapidly and pixelwise, allowing C2B cameras to simultaneously capture information at two discrete timescales—the conventional frame scale and the shorter sub-frame scale. We formulate the code design as the solution to an optimal multiplexing problem and use these optimized codes to demonstrate live, per-pixel 3D shape estimation—such as disparity and surface normals—with active illumination in a single exposure.

Second, we introduce a new imaging regime called passive ultra-wideband imaging that images a dynamic scene over a continuous range of timescales simultaneously from seconds to picoseconds. Drawing insights from stochastic calculus, we develop a flux probing theory that enables SPAD-based cameras to operate in this regime. Our experiments showcase several never-seen-before capabilities of passive ultra-wideband imaging including recording ultra-wideband videos, which can replay everyday motion at 30 Hz and reveal light propagation by slowing down to 250 GHz.

Finally, we introduce a method to detect and analyze asynchronous pulsed lasers using a free-running SPAD camera. By leveraging each laser's distinct frequency signature, we image at laser-specific timescales to characterize their pulses. This capability lets us visualize for the first time, multiple, asynchronously propagating light pulses within a single scene and enables a new form of passive 3D vision, by harvesting photons from two or more out-of-view lasers.

BUT even [silly] people often make fine scientists. However wonderful a receptive mind may be, it is not central to science. Nor is mental speed. To persist in your inquiry and to engage it fruitfully—this is what makes a first-rate scientist. Curiosity, diligence, and ambition are traits far more essential than imagination. And [silly] people are often remarkably well endowed with these traits.

— Eugene P. Wigner

Acknowledgements

The acknowledgements seem to me a place to immortalize my gratitude for people in the annals of my personal history, and so they are written as much for those I acknowledge as for me to remember when I grow older and wearier.

Two people have played a profound role during my PhD, and without whom, this thesis would not be possible.

Το πρώτο πρόσωπο είναι ο σύμβουλος και μέντοράς μου, ο Κύρος Κουτουλάκος. Παρόλο που οι στιγμές μας δεν ήταν πάντα εύκολες (μερικές φορές θα τολμούσα να πω ακόμα και εκνευριστικές) και είχαμε αρκετές διαφωνίες όλα αυτά τα δέκα και πλέον χρόνια, ποτέ δεν αμφέβαλλα για την ακεραιότητά του απέναντι στην έρευνα και την αριστεία, ούτε για την αφοσίωσή του στην προσωπική μου εξέλιξη και επιτυχία. Ο Κύρος ήταν ακούραστος στην προσπάθειά του να κατανοήσει ένα πρόβλημα και να εμβαθύνει σε αυτό. Αυτό το επίπεδο αφοσίωσης αποτελεί πρότυπο που ελπίζω να υιοθετήσω καθώς προχωρώ στην καριέρα μου. Είμαι βέβαιος πως δεν ήμουν πάντα ο πιο εύκολος φοιτητής για να συμβουλευθεί κανείς, δεδομένης της τάσης μου για πείσμα και σχολαστικότητα—κάτι στο οποίο οι φίλοι μου θα συμφωνούσαν—και γι’ αυτό, σε ευχαριστώ για την υπομονή και την κατανόησή σου.¹

Το δεύτερο πρόσωπο (και ήδη ξέρεις ποιος είσαι ;)) είναι ο Σωτήρης Νούσιας, ή μάλλον να πω καλύτερα ο Καθηγητής Σωτήρης Νούσιας. Αν ποτέ υπήρξε κάποιος που θα μπορούσε να χαρακτηριστεί «εργασιακός σύντροφος» μου κατά τη διάρκεια της θητείας μου στο ΔΓΠ, αυτός είναι σίγουρα ο Σωτήρης. Αλλά ακόμη και ο όρος «εργασιακός σύντροφος» δεν αποδίδει πλήρως τη σχέση μας, οπότε προτιμώ να τον αποκαλώ «το καλύτερο μου μισό». Έχουμε τσακωθεί, έχουμε κλάψει, έχουμε γελάσει, και έχουμε πει καφέ, με την ευρεία έννοια, αναρίθμητες φορές. Κατά κάποιον τρόπο, ενάντια σε κάθε πιθανότητα, συναντηθήκαμε στον ίδιο τόπο και χρόνο. Ο χρόνος που περάσαμε μαζί στο ΥοφΤ (ορ ΔΓΠ) ήταν πραγματικά η πλατωνική ιδέα μου για το πώς πρέπει να γίνεται η έρευνα, και ο χρόνος που περάσαμε εκτός εργασίας ήταν η πλατωνική ιδέα μου για τη φιλία.²³

I want to thank Alec Jacobson, David Fleet, and Roman Genov for being on my supervisory committee. A special thanks to David Lindell for his involvement with the single-photon projects. Also, a warm thank you to Srinivas Narasimhan, who agreed to be my external appraiser, and Aviad Levis, for being my internal-external. I want to thank my co-authors who helped with the work presented in this thesis.

¹The first of whom is my advisor and mentor, Kyros Kutulakos. Although times have not always been great (dare I even say frustrating), and we have had our fair share of arguments over the decade (and then some), I have never doubted his integrity to research and excellence nor his commitment to my growth and success. Kyros was unrelenting in his pursuit of understanding a problem and thinking about it deeply and that is a standard I hope I can bring to the table as I venture forward in my career. I’m sure I’ve not always been the easiest graduate student to advise, given my penchant for stubbornness and pedantry, as my friends will agree, so thank you for being patient and understanding with me.

²The second (and you already know who you are ;)) is Sotiris Nousias or should I say Professor Sotiris Nousias. If ever there was someone who could be considered my ‘work spouse’ during my time at DGP, it would be Sotiris. But even ‘work spouse’ does not do things justice so I prefer to call him my ‘better half’. We’ve argued, we’ve cried, we’ve laughed, and we’ve grabbed coffee uncountably (if you’ll allow me to use this term loosely) many times. Somehow, against all odds, we overlapped in both space and time. The time we spent working truly was my platonic ideal of conducting research, and the time we spent not working the platonic ideal of friendship.

³Translations courtesy of Sotiris Nousias.

I want to give a special shout-out to a few people in DGP who have made my time there immensely wonderful: Vismay Modi, Brenna Li, Rinat Abdrashitov, and Sarah Kushner. These are some of the best friends anyone could ask for, and they were a godsend for my sanity. To the many lives who have walked through the doors since I first started and kept my time here lively and fun, I can only enumerate a subset (in no particular order): Eva Mackamul, Emilie Yu, Moko, Vicky Bilbily, Siddharth Somasundaram, Karran Pandey, Bogdan Pikula, Yasaman Rohanifar, Zhecheng Wang, Selina Ling, Mengfei Liu, Kinjal Parikh, Fanny Chevalier, Nicole Sultanum, Michael Glueck, Silvia Sellán, Bruno De Araujo, Peter Hamilton, Varun Perumal, and Mazen Al Borno.

To those who were part of Kyros' group, I thank Parsa Mirdeghan, Robin Swanson, Huixuan Tang, Matthew O'Toole, Liviu-Mihai Calin, Howard Xiao, Pan Zhang, Zhao Wei, Yawen Ma, and John Mather. To Roman's amazing group, I thank Rahul Gulve, Navid Sarhangnejad, Zhengfan Xia, Nikita Gusev, Gairik Dutta, Roberto Rangel, Ayandev Barman, and Nikola Katic. To Oliver Cos-sairt's group, who hosted me wonderfully for a summer in Chicago, I thank Fengqiang Li, Zihao Wang, Florian Willomitzer. To Mohit Gupta's group, who hosted me on two separate occasions, I thank Anant Gupta, Felipe Gutierrez-Barragan, Sizhuo Ma, Jongho Lee, and Atul Ingle. To the wonderful people whom I got to meet outside Toronto and who made conferences fun: Mark Sheinin (Tokyo + Seattle), Vadim Holodovsky (Japan), Dorian Chan, Varun Sundar, Sacha Jungerman (Seattle + Paris).

To my support system, Kacy Lin, Cindy You, Haider Rizvi, and Alice Yang. Throughout the many years, they kept my ego in check and, more importantly, me grounded and helped me navigate the many hardships that naturally come with doing a PhD. I don't have enough words to express how much you have done for me and my love for you all <3. Kacy, we've both found our callings in life, and you've been a constant source of joy in my life. I'm so blessed we are in each other's lives. Cindy, I'm not sure I won the race (certainly not the one to motherhood), but congratulations on becoming a mother to two wonderful kids—Charlotte and Oliver Saracino. I want to thank the Glenforest crew (Miranda Shi, Brendan O'Sullivan, Mibo Zhao, Jacob Wanuch, and Krystal Beh) + Cottage crew (Wasi Rizvi, Kerrie Trottier, and Jessica Morales) for the many fun hangouts and board game sessions. I'm grateful to my landlords—Jack Budgell and Ann Camilleri—for being so wonderful and providing me a home to stay at for the many years. I want to especially thank the U of T Veg Club, particularly Lingling Deng, Emily Yuen, Naz Georgis-Yap, Joni Soropia, and Sanjeev Narayanaswamy, for giving me a home the last few years and many wonderful memories; I'll sorely miss my position as the resident old mian. I'm grateful to the Big Carrot, the (medium) Big Carrot, Food Basics, and Whole Foods for all the delicious and fresh food I was able to eat. To Ayal Sher, thank you for being a deer friend and amazing person; I'm in awe of your commitment to reducing food waste and always being on the right side; I hope you realize just how much fun I've had on the mian-y trips these past few months.

I am eternally grateful to my parents—林明星 and 魏景元—for supporting me through my time!

Finally, to borrow a page from Silvia Sellán's acknowledgements, I want to thank the Canadian and American taxpayers for supporting my research; it has been a tremendous privilege.

Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	vi
1 Introduction	1
1.1 Thesis roadmap	3
2 Image Formation and Cameras	4
2.1 Radiant Energy	4
2.2 Conventional Cameras	6
2.3 Coded-Exposure Imaging Cameras	7
2.4 SPAD-based Cameras	11
2.4.1 Photon number	11
3 Imaging Two Time Scales with Coded Two-Bucket Cameras	12
3.1 Introduction	12
3.2 Coded Two-Bucket Imaging	15
3.2.1 Acquiring two-bucket illumination mosaics	17
3.3 Per-Pixel Estimation of Normals and Depth	17
3.4 Code Matrices for Bucket Multiplexing	19
3.5 Empirical SNR Analysis	21
3.5.1 Empirical SNR gain for image intensities.	21
3.5.2 Empirical SNR gain for illumination ratios.	22
3.5.3 Theoretical SNR gain.	22
3.5.4 Empirical evaluation of SNR gain for illumination ratios.	22
3.6 One-Shot Shape from Two-Bucket Illumination Mosaics	23
3.6.1 Performance evaluation on synthetic data	23
3.7 C2B Pixel Design Considerations	25
3.8 Experiments	27
3.8.1 Experimental conditions	27
3.8.2 Quantitative experiments: one-shot versus multi-frame acquisition	27
3.8.3 Live 3D acquisition experiments: photometric stereo	30
3.8.4 Live 3D acquisition experiments: structured-light triangulation	30
3.9 Limitations	31
3.10 Concluding Remarks	32
4 Passive Ultra-Wideband Single-Photon Imaging	34
4.1 Introduction	34
4.2 The Passive Ultra-Wideband Imaging Regime	36

4.2.1	Imaging with photon timestamps	38
4.3	Probing Flux Functions	39
4.3.1	The photon counting process	39
4.3.2	Flux probing theory	40
4.4	Flux Function Reconstruction	43
4.4.1	Choosing the step size Δf for flux probing	44
4.5	Experiments	44
4.5.1	Hardware	46
4.5.2	1D passive ultra-wideband imaging of Figure 4.1.1 scene	46
4.5.3	1D passive ultra-wideband imaging over room-size distances	49
4.5.4	Passive reconstruction of laser pulse trains	51
4.5.5	2D fan experiment: freezing time at all timescales	52
4.5.6	2D Coca-Cola bottle experiment	59
4.5.7	Recovering passive non-line-of-sight (NLOS) videos	60
4.6	Simulations	63
4.6.1	Normal distribution of probing measurements	63
4.6.2	Passive reconstruction of laser pulse trains	65
4.6.3	Recovering passive non-line-of-sight (NLOS) videos	66
4.7	Limitations	69
4.8	Concluding Remarks	69
5	Flux Probing Theory for Non-Negligible Dead Time	71
5.1	Introduction	71
5.2	Preliminary definitions	71
5.3	Photon statistics	72
5.4	Mathematical Treatment of the Counting Process	73
5.5	Flux Probing under Non-Negligible Dead Time	74
5.5.1	Generalized flux probing	75
5.5.2	Generalized algorithm	76
5.5.3	Dead-time compensation step of Algorithm 5.1	77
5.6	Experiments	77
5.6.1	Probing with SPAD arrays and non-negligible dead time	77
5.7	Simulations	79
5.7.1	Flux probing under non-negligible dead time	79
5.8	Limitations	81
5.9	Conclusion	81
6	Opportunistic Single-Photon Time of Flight	82
6.1	Introduction	82
6.2	Opportunistic Time of Flight	84
6.3	Incident Flux Model for Pulsed Laser Light	85
6.4	Flux Probing Framework for Opportunistic ToF	87
6.5	Laser Discovery & Synchronization	89
6.5.1	Method pseudocode	91

6.6	Geometric Optimization	94
6.7	Experiments	95
6.7.1	Hardware, 2D acquisition procedure & algorithm details	96
6.7.2	Visualizing propagating wavefront of multiple asynchronous lasers	97
6.7.3	Opportunistic time-of-flight 3D imaging of room-scale scene	102
6.7.4	3D reconstruction with dynamic camera & pulsed laser locations.	103
6.7.5	Quantitative experiments: pulse delay estimation against frequency error . .	106
6.7.6	Quantitative experiments: assessing geometry with 3D-printed objects . . .	107
6.7.7	Opportunistic time-of-flight under strong ambient light	109
6.7.8	Quantitative experiments: frequency localization precision against SBR . . .	111
6.7.9	Quantitative experiments: 3D geometry accuracy under strong ambient light	113
6.8	Simulations	113
6.8.1	Simulation details	114
6.8.2	3D reconstruction and localization consistency	115
6.8.3	3D reconstruction accuracy	115
6.8.4	Two-laser detection test	116
6.9	Limitations	118
6.10	Concluding remarks	118
7	Conclusions	120
7.1	Limitations	121
7.2	Future Directions	121
A	Coded Two-Bucket Camera	123
A.1	Proofs and Derivations	123
A.1.1	Gaussian approximation of bucket-ratio random variables	123
A.1.2	Derivation of Equation 3.3.4	125
A.1.3	Derivation of Equation 3.3.5	125
A.1.4	Proof of Observation 3.1	125
A.1.5	Proof of Lemma 3.1	126
A.1.6	Proof of Proposition 3.1	130
A.1.7	Generalization of Equation 3.3.2 and Lemma 3.1 to bucket ratios	132
B	Flux Probing Theory	136
B.1	Preliminaries	136
B.2	Proofs and Derivations for Negligible Dead Time	137
B.2.1	Proof of Proposition 5.1	137
B.2.2	Flux probing operation (proof of Proposition 4.1)	138
B.2.3	Maximum frequency (proof of Proposition 4.2)	140
B.2.4	Distribution of probing measurements (proof of Proposition 4.3)	141
B.2.5	Covariance of probing measurements	143
B.2.6	Distribution of Fourier probing (proof of Corollary 4.1)	145
B.2.7	Distribution of Fourier probing energy (proof of Corollary 4.2)	146
B.2.8	Unbiased estimators from probing	147
B.2.9	Frequency detection	149

B.3	Generalization Proofs for Non-Negligible Dead Time	150
B.3.1	Generalized flux probing equation (proof of Proposition 5.2)	150
B.4	Derivations using Monte Carlo estimators	154
B.4.1	Alternate proof of Proposition 4.3	155
B.4.2	Alternate proof of Proposition B.1	157
C	Opportunistic Single-Photon Time of Flight	160
C.1	Proofs and Derivations	160
C.1.1	Distribution of periodic flux estimator (proof of Proposition 6.1)	160
C.1.2	Proof of Corollary 6.1	161
C.1.3	Derivation of CFAR sinc detector threshold	162
	Bibliography	164

1 Introduction

We live in an ever-changing world where events unfold over vast timescales. In just one minute, our hearts beat sixty to a hundred times, and a hummingbird flaps its wings thousands of times. Instruments vibrate hundreds of times a second to make beautiful melodies. Cars move from piston motion caused by the engine's internal combustion, which occurs within a few milliseconds. Proteins responsible for many biological and cellular processes fold at the microsecond to millisecond timescale. At the picosecond timescale, light propagation is slow enough to be observable. Observing these processes is crucial for our scientific understanding of the world.

Even though the human visual system is ill-equipped to observe events across all these different timescales, cameras have proved to be effective proxies for our eyes in scientifically studying many phenomena. Technological advancements have enabled cameras to capture images at extremely high speeds. Modern cameras are remarkably fast. For instance, a smartphone camera can capture 240 frames a second [71, 11, 210], while commercial cameras can reach frame rates of a million per second under natural illumination [201]. Using controlled lighting and specialized optics, cameras can capture at frame rates from billions to hundreds of trillions [261, 66, 129] per second, revealing light-matter interaction that can be studied. Beyond speed, recording images with high precision provides photographic evidence of observations—something concrete that can be shared and analyzed for scientific advancement; some examples include photographic plates used to study animal locomotion [153] and photo 51 used to discover the double helix structure of DNA [63]. Using cameras to observe events is therefore crucial for scientific progress.

Conventional cameras capture images by integrating incoming light over a time interval called the exposure time. This exposure time also limits the frame rate of a camera, as faster frame rates require shorter exposure times. One key drawback of conventional cameras is that they only provide a single cumulative measurement of light within the exposure time, and as a result, events occurring at timescales shorter than the exposure time become motion-blurred or aliased and therefore

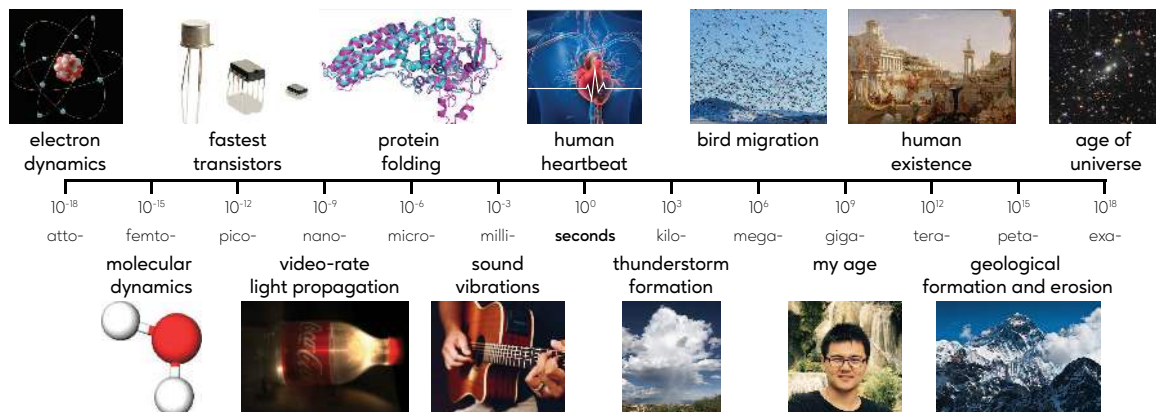


Figure 1.0.1: Timeline spanning over 36 orders of magnitude in time from attoseconds (10^{-18} seconds) to exaseconds (10^{18} seconds) showing different phenomena at their approximate timescales.

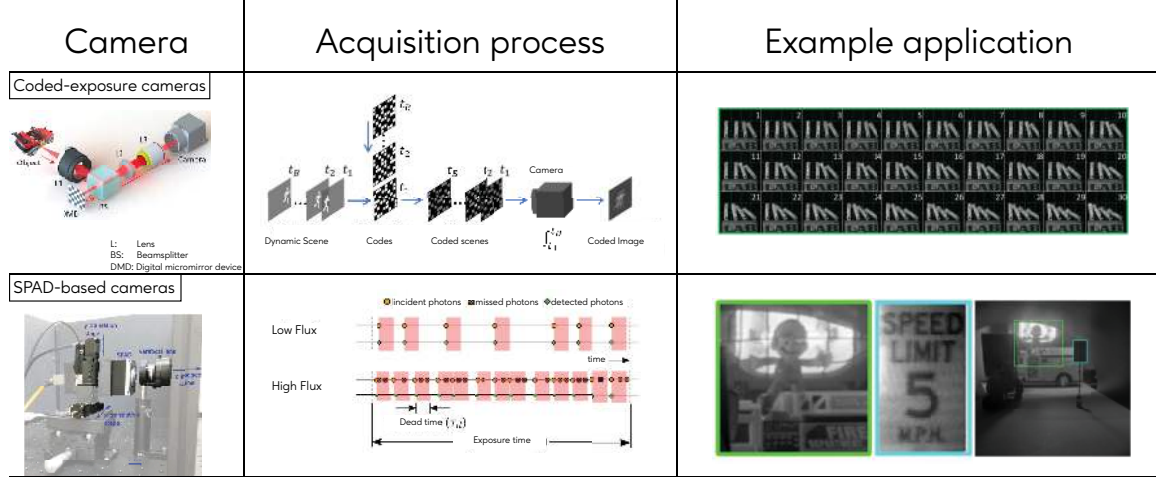


Figure 1.0.2: Some examples of computational cameras. **Coded-exposure cameras** output a coded image by applying codes to the scene. A video of a falling domino is recovered from a single coded image. **SPADs** output the time stamps of detected photons. A dead time occurs after each photon detection, during which incident photons are missed. High-dynamic range imaging with a SPAD. Figures taken from [184, 95, 94].

cannot be observed [166, 196, 176, 276]. As a result, recording events that occur at a particular timescale requires using cameras whose frame rates match or exceed the timescale to avoid aliasing. However, increasing the frame rate requires a proportional increase in the light level of the scene to maintain image quality and the data bandwidth for the same recording time. Therefore, to facilitate high frame rates, high-speed cameras limit their recording time and reduce their image resolution to reduce the data bandwidth. For example, the Phantom TMX 7510, one of the fastest commercially-available high-speed cameras, can record at a frame rate of 1.75 million frames per second but only for 4.6s and with an image resolution of 640×64 [201], both of which are significantly lower when compared to the recording time and image resolution, that are on the orders of minutes and megapixels, of even a regular smartphone camera that captures at 30 frames per second. Overall, conventional cameras are unsuited for recording scenes with events occurring at different timescales.

Recent technological advancements have led to the emergence of *computational cameras* that combine optics, sensing technology and computational techniques to process and capture information about the scene in unprecedented new ways (Figure 1.0.2). Among these advancements are cameras that simultaneously capture scenes across multiple timescales, enabling observations of various phenomena. For example, coded-exposure-imaging (CEI) cameras enable the use of codes that serially modulate the incident wavefront to capture coded images that can be decoded to recover videos [92, 200], allowing for two discrete timescales to be imaged together. Sheinin *et al.* [223] combined a global- and a rolling-shutter camera to recover sound in the scene optically. Velten *et al.* [250] recorded images from a streak camera for many hours to then computationally splice together a transient video of light propagating through a Coca-Cola bottle, lasting only 130 ps. These examples highlight the potential of these computational cameras to reveal details and dynamics that are invisible to or difficult to capture for conventional imaging methods.

The main contribution of this thesis is laying the mathematical groundwork for the control and operation of computational cameras to image across different timescales. The thesis proposes two

sets of frameworks—**(a)** one that optimizes the programming and control for a new class of CEI camera called *coded two-bucket (C2B)* cameras and **(b)** one that enables imaging across a much larger and continuous range of timescales by making use of single-photon cameras. These frameworks push the boundary regarding what these cameras can achieve, and we empirically demonstrate several never-seen-before imaging capabilities. In summary, we make the following contributions:

- a generalized form of optimal multiplexing for coded two-bucket imaging with applications in photometric stereo and structured-light triangulation;
- a new imaging regime called *passive ultra-wideband imaging* that enables imaging across a wide range of timescales spanning over nine orders of magnitudes from 10 frames per second to 250 billion frames per second;
- a mathematical theory that relates a stream of photon detections output from single-photon devices (SPDs) to its underlying photon flux governing photon arrivals; and
- a method for detecting light-source-specific timescales for imaging that enables visualization of multiple, asynchronously propagating light pulses and a generalized form of 3D lidar that removes the need for synchronization and co-located light sources.

1.1 Thesis roadmap

The following chapters of this thesis are organized as follows: **Chapter 2** defines radiometric terms and the cameras relevant to later chapters. **Chapter 3** discusses our work on coded two-bucket imaging using the C2B camera and optimizing its control. **Chapter 4** explores passive ultra-wideband imaging using single-photon cameras and demonstrates several novel imaging capabilities made only possible due to the extreme range of observable timescales in this regime. **Chapter 5** introduces our flux probing theory that establishes a fundamental connection between timestamp arrival data and the underlying photon flux function using tools from stochastic calculus. **Chapter 6** presents our work on opportunistic time-of-flight imaging that harvests and processes ambient photons from asynchronous time-of-flight light sources that pulse at different timescales for joint 3D recovery and light source localization. **Chapter 7** concludes this thesis by addressing current limitations and discussing potential avenues for future research.

2 Image Formation and Cameras

Digital cameras capture and measure light in various ways. Conventional cameras, for example, record the total amount of light for an exposure time, hyper-spectral cameras capture spectral information across a wide range of wavelengths, and polarization cameras capture polarization information about light in a scene. Conceptually, a digital camera can be viewed as a system that takes in light, alongside device-specific parameters such as exposure time and focal length, and maps these inputs to an array of digital signals, or more precisely, an image that represents specific characteristics of light at each pixel.

In this thesis, we use cameras to capture visual information that varies across different timescales. For this reason, the most relevant quantity to measure is the *radiant energy* at every pixel rather than other properties of light, such as polarization or wavelength. Radiant energy is a radiometric term that quantifies the total amount of light, in terms of energy, received by a sensor and can be interpreted as a measure of scene brightness. For this thesis, measuring radiant energy should be understood to mean that the camera’s sensor produces a quantifiable response—in the form of a digital signal and by extension, an image—that is proportionate to the amount of radiant energy received at each pixel, rather than the exact amount of radiant energy. This distinction reflects our emphasis on using these measurements to observe events rather than to calibrate camera response radiometrically.

Imaging at a particular timescale amounts to measuring radiant energy over time intervals on the same order of magnitude as that timescale. For instance, imaging at the picosecond timescale refers to the ability to measure radiant energy over time durations of a few picoseconds. We define *imaging across multiple timescales* then to be the *simultaneous* measurements of radiant energy over sub-intervals of the exposure time across *two or more* distinct timescales. More simply put, a camera capable of imaging across multiple timescales can capture visual information of the same scene at different temporal resolutions.¹ In this thesis, we specifically use two types of cameras for imaging across different timescales: CEI cameras and single-photon avalanche diode (SPAD)-based cameras.

This chapter begins with a precise definition of radiant energy and an overview of the image formation process in conventional cameras. It then discusses the operating principles of CEI cameras and SPADs.

2.1 Radiant Energy

In order to sense light, each pixel in a camera contains a photosensitive area that is exposed to the incoming light. The amount of light falling onto this area is known as the *irradiance* I_e [93] measured

¹Our definition of imaging across multiple timescales excludes cameras that can capture at different frame rates, perform burst photography, or capture exposure brackets, among others.

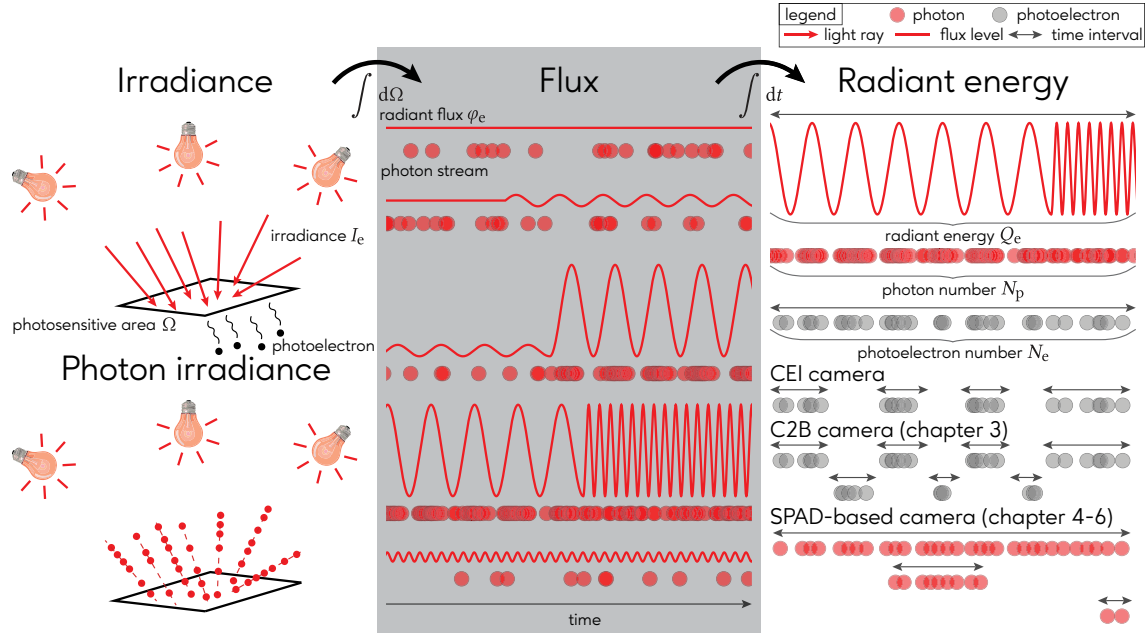


Figure 2.1.1: Relationship between various radiometric and photon quantities. **Left:** Light falling onto a surface is measured as irradiance, and the photosensitive area converts light into photoelectrons. We can also think of light as streams of photons striking the sensor, characterized by the photon flux. **Middle:** Radiant flux, the integral of irradiance over a surface. Radiant flux can vary over time, and different flux levels affect photon arrival. We can see how photon arrival correlates with the peaks of sinusoidal flux and varying flux levels. When flux is extremely low, fast variations in the flux cannot be observed in the photon arrival. **Top right:** Flux integrated over a time interval becomes radiant energy. Similarly, photon flux integrates to photon number. Modern sensors convert light into photoelectron number, with some conversion loss due to quantum efficiency, and then outputs a digital signal based on the photoelectron number. **Bottom right:** We show how various cameras measure radiant energy. CEI cameras measure radiant energy with photoelectron number across disjoint time intervals. C2B cameras (Chapter 3) measure radiant energy with photoelectron number across two sets of complementary time intervals. SPAD-based cameras (Chapters 4, 5, and 6) measure radiant energy with photon number of detected photons across varying time intervals.

in watts per square metre $[\text{W}/\text{m}^2]$,² and the total power received across the entire photosensitive region is referred to as the *radiant flux* ϕ_e , measured in watts $[\text{W}]$.³ Finally, the *radiant energy* Q_e , measured in joules $[\text{J}]$, is defined to be the time integral of the radiant flux over a given time interval $[a, b]$ [2]:

$$Q_e = \int_a^b \phi_e(t) dt. \quad (2.1.1)$$

Figure 2.1.1 shows a detailed overview of the relationship between these radiometric terms.

Most modern digital cameras convert incident light into *photoelectrons* using CCD (charged-coupled devices) [30, 241] or CMOS (complementary metal-oxide semiconductor) [59, 62] chips. These photoelectrons are then collected and read out as a digital number (or pixel intensity) for each pixel. Strictly speaking, pixels measure the number of photoelectrons generated by the incident light, not the radiant energy.

To relate the number of photoelectrons to the radiant energy, we treat incident light as a stream of photons, elementary particles, that strike the sensor. Each photon carries a discrete amount of

²We use 'e' as a subscript to denote radiometric or energetic quantities.

³The ISO symbol for radiant flux is Φ_e [2], but we use ϕ_e here for consistency with the later thesis chapters.

energy, i.e., the *photon energy* Q_p ,⁴ determined by the Planck-Einstein equation [2]:

$$Q_p = \frac{hc}{\lambda}, \quad (2.1.2)$$

where h is the Planck constant, defined to be $6.626\,070\,15 \times 10^{-34}$ joule-seconds [J·s], c is the speed of light, defined to be 299 792 458 metres per second [m/s], and λ is the photon's wavelength in metres [m]. When a photon hits the sensor, it can generate a photoelectron via the photoelectric effect [52]. For monochromatic light with a wavelength of λ , the radiant energy received by a pixel is directly proportional to the photoelectron number N_e during a given time interval [2]:

$$N_e = \eta(\lambda)N_p = \eta(\lambda)\frac{Q_e}{Q_p}, \quad (2.1.3)$$

where N_p is the incident photon number and $\eta(\lambda)$ is the quantum efficiency of a pixel that describes the ratio between the number of incident photons with a given wavelength λ and the number of generated photoelectrons.⁵

Pixels typically receive light composed of multiple wavelengths. For instance, the visible spectrum consists of light in the spectral band from 380 nm to 780 nm [1, term number 845-21-003].⁶ In such cases, the relationship between radiant energy and the number of photoelectrons becomes more complex, as photons with different wavelengths carry different amounts of energy. To simplify this, we approximate the relationship by assuming a mean wavelength $\bar{\lambda}$ for all incident photons [47]:

$$Q_e \approx \frac{hc}{\eta(\bar{\lambda})\bar{\lambda}}N_e. \quad (2.1.4)$$

Given Eq. (2.1.4), we can treat the photoelectron number itself as an approximate measure of radiant energy. By connecting photoelectron number to radiant energy, we can simplify the discussion to focus on how radiometric quantities, such as radiant energy and irradiance, affect sensor output, without considering the details of the photoelectron conversion process.

2.2 Conventional Cameras

Conventional cameras convert radiant energy into digital numbers, or intensities, for each pixel to form images. This conversion process introduces new sources of noise resulting in the following image formation model for the image intensity i :

$$i = \min\left(q\left(\frac{\alpha Q_e}{g} + n\right), i_{max}\right), \quad (2.2.1)$$

⁴We use 'p' as a subscript to denote photon quantities.

⁵Through the thesis, I also refer to photon number as photon counts so they should be treated synonymously.

⁶As stated by the IEC [1], "there are no precise limits for the spectral range of visible radiation since they depend upon the amount of radiant flux reaching the retina and the responsivity of the observer. The lower limit is generally taken between 360 nm and 400 nm and the upper limit between 760 nm and 830 nm". Since wavelengths above 780 nm are typically considered infrared [1, term number 845-21-004], we use 780 nm as our limit.

where α is the conversion factor from radiant energy to photoelectron number described in Eq. (2.1.4), g is the gain used to amplify the charge of photoelectrons, n is the noise, q is the quantization, and i_{\max} is the saturation level. Note that Eq. (2.2.1) only applies to the RAW images output by the sensor because modern digital cameras are equipped with image signal processors (ISPs) that apply a slew of transformations such as white balancing, tone mapping, denoising, *etc.*, that all affect noise in the final output image in a device-specific way. We can approximately decompose the sources of noise into four additive term⁷—shot noise n_{shot} , thermal noise n_{thermal} , read noise n_{read} , and quantization noise $n_{\text{quantization}}$ [84, 127]:

$$i \approx \frac{\alpha Q_e}{g} + n_{\text{shot}} + n_{\text{thermal}} + n_{\text{read}} + n_{\text{quantization}}. \quad (2.2.2)$$

The shot noise results from the fact that incident photon arrivals are stochastic in nature. In particular, the photoelectron number can be modelled as a Poisson distribution with parameter αQ_e [169]. When the photon counts are large, n_{shot} can be approximated as a normal random variable whose standard deviation is signal dependent, *i.e.* $n_{\text{shot}} \sim \mathcal{N}(0, \alpha Q_e / g^2)$ [56].

For many cameras, photoelectrons are generated even in the absence of light. This is due to the thermal energy in the system, which can be reduced by cooling the camera. As a result, this form of noise is aptly referred to as thermal noise. Similarly to incident photon arrivals, the photoelectron number due to dark current, N_{dc} can also be modelled as a Poisson random variable, $N_{\text{dc}} \sim \text{Pois}(\varphi_{\text{dc}} \cdot t)$ where φ_{dc} is the dark current rate and t is the exposure time. Then, $n_{\text{thermal}} = N_{\text{dc}} / g$. Under low-flux settings, thermal noise can dominate the shot noise.

During the charge transfer from pixel to the camera, noise is introduced as a result of imperfections in the camera's electronics. This noise is called read noise and is signal-independent zero-mean Gaussian noise, $n_{\text{read}} \sim \mathcal{N}(0, \sigma_{\text{read}}^2)$.

During the analog-to-digital conversion process, the signal must be quantized to discrete integer values. The rounding error introduced in this case is called quantization noise and can generally be modelled as a uniform distribution over the range $[-0.5, 0.5]$. Therefore, it has a mean equal to 0 with a variance equal to $1/12$. Typically, the quantization noise is significantly weaker than the other sources of noise.

2.3 Coded-Exposure Imaging Cameras

Coded-exposure imaging (CEI) cameras capture images by using controllable masks or "codes" to modulate the incident light on the sensor. The advantage here is that many masks can be applied within a single exposure, enabling these cameras to capture coded images, an example of which we show in Figure 2.3.1. As a result, the codes provide a great degree of freedom in the images CEI cameras can capture compared to conventional cameras. Furthermore, a judicious choice of masks can be used to capture coded images that encode additional information about the scene beyond light intensity.

⁷There is also fixed pattern noise which results from defects in the manufacturing process which we ignore here as they are often suppressed or corrected before image acquisition.



Figure 2.3.1: Capturing a coded image. (a) Subframes of the scene during exposure. (b) Sequence of codes used during exposure; each code lasts for one subframe. (c) The codes block specific pixels from collecting light during each subframe so the camera sees a partly obscured view of the scene. (d) Coded image output at the end of the exposure.

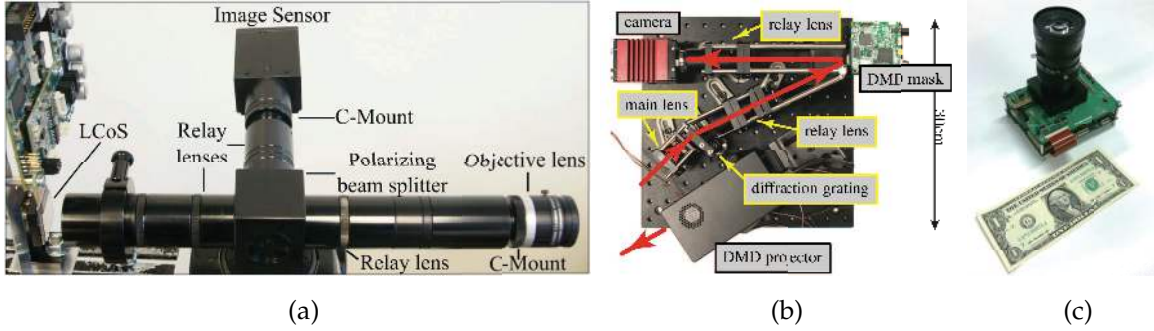


Figure 2.3.2: Optical set-ups of different CEI cameras. (a) Masking using spatial light modulators [92]. (b) Masking using digital micromirror devices [173]. (c) In-pixel electronic masking (row- and column-wise programmable) [275].

Initial CEI cameras used physical masks such as spatial light modulators or digital micromirror devices [159, 196, 92, 200] placed in the optical path between the scene and the image sensor. These cameras were bulkier due to the additional optical elements required to implement coding, as shown in Figures 2.3.2(a) and 2.3.2(b). Recent advancements have allowed for in-pixel electronic masking [275, 91] as shown in Figure 2.3.2(c), although not all sensors are fully programmable, i.e., pixels cannot be set individually. For example, the sensor by [275] only allows codes for pixels to vary along columns and rows. More recently, CEI cameras with per-pixel electronic masking have been developed reducing their form factor significantly.

In general, we can model the resulting radiant flux $\varphi_e^{(c)}(t)$ on the sensor due to the coding capability of CEI cameras by treating it as a multiplication between the incident radiant flux $\varphi(t)$ and the masking or "coding" function $m(t)$ [240]:

$$\varphi_e^{(c)}(t) = m(t)\varphi_e t, \quad (2.3.1)$$

where $0 \leq m(t) \leq 1$ for all $t \in \mathbb{R}$. While the formulation in Eq. (2.3.1) is quite general, in practice, the masking function $m(t)$ is defined over a subdivision of the exposure time into F smaller exposure time called sub-frames wherein $m(t)$ is constant within each sub-frame. In this case, it is more apt to treat $m(t)$ as a F -dimensional vector \mathbf{m} . Therefore, the radiant energy incident to a pixel is:

$$Q_e^{(c)} = \sum_{k=1}^F \mathbf{m}[k] Q_e(k), \quad (2.3.2)$$

where $Q_e(k)$ is the radiant energy of the exposure period corresponding to the k^{th} sub-frame. The sources of noise for CEI cameras remain the same as those of conventional cameras so Eqs. 2.3.2 and 2.2.2 yield the following noise model for the coded image i_c :

$$i_c \approx \sum_{k=1}^F \mathbf{m}[k] \mathbf{i}[k] + n_{\text{shot}} + n_{\text{thermal}} + n_{\text{read}} + n_{\text{quantization}}, \quad (2.3.3)$$

where $\mathbf{i}[k] = \frac{\alpha Q_e(k)}{g}$. Equation 2.3.3 tells us that we can treat the coded image i_c output by CEI cameras as the linear combination of F images $\mathbf{i}[1], \mathbf{i}[2], \dots, \mathbf{i}[F]$. In principle, we could capture each of the F images separately by using a high-speed camera, but this has several disadvantages [174]. First, we would incur read noise F times, once for each captured images as opposed to once for the entire coded image. Furthermore, applying F different masks within an exposure is much quicker than reading out F times.

Certain masks, such as random codes [154, 275], enable recovery of each image $\mathbf{i}[k]$, which measures the radiant energy over time intervals F times shorter than the exposure time. Additionally, their sum $\sum_{k=1}^F \mathbf{i}[k]$ measures the radiant energy over the exposure time. As a result, CEI cameras satisfy our definition of cameras that image across multiple timescales, particularly across two discrete timescales.

CEI cameras can recover additional measurements about the scene beyond just the pixel intensity depending on the codes used, as seen in Figure 2.3.3. By coding the incoming photons, temporal variations in the scene become multiplexed spatially, allowing more measurements to be encoded in the coded image. Coded-exposure imaging methods then decode the image to recover the task-specific unknowns. Different codes enable CEI cameras to be used for various applications. For example, by spatially varying the amount of light each pixel in a neighbourhood sees, the captured coded image can recover an image with a much greater dynamic range [159, 160, 101]. Direct-indirect imaging of scenes [173, 170, 111] can be achieved using codes that block corresponding epipolar lines with a projector. Examples of applications with the aforementioned codes are shown in Figure 2.3.3.

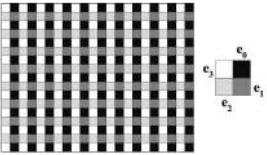


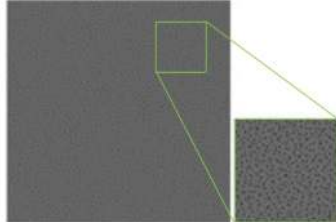
Codes	Example application		
Spatially Varying Exposure 	High-dynamic range image (reconstructed)	High exposure image (conventional)	Low exposure image (conventional)
Random 	Frame 1	Frame 9	Frame 18
Epipolar 	Scene under ambient light	3D from conventional imaging	3D from indirect-invariant imaging
Learned 	High-dynamic range image (reconstructed)	High exposure image (conventional)	Low exposure image (conventional)

Figure 2.3.3: Some examples of codes. **Spatially varying exposure** codes [159] allow different exposures to be sampled locally enabling a high-dynamic range image of the scene to be captured. **Random** codes [92] enable compressive video sensing, allowing a video to be recovered from a single coded image. An example of a falling baseball is provided. **Epipolar** codes [173] leverage epipolar geometry between a projector and CEI camera to block out indirect light paths, i.e. light paths with two or more bounces. These codes enable 3D measurements to be captured even in the presence of reflective surfaces like the mirror. **Learned** codes [145] allow us to use task-driven objective functions to determine what codes to use and improve performance. An example of HDR imaging with learned codes is provided. Figures are from their respective paper.

2.4 SPAD-based Cameras

Single-photon avalanche diodes (SPADs) are highly sensitive semiconductor devices capable of detecting individual photons. By leveraging impact ionization—whereby a single photoelectron can trigger an avalanche of additional photoelectrons due to weakly-bonded electron pairs in the photo-sensitive region—SPADs convert incident photons into discrete, measurable electrical pulses. The increased sensitivity to single photons confers many advantages to SPAD-based cameras compared to conventional cameras, such as the absence of read noise, high-speed imaging, and low-light sensitivity. Among the various operational modes of SPADs, we focus on time-stamping mode, where paired time-to-digital converters (TDC) convert the discrete electrical pulses into timestamps with extremely high temporal precision. In this mode, SPADs output a timestamp stream, which encodes both the number of detected photons and their corresponding arrival times. The high temporal precision has enabled many imaging applications such as 3D lidar [77, 225, 106], non-line-of-sight (NLOS) imaging [34, 270, 172], transient imaging [126, 171], and fluorescence lifetime imaging (FLIM) [43, 285].

2.4.1 Photon number

Unlike conventional and CEI cameras, which measure radiant energy through photoelectrons that accumulate over a fixed exposure time, timestamp streams $\mathcal{T} = \{\tau\}$ offer a much richer set of information about radiant energy in the scene. Since we treat the photoelectron number as a proxy for radiant energy (Eq. (2.1.4)), we can similarly use photon number as a proxy. Furthermore, since we also have the timestamp for each photon detection, we can measure the photon number $N_p(s, t)$ for any time interval $[s, t]$ by counting the photon detections:

$$N_p(s, t) = \sum_{\tau \in \mathcal{T}} \mathbb{1}_{\tau \geq s \ \& \ \tau \leq t}, \quad (2.4.1)$$

where $\mathbb{1}_X = 1$ if X is true and 0 otherwise. This capability to measure photon number across different time intervals satisfies our definition of imaging across different timescales, and indeed, SPADs have been used for imaging at variable frame rates [218]. While counting detections can measure photon number over varying time intervals, this approach is fundamentally constrained by the photon arrival rate. At low flux levels, photon arrivals are incredibly sparse, so the photon number is 0 for many time intervals. Meanwhile, at high flux levels, the dead time of SPADs limits the maximum photon arrival rate to a few million per second, limiting the time intervals to tens or hundreds of nanoseconds, depending on the SPAD device.⁸ This counting-based approach neglects a key advantage of SPADs: their exceptional temporal resolution.

Chapters 4 and 5 introduce the flux probing theory that allows us to measure the flux from timestamp streams by leveraging the SPAD’s high temporal resolution. We can then measure the photon number by integrating the flux rather than counting photon detections. This approach enables photon number to be measured over many time intervals including at the picosecond scale, which is several orders of magnitude shorter than the counting-based approach can achieve.

⁸The dead time is a brief, refractory period after detecting a photon during which the SPAD cannot detect additional photons. We will discuss its effect in greater detail in Chapter 4. For the SPAD device used in this thesis, the dead time was around 250 ns.

3 Imaging Two Time Scales with Coded Two-Bucket Cameras

3.1 Introduction

New camera designs—and new types of imaging sensors—have been instrumental in driving the field of computer vision in exciting new directions. In the last decade alone, time-of-flight cameras [114, 19] have been widely adopted for vision [161] and computational photography tasks [86, 99, 227, 35]; event cameras [121] that support asynchronous imaging have led to new vision techniques for high-speed motion analysis [104] and 3D scanning [147]; high-resolution sensors with dual-pixel [97] and assorted-pixel [273] designs are defining the state of the art for smartphone cameras; and sensors with pixel-wise coded-exposure capabilities are starting to appear [281, 233] for compressed sensing applications [20].

Against this backdrop, we introduce a new type of computational video camera to the vision community—the *coded two-bucket (C2B) camera* (Figure 3.1.1). The C2B camera is a pixel-wise coded-exposure camera that never blocks the incident light. Instead, each pixel in its sensor contains two charge-collection sites—two “buckets”—as well as a one-bit writeable memory that controls which bucket is active. The camera outputs two images per video frame—one per bucket—and performs exposure coding by rapidly controlling the active bucket of each pixel, via a programmable sequence of binary 2D patterns. Key to this unique functionality is a novel programmable CMOS sensor that we designed from the ground up, fabricated in a CMOS image sensor (CIS) process technology [60] for the first time, and turned into a working camera system.

The light efficiency and electronic per-pixel coding capabilities of C2B cameras open up a range of applications that go well beyond what is possible today. This potentially includes compressive acquisition of high-speed video [92] with optimal light efficiency; simultaneous acquisition of both epipolar-only [173] and non-epipolar video streams; fully-electronic acquisition of high-dynamic-range AC-flicker videos [224]; conferring EpiScan3D-like functionality [170] to non-rectified imaging systems; and performing many other coded-exposure imaging tasks [20, 88, 196] with a compact camera platform.

Our focus here, however, is to highlight the novel capabilities of C2B cameras for live dense one-shot 3D reconstruction: we show that from just one grayscale C2B video frame of a dynamic scene under active illumination, it is possible to reconstruct the scene’s 3D snapshot (*i.e.*, per-pixel disparity or normals, plus albedo) at a resolution comparable to the sensor’s pixel array. We argue that C2B cameras allow us to reduce this very difficult 3D reconstruction problem [89, 103, 64, 246, 207, 157] to the potentially much easier 2D problems of image demosaicing [69, 87] and illumination multiplexing [215].

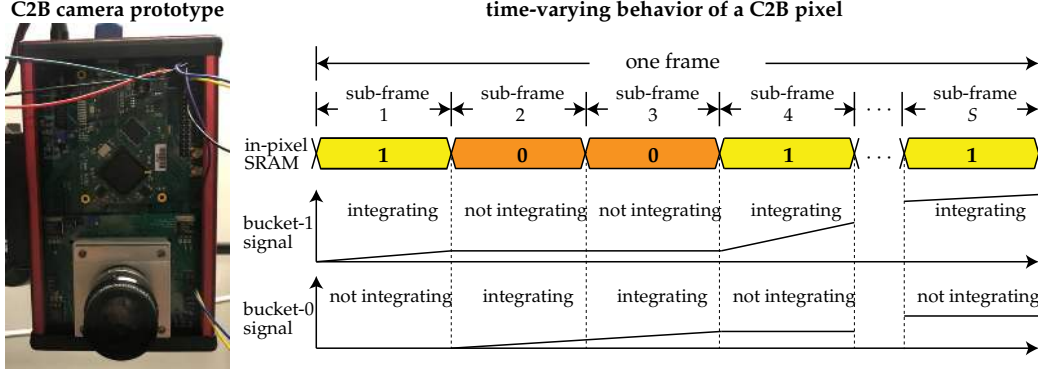


Figure 3.1.1: *The C2B camera.* **Left:** Our prototype’s sensor outputs video at 20 frames per second and consists of two arrays: a 244×160 -pixel array that supports relatively slow bucket control (up to 4 sub-frames per frame) and a 35×48 array with much faster control (up to 120 sub-frames per frame). **Right:** Each frame is divided into S sub-frames during which the pixel’s SRAM memory remains unchanged. A user-specified sequence of 2D binary patterns determines the SRAM’s value at each pixel and sub-frame. Note that the two buckets of a pixel are never in the same state (*i.e.*, both active or both inactive) as this would degrade imaging performance—see Section 3.7 for a discussion of this and other related CMOS design issues. The light-generated charges of both buckets are read, digitized and cleared only once, at the end of each frame.

In particular, we show that C2B cameras can acquire—in one frame—images of a scene under $S \geq 3$ linearly independent illuminations, multiplexed across the buckets of $S - 1$ neighboring pixels. We call such a frame a *two-bucket illumination mosaic*. In this setting, reconstruction at full sensor resolution involves four steps (Figure 3.1.2): (1) control bucket activities and light sources to pack $2(S - 1)$ distinct low-resolution images of the scene into one C2B frame (*i.e.*, $S - 1$ images per bucket); (2) upsample these images to full resolution by demosaicing; (3) demultiplex all the upsampled images jointly, to obtain up to S linearly-independent full-resolution images; and (4) use these images to solve for shape and albedo at each pixel independently. We demonstrate the effectiveness of this procedure by recovering dense 3D shape and albedo from one shot with two of the oldest and simplest active 3D reconstruction algorithms available—multi-pattern cosine phase shifting [209, 208] and photometric stereo [267].

From a hardware perspective, we build on previous attempts to fabricate sensors with C2B-like functionality [211, 136, 137], which did not rely on a CMOS image sensor process technology. More broadly, our prototype can be thought of as generalizing three families of sensors. *Programmable coded-exposure sensors* [281] allow individual pixels to be “masked” for brief periods during the exposure of a video frame (Figure 3.1.3, left). Just like the C2B sensor, they have a writeable one-bit memory inside each pixel to control masking, but their pixels lack a second bucket so light falling onto “masked” pixels is lost. *Continuous-wave time-of-flight sensors* [114, 19] can be thought of as having complementary functionality to coded-exposure sensors: their pixels have two buckets whose activity can be toggled programmatically (so no light is lost), but they have no in-pixel writeable memory. As such, the active bucket is constrained to be the same for all pixels (Figure 3.1.3, middle). This makes programmable per-pixel coding—and acquisition of illumination mosaics in particular—impossible without specialized optics (*e.g.*, [92]). *Multi-bucket* (a.k.a., “multi-tap”) sensors [258, 264, 257, 220] have more than two buckets in each pixel but they have no writeable memory either, so per-pixel coding is not possible. In theory, an S -bucket sensor would be uniquely suited for dense one-shot reconstruction because it can acquire in each frame S full-resolution im-

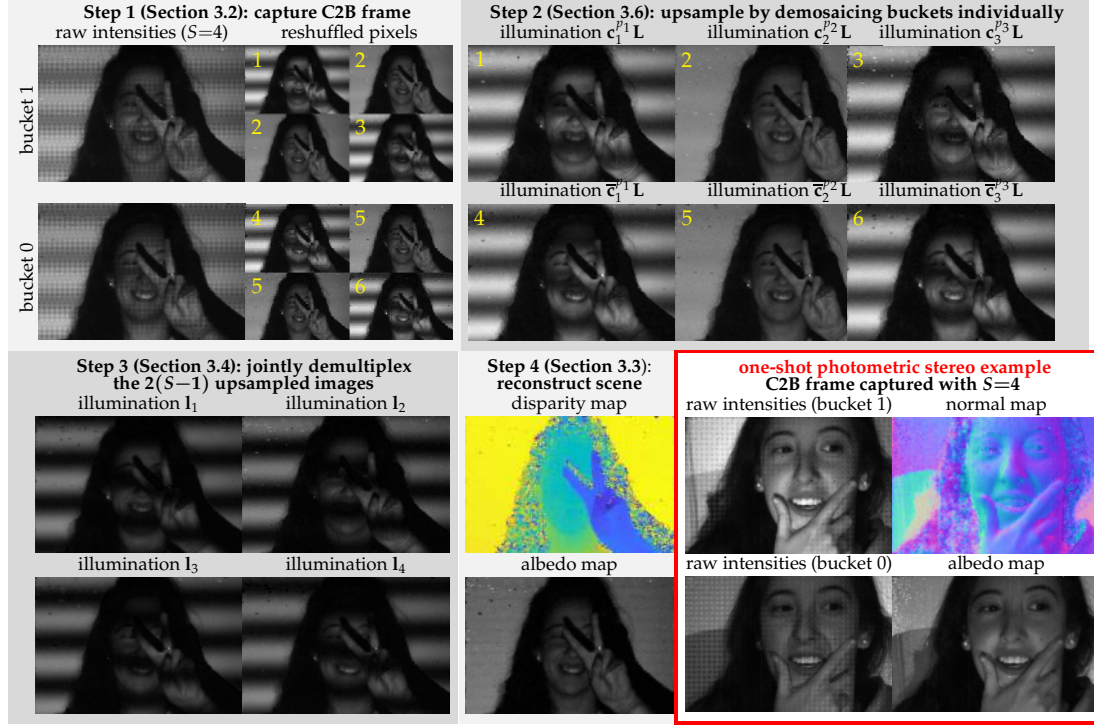


Figure 3.1.2: *Dense one-shot reconstruction with C2B cameras.* The procedure runs in real time and is illustrated for structured-light triangulation. Please zoom in to the electronic copy to see individual pixels of the C2B frame and refer to the listed sections for notation and details. Photometric stereo is performed in an analogous way, by replacing the structured-light projector with a set of S directional light sources (a reconstruction example is shown in the lower right).

ages corresponding to any set of S illuminations [274]. In practice, however, C2B sensors have several advantages: they are scalable because they can pack S linearly-independent images into one frame for any value of S —without hard-wiring this value into the pixel’s CMOS design; they are much more light efficient because each extra bucket reduces the pixel’s photo-sensitive region significantly for a given pixel size; and they have a broader range of applications because they enable per-pixel coding. To our knowledge, 2D sensors with more than four buckets have not been fabricated in a standard CMOS image process, and it is unclear if they could offer acceptable imaging performance.

On the conceptual side, our contributions are the following: (1) we put forth a general model for the C2B camera that opens up new directions for coded-exposure imaging with active sources; (2) we formulate its control as a novel multiplexing problem [263, 198, 33, 215, 197, 167, 150] in the bucket and pixel domains; (3) we draw a connection between two-bucket imaging and algorithms that operate directly on intensity ratios [132]; and (4) we provide an algorithm-independent framework for dense one-shot reconstruction that is simpler than earlier attempts [173] and is compatible with standard image processing pipelines.

Last but not least, we demonstrate all the above experimentally, on the first fully operational C2B camera prototype.

		coded-exposure imaging	continuous-wave time-of-flight imaging		coded two-bucket imaging	
		pixel masks \mathbf{c}_{fs}	bucket-1 masks \mathbf{c}_{fs}	bucket-0 masks $\bar{\mathbf{c}}_{fs}$	bucket-1 masks \mathbf{c}_{fs}	bucket-0 masks $\bar{\mathbf{c}}_{fs}$
sub-frame	$s = 1$					
sub-frame	$s = 2$					
...
sub-frame	$s = S$					

Figure 3.1.3: Comparison of basic sensor abilities. Coded-exposure sensors can rapidly mask individual pixels but cannot collect all the incident light; continuous-wave ToF sensors always collect all the incident light but they cannot mask pixels individually; C2B sensors can do both. The column vectors \mathbf{c}_{fs} and $\bar{\mathbf{c}}_{fs}$ denote bucket-1 masks/activities and their binary complement, respectively.

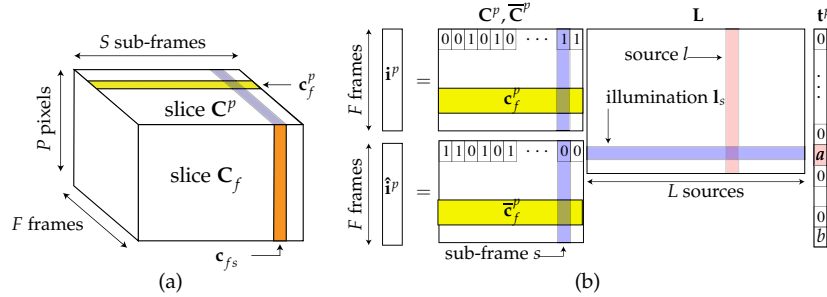


Figure 3.2.1: (a) Structure of the code tensor \mathbf{C} . (b) Image formation model for pixel p . We show the transport vector \mathbf{t}^p for a structured-light setting, where the contribution of ambient illumination is $\mathbf{t}^p[L] = b$, the corresponding projector pixel is l , and its albedo is $\mathbf{t}^p[l] = a$.

3.2 Coded Two-Bucket Imaging

We begin by introducing an image formation model for C2B cameras. We consider the most general setting in this section, where a whole sequence of C2B frames may be acquired instead of just one. C2B cameras output *two* images per video frame—one for each bucket (Figure 3.1.2). We refer to these images as the *bucket-1 image* and *bucket-0 image*.

The code tensor. Programming a C2B camera amounts to specifying the time-varying contents of its pixels’ memories at two different timescales: (1) at the scale of *sub-frames* within a video frame, which correspond to the updates of the in-pixel memories (Figure 3.1.1, right), and (2) at the scale of frames within a video sequence. For a video sequence with F frames and a camera that has P pixels and supports S sub-frames, bucket activities can be represented as a three-dimensional binary tensor \mathbf{C} of size $P \times F \times S$. We call \mathbf{C} the *code tensor* (Figure 3.2.1a).

We use two specific 2D “slices” of the code tensor in our analysis below and have special notation for them. For a specific pixel p , slice \mathbf{C}^p describes the activity of pixel p ’s buckets across all frames and sub-frames. Similarly, for a specific frame f , slice \mathbf{C}_f describes the bucket activity of all pixels

across all sub-frames of f :

$$\mathbf{C}^p = \underbrace{\begin{bmatrix} \mathbf{c}_1^p \\ \mathbf{c}_2^p \\ \vdots \\ \mathbf{c}_F^p \end{bmatrix}}_{F \times S} \quad \mathbf{C}_f = \underbrace{\begin{bmatrix} \mathbf{c}_{f1} & \mathbf{c}_{f2} & \cdots & \mathbf{c}_{fS} \end{bmatrix}}_{P \times S}, \quad (3.2.1)$$

where \mathbf{c}_f^p is an S -dimensional row vector that specifies the active bucket of pixel p in the sub-frames of frame f ; and \mathbf{c}_{fs} is a P -dimensional column vector that specifies the active bucket of all pixels in sub-frame s of frame f .

The illumination matrix. Although C2B cameras can be used for passive imaging applications [20], we model the case where illumination is programmable at sub-frame timescales too. In particular, we represent the scene’s time-varying illumination condition as an illumination matrix \mathbf{L} that applies to all frames:

$$\mathbf{L} = \underbrace{\begin{bmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \\ \vdots \\ \mathbf{l}_S \end{bmatrix}}_{S \times L}, \quad (3.2.2)$$

where row vector \mathbf{l}_s denotes the scene’s illumination condition in sub-frame s of every frame. We consider two types of scene illumination in this work: a set of L directional light sources whose intensity is given by vector \mathbf{l}_s ; and a projector that projects a pattern specified by the first $L - 1$ elements of \mathbf{l}_s in the presence of ambient light, which we treat as an L -th source that is “always on” (*i.e.*, element $\mathbf{l}_s[L] = 1$ for all s).

Two-bucket image formation model for pixel p . Let \mathbf{i}^p and $\hat{\mathbf{i}}^p$ be column vectors holding the intensities of pixel p ’s bucket 1 and bucket 0, respectively, in F frames. We model these intensities as the result of light transport from the L light sources to the pixel’s two buckets (Figure 3.2.1b):

$$\underbrace{\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix}}_{2F \times 1} = \underbrace{\begin{bmatrix} \mathbf{C}^p \\ \bar{\mathbf{C}}^p \end{bmatrix}}_{2F \times S} \underbrace{\mathbf{L}}_{S \times L} \underbrace{\mathbf{t}^p}_{L \times 1}, \quad (3.2.3)$$

where \bar{b} denotes the binary complement of matrix or vector b , \mathbf{C}^p is the slice of the code tensor corresponding to p , and \mathbf{t}^p is the pixel’s *transport vector*. Element $\mathbf{t}^p[l]$ of this vector describes the transport of light from source l to pixel p in the timespan of one sub-frame, across all light paths.

To gain some intuition about Eq. (3.2.3), consider the buckets’ intensity in frame f :

$$\mathbf{i}^p[f] = \underbrace{(\mathbf{c}_f^p \mathbf{L})}_{\text{illumination condition of pixel } p, \text{ bucket 1, frame } f} \mathbf{t}^p \quad \hat{\mathbf{i}}^p[f] = \underbrace{(\bar{\mathbf{c}}_f^p \mathbf{L})}_{\text{illumination condition of pixel } p, \text{ bucket 0, frame } f} \mathbf{t}^p. \quad (3.2.4)$$

In effect, the two buckets of pixel p can be thought of as “viewing” the scene under two potentially different illuminations given by vectors $\mathbf{c}_f^p \mathbf{L}$ and $\bar{\mathbf{c}}_f^p \mathbf{L}$, respectively. Moreover, if \mathbf{c}_f^p varies from frame to frame, these illumination conditions may vary as well.

Bucket ratios as albedo “quasi-invariants.” Since the two buckets of pixel p generally represent different illumination conditions, the two ratios

$$r = \frac{\mathbf{i}^p[f]}{\mathbf{i}^p[f] + \hat{\mathbf{i}}^p[f]} \quad , \quad \hat{r} = \frac{\hat{\mathbf{i}}^p[f]}{\mathbf{i}^p[f] + \hat{\mathbf{i}}^p[f]} \quad , \quad (3.2.5)$$

defined by p ’s buckets are *illumination ratios* [132, 259, 182]. Moreover, we show in Appendix A.1.1 that under zero-mean Gaussian image noise, these ratios are well approximated by Gaussian random variables whose mean is the ideal (noiseless) ratio and whose standard deviation depends weakly on albedo. In effect, C2B cameras provide two “albedo-invariant” images per frame. We exploit this feature of C2B cameras for both shape recovery and demosaicing in Sections 3.3 and 3.6, respectively.

3.2.1 Acquiring two-bucket illumination mosaics

A key feature of C2B cameras is that they offer an important alternative to multi-frame acquisition: instead of capturing F frames in sequence, they can capture a spatially-multiplexed version of them in a single C2B frame (Figure 3.1.2). We call such a frame a *two-bucket illumination mosaic* in analogy to the RGB filter mosaics of color image sensors [25, 156, 273]. Unlike filter mosaics, however, which are attached to the sensor and cannot be changed, acquisition of illumination mosaics is programmable for any F .

The bucket-1 and bucket-0 image sequences. Collecting the two buckets’ intensities in Eq. (3.2.4) across all frames and pixels, we define two matrices that hold all this data:

$$\mathbf{I} = \underbrace{\begin{bmatrix} \mathbf{i}^1 & \mathbf{i}^2 & \dots & \mathbf{i}^P \end{bmatrix}}_{F \times P} \quad \hat{\mathbf{I}} = \underbrace{\begin{bmatrix} \hat{\mathbf{i}}^1 & \hat{\mathbf{i}}^2 & \dots & \hat{\mathbf{i}}^P \end{bmatrix}}_{F \times P} . \quad (3.2.6)$$

Code tensor for mosaic acquisition. Formally, a two-bucket illumination mosaic is a spatial sub-sampling of the sequences \mathbf{I} and $\hat{\mathbf{I}}$ in Eq. (3.2.6). Acquiring it amounts to specifying a one-frame code tensor $\tilde{\mathbf{C}}$ that spatially multiplexes the corresponding F -frame tensor \mathbf{C} in Figure 3.2.1(a). We do this by (1) defining a regular tiling of the sensor plane and (2) specifying a correspondence $(p_i \rightarrow f_i), 1 \leq i \leq K$, between the K pixels in a tile and frames. The rows of $\tilde{\mathbf{C}}$ are then defined to be

$$\tilde{\mathbf{c}}_1^{p_i} \stackrel{\text{def}}{=} \mathbf{c}_{f_i}^{p_i} . \quad (3.2.7)$$

Mosaic acquisition example. The C2B frames in Figure 3.1.2 were captured using a 2×2 -pixel tile to spatially multiplex a three-frame code tensor. The tensor assigned identical illumination conditions to all pixels within a frame and different conditions across frames. Pixels within each tile were assigned to individual frames using the correspondence $\{(1,1) \rightarrow 1, (1,2) \rightarrow 2, (2,1) \rightarrow 2, (2,2) \rightarrow 3\}$.

3.3 Per-Pixel Estimation of Normals and Depth

Let us now turn to the problem of normal and depth estimation using photometric stereo and structured-light triangulation, respectively. We consider the most basic formulation of these tasks, where all computations are done independently at each pixel and the relation between observations and unknowns is expressed as a system of linear equations. These formulations should be treated

	Lambertian photometric stereo	Structured-light triangulation w/ cosine patterns
<i>Assumptions</i>	Lambertian reflectance, non-uniform albedo; calibrated light sources; no ambient or indirect light	reflectance has non-negligible diffuse component; robustness to indirect light depends on frequency choice
<i>Illumination vectors \mathbf{l}_s</i>	each \mathbf{l}_s corresponds to illumination with only source s turned on, i.e., element $\mathbf{l}_s[l]$ is non-zero iff $s=l$	$\mathbf{l}_s[l] = \cos(\phi_s + \theta_l)$, where ϕ_s is phase shift of s -th pattern, θ_l is phase of projector pixel l
<i>Vector \mathbf{d}_s</i>	orientation and intensity of source s , expressed as a 3D row vector	$\mathbf{d}_s = [\cos(\phi_s) \quad -\sin(\phi_s) \quad 1]$
<i>Matrix \mathbf{D}</i>	matrix whose rows are the vectors $\mathbf{d}_1, \dots, \mathbf{d}_S$	matrix whose rows are the vectors $\mathbf{d}_1, \dots, \mathbf{d}_S$
<i>Transport vector \mathbf{t}</i>	$\mathbf{t}[s] = a \mathbf{d}_s \mathbf{n}$ where a is the albedo and \mathbf{n} is the unit surface normal	$\mathbf{t} = [a \mathbf{m} \quad b]^T$, where a is albedo, b is the contribution of ambient light, and binary row vector \mathbf{m} indicates the matching projector pixel, i.e., $\mathbf{m}[l]=1$ iff that pixel is l (see Figure 3.2.1b)
<i>Vector \mathbf{x}</i>	$\mathbf{x} = \mathbf{n}$	$\mathbf{x} = [\cos(\theta) \quad \sin(\theta) \quad \frac{b}{a}]^T$ if the same cosine frequency is used for all patterns; additional frequencies contribute two unknowns each; θ is the phase of the matching projector pixel

Table 3.1: The two basic multi-image reconstruction techniques considered in this work.

merely as examples that showcase the special characteristics of two-bucket imaging; as with conventional cameras, using advanced methods to handle more general settings [185, 78] is certainly possible.

From bucket intensities to demultiplexed intensities. As a starting point, we expand Eq. (3.2.3) to get a relation that involves only intensities:

$$\underbrace{\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix}}_{\substack{\text{bucket measurements} \\ (2F \times 1)}} = \underbrace{\begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix}}_{\substack{\text{bucket-multiplexing} \\ \text{matrix } \mathbf{W} \\ (2F \times S)}} \stackrel{\text{def}}{=} \underbrace{\begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix}}_{\substack{\text{pixel intensity under} \\ \text{illuminations } \mathbf{l}_1, \dots, \mathbf{l}_S \\ (S \times 1)}} \begin{bmatrix} i_1^p \\ \vdots \\ i_S^p \end{bmatrix}. \quad (3.3.1)$$

Each scalar i_s^p in the right-hand side of Eq. (3.3.1) is the intensity that a conventional camera pixel would record if the scene's illumination condition was \mathbf{l}_s . Therefore, Eq. (3.3.1) tells us that as far as a single pixel p is concerned, C2B cameras capture the same S measurements a conventional camera would capture for 3D reconstruction—except that those measurements are multiplexed over $2F$ bucket intensities. To retrieve them, these intensities must be demultiplexed by inverting Eq. (3.3.1):

$$\begin{bmatrix} i_1^p \\ \vdots \\ i_S^p \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}' \begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix}, \quad (3.3.2)$$

where $'$ denotes matrix transpose. This inversion is only possible if $(\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'$ is non-singular. Moreover, the signal-to-noise ratio (SNR) of the demultiplexed intensities depends heavily on \mathbf{W} and \mathbf{C}^p (Section 3.4). Setting aside this issue for now, we consider below the task of shape recovery from already-demultiplexed intensities. For notational simplicity we drop the pixel index p from the equations below.

Per-pixel constraints on 3D shape. The relation between demultiplexed intensities and the pixel's

unknowns takes exactly the same form in both photometric stereo and structured-light triangulation with cosine patterns:

$$\begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = a \mathbf{D} \mathbf{x} + \mathbf{e} , \quad (3.3.3)$$

where \mathbf{D} is known; \mathbf{x} is a 3D vector that contains the pixel's shape unknowns; a is the unknown albedo; and \mathbf{e} is observation noise. See Table 3.1 for a summary of each problem's assumptions and for the mapping of problem-specific quantities to Eq. (3.3.3).

There are (at least) three ways to turn Eq. (3.3.3) into a constraint on normals and depths under zero-mean Gaussian noise. The resulting constraints are *not* equivalent when combining measurements from small pixel neighborhoods—as we implicitly do—because they are not equally invariant to spatial albedo variations:

1. *Direct method (DM)*: treat Eq. (3.3.3) as providing S independent constraints on vector $a\mathbf{x}$ and solve for both a and \mathbf{x} . The advantage of this approach is that errors are Gaussian by construction; its disadvantage is that Eq. (3.3.3) depends on albedo.
2. *Ratio constraint (R)*: divide individual intensities by their total sum to obtain an illumination ratio, as in Eq. (3.2.5). This yields the following constraint on \mathbf{x} :

$$r_l \mathbf{1} \mathbf{D} \mathbf{x} = \mathbf{d}_l \mathbf{x} , \quad (3.3.4)$$

where $r_l = i_l / \sum_k i_k$ and $\mathbf{1}$ is a row vector of all ones. The advantage here is that both r_l and Eq. (3.3.4) are approximately invariant to albedo.

3. *Cross-product constraint (CP)*: instead of computing an explicit ratio from Eq. (3.3.3), eliminate a to obtain

$$i_l \mathbf{d}_k \mathbf{x} = i_k \mathbf{d}_l \mathbf{x} . \quad (3.3.5)$$

Since Eq. (3.3.5) has intensities i_l, i_k as factors, it does implicitly depend on albedo.

Solving for the unknowns. Both structured light and photometric stereo require at least $S \geq 3$ independent constraints for a unique solution. In the DM method we use least-squares to solve for $a\mathbf{x}$; when using the R or CP constraints, we apply singular-value decomposition to solve for \mathbf{x} .

3.4 Code Matrices for Bucket Multiplexing

The previous section gave ways to solve for 3D shape when we have enough independent constraints per pixel. Here we consider the problem of controlling a C2B camera to actually obtain them for a pixel p . In particular, we show how to choose (1) the number of frames F , (2) the number of sub-frames per frame S , and (3) the pixel-specific slice \mathbf{C}^p of the code tensor, which defines the multiplexing matrix \mathbf{W} in Eq. (3.3.1).

Determining these parameters can be thought of as an instance of the *optimal multiplexing* problem [263, 33, 215, 197, 198, 167, 150]. This problem has been considered in numerous contexts before, as a one-to-one mapping from S desired measurements to S actual, noisy observations. In the case of coded two-bucket imaging, however, the problem is slightly different because each

# sub-frames	S=3	S=4	S=5	S=6	S=7
Eq. (3.4.1) bound for $\sigma=1$	0.5556	0.4167	0.34	0.2889	0.2517
Optimal MSE for $\sigma=1$	0.8333	0.4167	0.3778	0.3467	0.3210
Optimal \mathbf{C}^p	1 0 0 0 1 0	1 1 0 0 1 0 1 0 1 0 0 1	1 1 0 0 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1	1 1 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 1 1 0 0 1 0 1	1 1 1 1 1 0 0 1 1 1 0 0 0 1 1 1 0 0 1 1 0 1 0 1 0 1 1 0 1 0 0 1 0 1 0 1 0 0 0 1 0 1

Table 3.2: Optimal matrices \mathbf{C}^p for small S . Note that the lower bound given by Eq. (3.4.1) is attained only for $S = 4$, i.e., for the smallest Hadamard-based construction of \mathbf{C}^p .

frame yields two measurements instead of just one.

The results below provide further insight into this particular multiplexing problem (see Appendix A.1 for proofs). Observation 3.1 implies that even though a pixel's two buckets provide $2F$ measurements in total across F frames, at most $F + 1$ of them can be independent because the multiplexing matrix \mathbf{W} is rank-deficient:

Observation 3.1

$\text{rank } \mathbf{W} \leq \min(F + 1, S)$.

Intuitively, a C2B camera should not be thought of as being equivalent to two coded-exposure cameras that operate completely independently. This is because the activities of a pixel's two buckets are binary complements of each other, and thus not independent.

Corollary 3.1

Multiplexing S intensities requires $F \geq S - 1$ frames.

Corollary 3.2

The minimal configuration for fully-constrained reconstruction at a pixel p is $F=2$ frames, $S=3$ sub-frames per frame, and $S=3$ linearly-independent illumination vectors of dimension $L \geq 3$. The next-highest configuration is 3 frames, 4 subframes/illumination vectors.

We now seek the optimal $(S - 1) \times S$ matrix \mathbf{C}^p , i.e., the matrix that maximizes the SNR of the demultiplexed intensities in Eq. (3.3.2). Lemma 3.1 extends the lower-bound analysis of Ratner *et al.* [198] to obtain a lower bound on the mean-squared error (MSE) of two-bucket multiplexing:

Lemma 3.1

For every multiplexing matrix \mathbf{W} , the MSE of the best unbiased linear estimator satisfies the lower bound

$$\text{MSE} = \frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \geq 2\sigma^2 \frac{(S-1)^2 + 1}{(S-1)S^2} . \quad (3.4.1)$$

Although Lemma 3.1 does not provide an explicit construction, it does ensure the optimality of \mathbf{W} matrices whose MSE is the lower bound. We used this observation to verify the optimality of

# Frames	S=4				S=5				S=6				S=7			
MSE _{sub}	0.9167				1				1.0667				1.1190			
SNR Gain, Eq. (3.5.1)	1.4833				1.6269				1.7541				1.8553			
Sup-optimal \mathbf{C}^p	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
					0	0	0	1	0	0	0	0	1	0	0	0
									0	0	0	1	0	0	0	0
									0	0	0	0	1	0	0	0
													0	0	1	0
													0	0	0	1

Table 3.3: Sup-optimal \mathbf{C}^p used for comparison. MSE and SNR gain for image intensities are computed for $\sigma = 1$ in the minimal configurations ($F = S - 1$).

matrices derived from the standard Hadamard construction [215]:

Proposition 3.1

Let $\mathbf{C}^p = \frac{1}{2}(\tilde{\mathbf{H}} + \mathbf{1})$ where $\tilde{\mathbf{H}}$ is derived from the $S \times S$ Hadamard matrix by removing its row of ones to create an $(S - 1) \times S$ matrix. The bucket-multiplexing matrix \mathbf{W} defined by \mathbf{C}^p is optimal.

The smallest S for which Proposition 3.1 applies are $S = 4$ and $S = 8$. Since our main goal is one-shot acquisition, optimal matrices for other small values of S are also of significant interest. To find them, we conducted a brute-force search over the space of small $(S - 1) \times S$ binary matrices to find the ones with the lowest MSE. These matrices are shown in Table 3.2.

3.5 Empirical SNR Analysis

We considered the cases where $S = \{4, 5, 6, 7\}$ and $F = S - 1$. The case $S = 3$ was not evaluated because all non-degenerate 2×3 matrices \mathbf{C}^p have the same expected SNR. We present the sub-optimal matrices we used as shown in Table 3.3. We evaluated the SNR gain for both image intensities and illumination ratios which is defined in Eq. (3.3.4) from Section 3.3.

3.5.1 Empirical SNR gain for image intensities.

We applied bucket multiplexing to the scene shown in Figure 3.5.1(b) and empirically measured the average SNR of demultiplexed images when (1) \mathbf{C}^p is given by Table 3.3 and (2) $\mathbf{C}^p = [\mathbf{1}_{(S-1) \times (S-1)} \mathbf{0}]$, which is a non-degenerate and sub-optimal matrix according to Proposition 3.1 ($\mathbf{1}_{(S-1) \times (S-1)}$ is the identity matrix). We plot the ratio of these SNRs against the theoretical ratio in Figure 3.5.2(a), suggesting that SNR gains are possible.

To measure average SNR of a given \mathbf{C}^p , we (1) sequentially capture F C2B frames of a static scene using $S = F + 1$ subframes per frame, S structured-light patterns $\mathbf{I}_1, \dots, \mathbf{I}_S$, and the same matrix \mathbf{C}^p for all pixels—see Eq. (3.3.1) in Section 3.3; (2) demultiplex the $2F$ intensities collected in each pixel’s buckets by multiplying with the $S \times 2F$ pseudoinverse of $\begin{bmatrix} \mathbf{C}^p \\ \mathbf{C}^p \end{bmatrix}$ to get intensities i_1^p, \dots, i_S^p for $\mathbf{I}_1, \dots, \mathbf{I}_S$, respectively; (3) repeat steps 1&2 300 times; (4) compute $\frac{\text{mean}}{\text{standard deviation}}$ at each pixel across trials; and (5) average this ratio over all valid pixels in the S demultiplexed images using a binary foreground mask (Figure 3.5.1(b)).

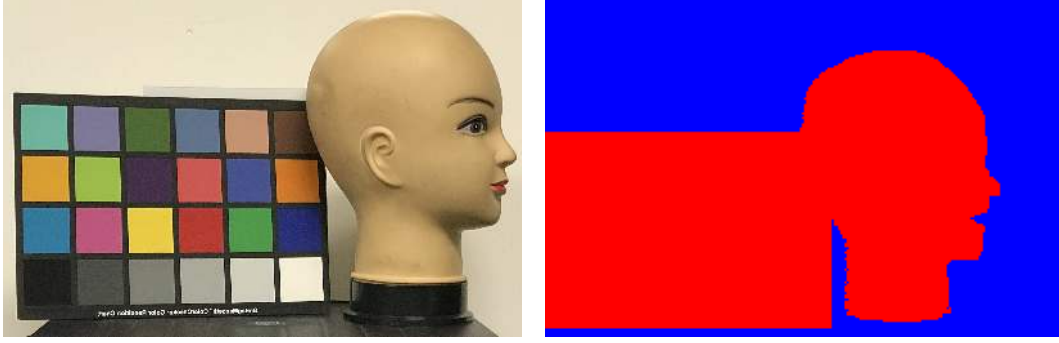


Figure 3.5.1: *Objects used for performance evaluation. (a) Color view of the objects, view not from C2B camera. (b) The binary foreground mask (red is foreground). The mask excludes pixels not belonging to the objects as well as shadows.*

3.5.2 Empirical SNR gain for illumination ratios.

We compute the empirical SNR gain for illumination ratios. We use the same steps as those used to compute the empirical SNR gain for image intensities. The only difference here is that instead of demultiplexing bucket intensities, we demultiplex the bucket ratios.

3.5.3 Theoretical SNR gain.

We define the theoretical SNR gain [47] as:

$$\text{SNR Gain} = \sqrt{\frac{\text{MSE}_{\text{triv}}}{\text{MSE}_{\text{opt}}}}, \quad (3.5.1)$$

where MSE_{triv} and MSE_{opt} are computed using Eq. (3.4.1) from Section 3.4 for the sub-optimal matrix and optimal matrix, respectively. We list the theoretical SNR gains in Table 3.3.

3.5.4 Empirical evaluation of SNR gain for illumination ratios.

We visualize the SNR gain for the illumination ratios in Figure 3.5.2(a). The same multiplexing matrices as those used for evaluating the SNR gain of image intensities are used here, however we use the approximate MSE for illumination ratios defined in Proposition A.3 in the calculation of the SNR gain. We also show examples of demultiplexed illumination ratios with optimal and sub-optimal multiplexing matrices in Figure 3.5.2(b). The gap between the theoretical SNR gain and the empirical SNR gain could be explained by the noise model used. In particular, the theoretical SNR gain does not take into account Poisson noise which might cause the SNR gain to decrease in practice. We can see that similar trends appear for both the empirical SNR gains in illumination ratios and the empirical SNR gains in image intensities (Figure 3.5.2(a)). However, the theoretical and empirical performance gains suggest that while bucket ratios do provide albedo-invariant images, demultiplexing image intensities provides a greater SNR gain than demultiplexing bucket ratios.

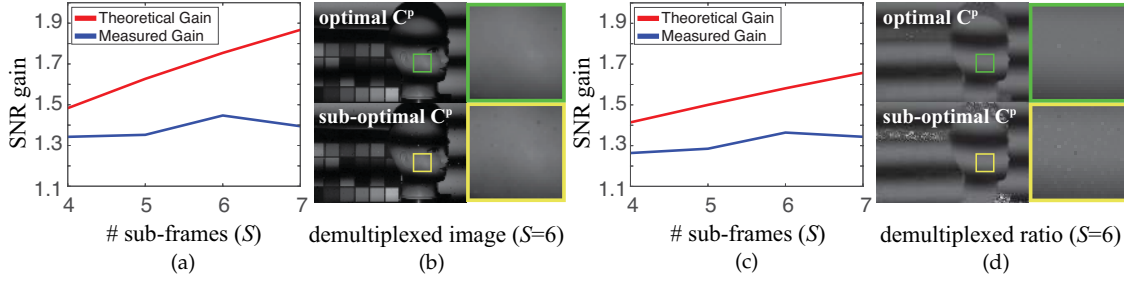


Figure 3.5.2: *Optimal versus sub-optimal multiplexing.* (a) Plot of empirical SNR gains (in blue) and theoretical SNR gains (in red) for image intensities. (b) One out of S demultiplexed images obtained with each C^p . (c) Plot of empirical SNR gains (in blue) and theoretical SNR gains (in red) for illumination ratios. (d) One out of S illumination ratios obtained with each C^p . The optimal C^p yielded visibly less noisy images (please zoom in to the electronic copy).

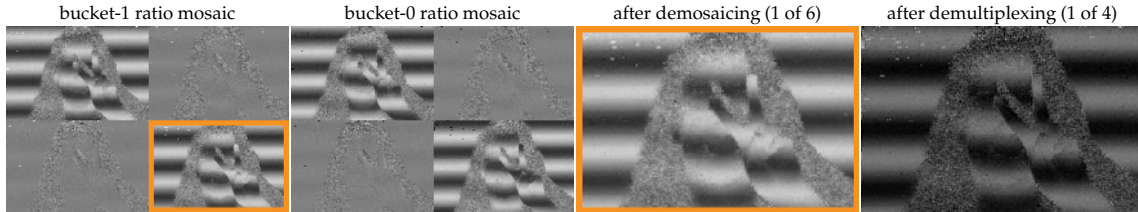


Figure 3.6.1: *Processing ratio mosaics.* Left to right: Intermediate results of the BRD reconstruction procedure of Section 3.6, starting from the raw C2B frame shown in Figure 3.1.2, Step 1. In contrast to the result of Steps 2 and 3 in Figure 3.1.2, the images above are largely unaffected by albedo variations.

3.6 One-Shot Shape from Two-Bucket Illumination Mosaics

We use three different ways of estimating shape from a two-bucket illumination mosaic:

1. *Intensity demosaicing (ID)*: treat the intensities in a mosaic tile as separate “imaging dimensions” for the purpose of demosaicing; upsample these intensities by applying either an RGB demosaicing algorithm to three of these dimensions at a time, or by using a more general assorted-pixel procedure [156, 273] that takes all of them into account; demultiplex the $2F$ upsampled images using Eq. (3.3.2); and apply any of the estimation methods in Section 3.3 to the result. Figure 3.1.2 illustrates this approach.
2. *Bucket-ratio demosaicing (BRD)*: apply Eq. (3.2.5) to each pixel in the mosaic to obtain two albedo-invariant “ratio mosaics”; demosaic and demultiplex them; and compute 3D shape using the ratio constraint of Section 3.3. See Figure 3.6.1 for an example.
3. *No demosaicing (ND)*: instead of upsampling, treat each mosaic tile as a “super-pixel” whose unknowns (*i.e.*, normal, disparity, *etc.*) do not vary within the tile; compute one shape estimate per tile using any of the methods of Section 3.3.

3.6.1 Performance evaluation on synthetic data

Scene details. We set the image resolution to 300×300 . We use orthographic projection for the camera and set the camera centre as the origin. We use a Lambertian model (Eq. (3.3.3) and Table 3.1 from Section 3.3 to render the scene so there are no interreflections and no occlusions in the scene. Each pixel is assumed to be in the line of sight of the light source. We set the mean reflectance intensity to be 80 and the variance to be 400. We use OpenCV’s demosaicing function with

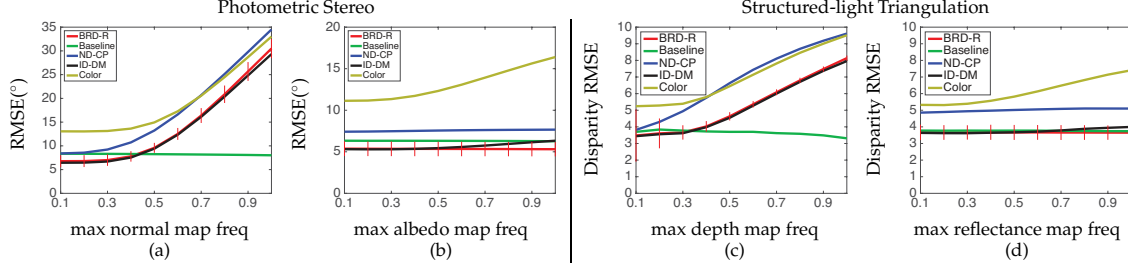


Figure 3.6.2: Performance comparisons of photometric stereo (a-b) and structured-light triangulation (c-d) on synthetic scenes. **(a)** Angular root-mean-squared error (RMSE) of normal estimates as a function of the normal map’s highest spatial frequency. Frequency 1.0 corresponds to the Nyquist limit. The highest spatial frequency of albedos was set to 0.3 the Nyquist limit. **(b)** Angular error as a function of the spatial frequency of the albedo map, with the maximum spatial frequency of the normal map set to 0.3 the Nyquist limit. Line colors are as indicated in (a). **(c)** Disparity root-mean-squared error (RMSE) as a function of the depth map’s highest spatial frequency. Reflectance spatial frequency is set to 0.3 of Nyquist. **(d)** Disparity root-mean-squared error of disparity estimation as a function of the reflectance map’s highest spatial frequency. Depth spatial frequency is set to 0.3 of Nyquist.

BayerBG2RGB_EA as the flag. We sample the RGB color space with Gaussians centred at $[0.5, 0.5, 0.5]$ and variance 0.1. color images have uniform color. Each frequency uses 20 samples to compute the RMSE. We run two set of experiments. One experiment keeps the reflectance frequency fixed to 0.3 of the Nyquist limit and varies the geometric frequency (normal for photometric stereo and depth for structured light). The other experiment keeps the geometric frequency fixed to 0.3 of the Nyquist limit while varying the reflectance frequency.

Photometric stereo. Figures 3.6.2(a) and (b) analyze the effective resolution and albedo invariance of normal maps computed by several combinations of methods from Sections 3.3 and 3.6, plus two more—*Baseline*, which applies basic photometric stereo to three full-resolution images; and *Color*, the one-shot color photometric stereo technique in [89]. To generate synthetic data, we (1) generated scenes with random spatially-varying normal maps and RGB albedo maps, (2) applied a spatial low-pass filter to albedo maps and the spherical coordinates of normal maps, (3) rendered them to create three sets of images—a grayscale C2B frame; three full-resolution grayscale images; and a Bayer color mosaic—and (4) added zero-mean Gaussian noise to each pixel, corresponding to a peak SNR of 30dB. Since all calculations except demosaicing are done per pixel, any frequency-dependent variations in performance must be due to this upsampling step. Our simulation results do match the intuition that performance should degrade for very high normal map frequencies regardless of the type of neighborhood processing. For spatial frequencies up to 0.3 the Nyquist limit, however, one-shot C2B imaging confers a substantial performance advantage. We multiplex the 4 light sources into 3 pixels using the 3 by 4 optimal C^P (Table 3.2 from Section 3.4) for the C2B frame and Bayer color mosaic. We uniformly sample 4 light source directions until their condition number is below 3.

Structured light triangulation. We multiplex 4 phase-shifted sinusoidal patterns with spatial frequency of 8 using the 3 by 4 optimal C^P (Table 3.2 from Section 3.4). We set the centre of the projector to $(20, 0, 0)$ and rotate it 10° in the XZ plane towards the camera. The camera image plane is a rectangle of length 50 and width 50 centred at the origin. For the depth map, we generate a 300×300 image of uniformly-sampled depths. We set the maximum depth frequency by band-pass filtering this depth map. We then rescale and translate the image to a depth map with a mean depth

100 (unitless in MATLAB) and a depth variance of 8.333. Phase unwrapping was done by finding the maximum and minimum disparity with depth bounds of 92.5 and 107.

We show the results of our simulations for structured light triangulation in Figure 3.6.2 with 95% confidence intervals on BRD-R. The experiments obtained by fixing the reflectance frequency show that when the reflectance frequency is low, ID-DM outperforms other techniques. The experiments fixing the depth frequency show that BRD-R outperforms ID-DM when there is high frequency reflectance variation. In general, these experiments show that the choice of shape estimation method has trade-offs. Reflectance-invariant methods like BRD-R are robust to arbitrary reflectance frequencies but may not always help improve performance.

3.7 C2B Pixel Design Considerations

Two- and four-bucket pixel layouts from the literature. The design of our sensor’s pixels builds upon several earlier pixel designs that include more than one bucket. Figure 3.7.1 shows four such designs, including an earlier C2B design from our group [211]. These layouts are shown here to provide some basic intuition about the two key design choices we made in our prototype’s new sensor—and in C2B sensing more generally.

Design choice #1: per-pixel vs. per-bucket coding. As illustrated in Figures 3.1.1 and 3.1.3, the two buckets inside a C2B pixel are never in the same state at the same time (*i.e.*, both active or both inactive). This can be viewed as a form of coding that assigns a single bit to each pixel, so that bucket activities can be toggled. A natural question is whether it would be possible and/or desirable to control these buckets individually, so that they could be activated or de-activated independently of each other.

Unfortunately, although more flexible, per-bucket coding would cause poor CMOS performance for two fundamental reasons. First, it would require doubling the pixel’s SRAM block in order to store the required two bits per pixel (*i.e.*, one per bucket). Such a block would be very large relative to the pixel’s overall dimensions because even a one-bit memory uses lots of gates (Fig 3.7.1, lower right). A two-bit in-pixel memory would therefore reduce the photo-sensitive region’s relative size severely—and thus the pixel’s overall quantum efficiency. Second, even ignoring the size of the photo-sensitive region, allowing both buckets to be active simultaneously would cause sluggish/inefficient pixels. This is because pixels operate by inducing a gradient field in the photo-sensitive region that “pulls” the light-generated charges toward the active bucket (see dashed lines in top row of Figure 3.7.1). Activating both buckets would induce competing gradient fields, causing charges to move slowly toward the individual buckets.

Design choice #2: C2B pixels versus memory-less multi-bucket pixels. We believe that C2B sensors have several advantages: (1) they are scalable because they can pack S views into one C2B shot via coding, without hard-wiring S into the pixel’s design. This is a particularly important feature because both structured-light and photometric-stereo systems typically use more than four illumination conditions for better accuracy and robustness. C2B sensors can achieve this—possibly at the expense of some spatial resolution compared to the results shown in this chapter—whereas a four-bucket sensor could not. (2) C2B pixels are much more light-efficient because each extra

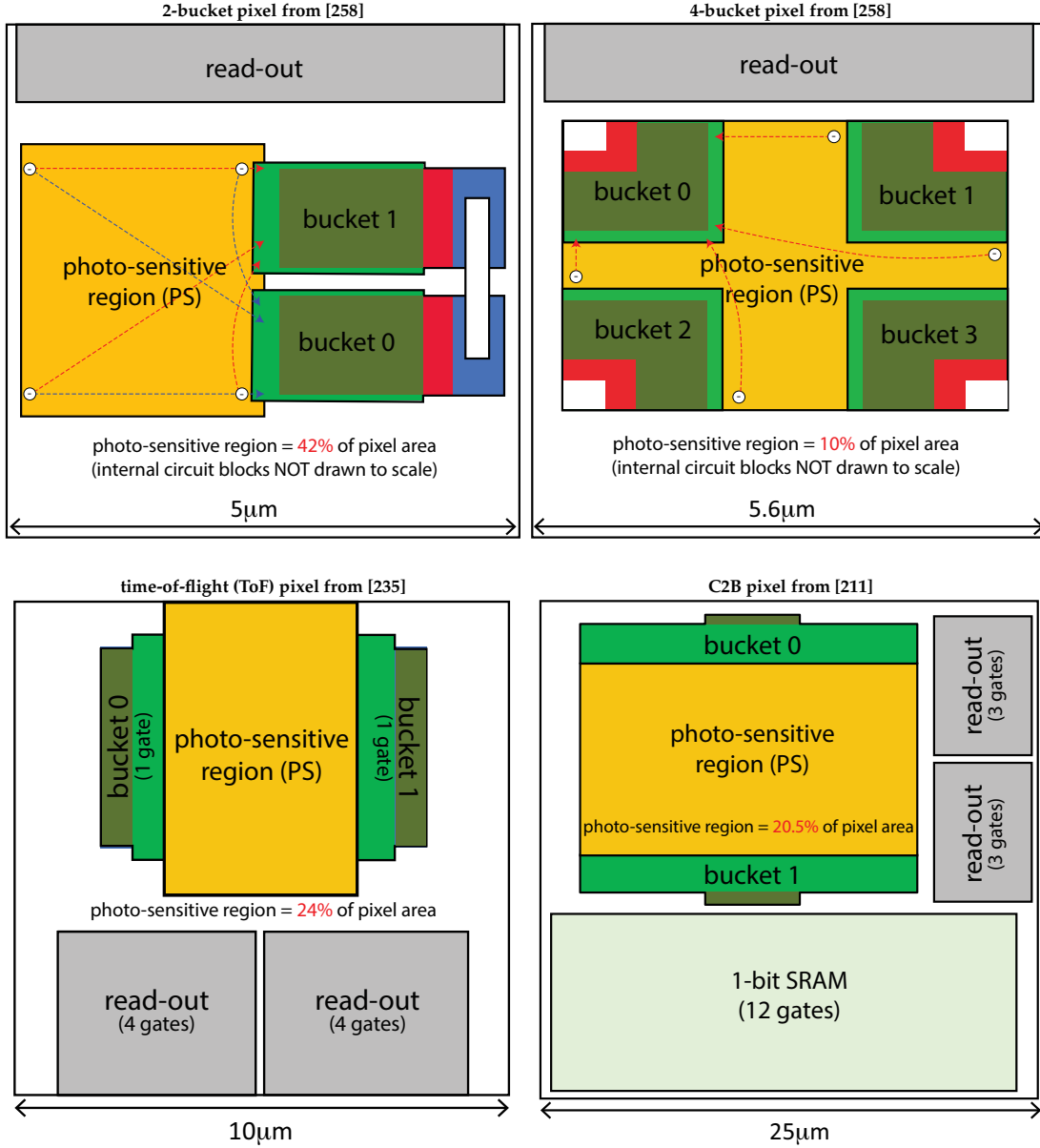


Figure 3.7.1: CMOS layout of two- and four-bucket pixels from the literature. Colors indicate circuit blocks of similar functionality across these layouts. Orange blocks denote the pixel’s photo-sensitive region. White regions are occupied by wires, etc.. Except for the relative size of the photo-sensitive region, the dimensions of internal blocks were not included in [258]. The number of gates in each block is listed where known. Dashed curves in the top row indicate the approximate path traced by photo-generated charges when bucket 1 (red) or bucket 0 (blue), is active, respectively.

bucket takes up lots of space within a pixel. For example, the photo-sensitive region of the four-bucket pixel in the top for Figure 3.7.1 is four times smaller than its two-bucket counterpart (10% vs. 42% of pixel area, respectively). Adding even more buckets to a 2D sensor could therefore lead to significant inefficiencies. (3) Moreover, it is unclear if 2D sensors with more than four buckets could offer acceptable imaging performance. This is because the charge-collection gradient fields of the S buckets must be identical up to rotation/reflection in order to ensure similar performance (e.g., see charge-collection paths in Figure 3.7.1, top row), but the rectilinear CMOS design rules

generally prevent achieving anything beyond four-fold gradient field symmetries. (4) S-bucket sensors cannot be used for applications that require sensor-plane coding because they lack the in-pixel memory required for this capability.

3.8 Experiments

3.8.1 Experimental conditions

Both C2B frame acquisition and scene reconstruction run at 20Hz for all experiments, using $F=3$, $S=4$, the corresponding optimal C^p from Table 3.2, and the 2×2 mosaic tile defined in Section 3.2.1. C2B frames are always processed by the same sequence of steps—demosaicing, demultiplexing and per-pixel reconstruction. For structured light, we fit an 8mm Schneider Cinegon $f/1.4$ lens to our camera with its aperture set to $f/2$ and use a TI LightCrafter for projecting 684×608 -pixel, 24-gray-level patterns at a rate of $S \times 20\text{Hz}$ in sync with sub-frames. The stereo baseline was approximately 20cm, the scene was 1.1~1.5m away, and the cosine frequency was 5 for all patterns and experiments. For photometric stereo we switch to a 23mm Schneider APO-Xenoplan $f/1.4$ lens to approximate orthographic imaging conditions and illuminate a scene 2~3m away with four sub-frame synchronized Luxdrive 7040 Endor Star LEDs, fitted with 26.5mm Carclo Technical Plastics lenses.

Configuration for photometric stereo. Light source calibration consisted of two steps. First, we computed the direction of incidence of each source onto the scene using a chrome sphere. Second, to account for the non-uniform illumination spot produced by the LED lenses, we placed a fronto-parallel, uniformly-white and diffuse poster board in front of the camera and computed the light source intensity at each pixel. This intensity was subsequently used for all shape computations at that pixel.

Configuration for structured-light triangulation. Our projector can store up to 96 binary patterns in its memory, and projects gray-scale patterns by dithering those binary patterns at rates up to 96 patterns per 20ms video frame. This limited projection rate, along with limitations of our camera’s triggering firmware, made it impossible to project more than 24 binary patterns onto the scene within each sub-frame. As a result, our structured-light projection patterns were quantized significantly, taking on integer values in the range $[0, 23]$. Phase unwrapping was done by using depth bounds.

3.8.2 Quantitative experiments: one-shot versus multi-frame acquisition

Our goal was to compare the 3D accuracy of one-shot C2B imaging against that of full-resolution sequential imaging—using the exact same system and algorithms. Figure 3.8.1 shows the static scenes used for these experiments, along with example reconstructions for photometric stereo and structured light, respectively. The “ground truth,” which served as our reference, was computed by averaging 1000 sequentially-captured, bucket-1 images per illumination condition and applying the same reconstruction algorithm to the lower-noise, averaged images. To further distinguish the impact of demosaicing from that of sensor-specific non-idealities, we also compute shape from a *simulated* C2B frame; to create it we spatially multiplex the S averaged images computationally

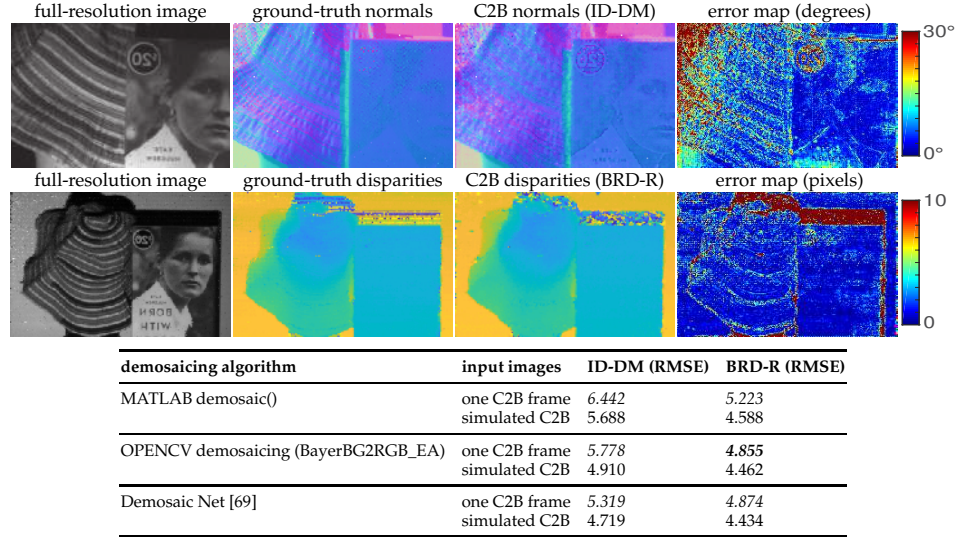


Figure 3.8.1: Quantitative experiments for photometric stereo (Row 1) and structured light (Rows 2, 3). Per-pixel unit normals \mathbf{n} are visualized by assigning them the RGB color vector $0.5\mathbf{n} + 0.5$.

in a way that simulates the operation of our C2B sensor. Row 3 of Figure 3.8.1 shows some of these comparisons for structured light. The BRD-R method, coupled with OpenCV’s demosaicing algorithm, yields the best performance in this case, corresponding to a disparity error of 4%.

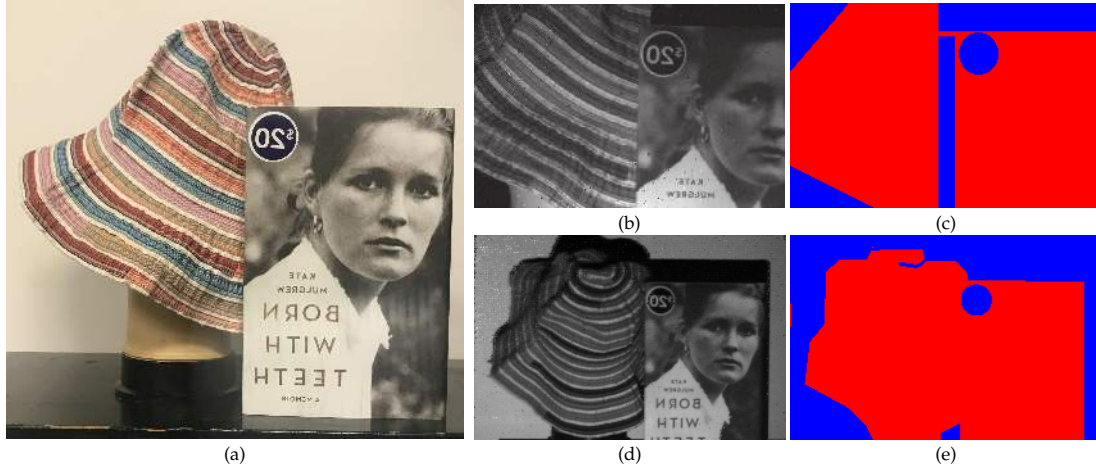


Figure 3.8.2: Objects used for ground-truth evaluation. (a) Color view of the objects. (b) C2B view of the scene for photometric stereo. (c) The binary foreground mask (red is foreground). The mask excludes pixels in shadow or pixels that exhibit glossy reflection. (d) C2B view of the scene for structured-light triangulation. (e) Binary foreground mask used for structured-light error calculations (red is foreground), non-object pixels are masked.

Definition of “ground truth.” The goal of our quantitative experiments was to evaluate how one-shot 3D estimates from a C2B camera differ from those computed by applying the identical shape estimation algorithm to full-resolution images acquired sequentially. Any differences between these estimates would necessarily be due to the demosaicing process or the coded two-bucket imaging procedure itself—not the accuracy of the underlying algorithms employed. We therefore treat the 3D estimates computed by multi-frame acquisition to be our reference, “ground-truth” shape. Because several pixels in the field of view do not produce valid “ground-truth” 3D

demaicing algorithm	input images	ID-DM (degrees)	BRD-R (degrees)
MATLAB demosaic()	one-shot	12.079	11.883
	simulated one-shot	10.539	10.140
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	12.224	12.322
	simulated one-shot	10.057	9.703
Demaic Net [69]	one-shot	11.675	11.936
	simulated one-shot	10.215	9.758

Table 3.4: *Angular RMSE for photometric stereo.*

demaicing algorithm	input images	ID-DM (degrees)	BRD-R (degrees)
MATLAB demosaic()	one-shot	5.917	4.901
	simulated one-shot	4.520	4.096
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	5.453	4.980
	simulated one-shot	3.947	3.745
Demaic Net [69]	one-shot	5.357	5.375
	simulated one-shot	3.993	3.990

Table 3.5: *Median angular error for photometric stereo.*

data (due to shadows, isolated regions with specular reflectance, *etc.*) we define a foreground mask to exclude them (see Figures 3.8.2(c) and 3.8.2(e)).

Evaluation procedure. For structured light we use one metric: (1) the percentage of bad-matching pixels [214]. We define bad-matching pixels to be pixels whose disparity is more than one pixel away from the ground-truth disparity. For photometric stereo, we use root-mean-squared angular error and median angular error, both measured in degrees. We compare the impact of three different RGB demosaicing procedures: MATLAB’s built-in function; OPENCV’s function with edge-aware demosaicing; and the deep demosaicing system of Gharbi *et al.* [69] which represents the current state of the art on the problem. We used the same test objects for both photometric stereo and structured-light triangulation. The objects, shown in Fig 3.8.2(a), consist of a book and a colorful hat whose high-frequency albedo variations make its reconstruction challenging to algorithms—such as color photometric stereo—as it breaks the constant-albedo assumption.

Photometric stereo results. Table 3.4 compares the RMSE obtained by the various methods, for both one-shot and simulated one-shot images. These results suggest that the ID-DM method has the best performance when coupled with the deep-learning-based demosaicing algorithm [69]. As shown in Table 3.5, which considers median error, the BRD-R method performs the best with MATLAB’s demosaic function. While there is a discrepancy in the best reconstruction method, we note that the errors are close between ID-DM and BRD-R, which is consistent with our simulations (Figure 3.6.2). This suggests that even ID-DM is able to handle the albedo variations present in the scene. One thing to note is that the demosaicing algorithms are designed for image intensities and not illumination ratios so it is possible for a demosaicing algorithm more optimized for illumination ratios to improve the performance of the BRD-R method. In terms of overall magnitude, the differences between one-shot and multi-shot are within the range of previously-reported ground-truth evaluations of previous one-shot photometric stereo algorithms [64, 40]. Comparing to the results obtained from simulated one-shot images it is clear that most of the reconstruction error is indeed due to demosaicing rather than our sensor’s performance under C2B imaging conditions.

Structured-light triangulation results. Tables 3.6 compares the percentage of bad-matching pixels for structured-light triangulation. Here the BRD-R method yields the best performance with OpenCV’s demosaic function giving the lowest bad-pixel percentage. There is a 10-20% gap be-

demosaicing algorithm	input images	ID-DM (bad%)	BRD-R (bad%)
MATLAB demosaic()	one-shot	61.30	48.23
	simulated one-shot	50.65	30.27
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	56.62	45.29
	simulated one-shot	48.34	30.84
Demosaic Net [69]	one-shot	55.6	47.16
	simulated one-shot	46.51	33.94

Table 3.6: *Disparity bad-pixel percentage for structured-light triangulation.*

tween the bad-pixel percentage of simulated one-shot and actual one-shot mosaics. This discrepancy is likely due to the fact that disparity errors in the actual one-shot results generally decrease for the simulated one-shot results. Since all disparity errors greater than 1 are treated equally in the bad-pixel percentage metric, this gap can be explained by disparity errors close to, but greater than 1 becoming lower for simulated one-shots. Since BRD-R methods has lower RMSE (Figure 3.8.1) than ID-DM methods, this also explains the gap there.

3.8.3 Live 3D acquisition experiments: photometric stereo

We show several examples of reconstructing dynamic scenes in Figure 3.8.3.

Person-PS. Errors in the hair result from the appearance of hair violating the Lambertian image formation model. Sectioned colors in the background result from shadows. Since we are using 4 subframes and 3 pixels, we are not robust to shadows and occlusions in the scene. The different expressions made by the subject show the ability of the system to capture surface normals at different orientations. Notice that the estimated normals in areas in the line of sight of all LEDs are not severely affected by neighbouring shadowed pixels. This is because our technique does not rely on the integrability of the surface in order to estimate normals. Our normal estimates are stable over time and do not exhibit any clear temporal noise. We try a variety of hand gestures to test the robustness of our system.

Hat-PS. Since we use ID-DM, the normal estimates should not be robust to high reflectance frequencies. We observe this in the form of temporal jitter in the video, where we can see various artefacts that appear on the hat as it is being rotated and moved. However, even though there are artefacts, the estimated normals overall look reasonable. We also place a hand in the scene for comparison. Since the hand is textureless, there are little, if any, artefacts in the estimated normals of the hand. The hand crushes the hat and shows that our system is able to handle non-rigid deformations of objects in the scene. It is important to point out that colorful and textured objects are not apt for traditional one-shot methods, so this is a hard object from which to estimate normals.

Dress-PS. The black edges in the normal map are caused by black edges in the original image and these are not artefacts from our algorithm. We see that the normal estimates are consistent during each period of the swing.

3.8.4 Live 3D acquisition experiments: structured-light triangulation

Person-SL. There are pixels with poor performance, and they appear as temporal noise in the disparity map. These erroneous disparities are strictly due to sensor imperfections and not the algorithms proposed. Because we are using a single projector, we are not able to estimate the disparity

of the pixels in the areas that are shadowed by the hand. We show the subject making different hand gestures. These hand gestures introduce discontinuities in the disparity map between the face and the hand as well as between different fingers. Even under these different deformations and discontinuities, the disparity map is quite stable.

Handkerchief-SL. We can see that while the object has texture, the resulting disparity map are largely invariant to the albedo. Due to motion in the scene however, they are not completely invariant, and this is apparent as the edges of the textures can be traced out in the scene. We also place the handkerchief under different deformations, showing that albedo-invariance holds in many cases.

Toy Cloud-SL. Since the cloud is colorful, this is an object that is traditionally difficult for color-based one shot techniques. We can see that the disparity in our case is still smooth, and the quality of the disparity map is not affected by the different colors in the toy cloud. We do note that right side of the cloud has a color that is very dark, so the disparity map is naturally more noisy for that region; this is not a problem with our algorithm.

3.9 Limitations

While our framework provides a principled way for optimizing the control of C2B cameras, there are several limitations to our approach.

First, our lower bound on the objective function is only tight for a subset of sub-frame values, namely those for which a Hadamard matrix can be defined. This means that, in general, we cannot use the lower bound to evaluate the optimality of a multiplexing matrix. Furthermore, since the objective function is non-convex [197], finding optimal matrices can be challenging, exacerbated by the fact that the matrix must be binary [282]. For this work, we circumvent this issue by limiting our problem to small values of S , for which searching over all possible binary matrices is tractable.

Second, while we define the objective function based on minimizing the MSE of the recovered S sub-frame images, it is apparent that for one-shot shape recovery, minimizing the MSE of the image intensities is not exactly what we want. Instead, we want to minimize the MSE of the reconstructed depth or surface normal. For example, Mirdeghan et al [149] showed that minimizing the expected number of incorrect correspondences can lead to more optimal structured light patterns. However, we found it simpler to analyze the behavior of the multiplexing matrix through its effect on the noise of the sub-frame images, rather than the recovered geometry, which led to mathematically rich results, including closed-form expressions for optimal multiplexing matrices and lower bounds. Implicitly, we are therefore assuming that less noisy measurements of the image intensities will result in less noisy depth measurements. Beyond this, it would also be helpful to consider task-specific objective functions [145, 164, 248]. Multiplexing matrices that recover perceptually pleasant video sequences should differ from those that recover depth, so it makes sense to use objective functions that incorporate these differences.

So far, we have only considered the effects of read noise, and as a result, there is a gap between our theoretical and experimental gains. Incorporating shot noise into the MSE calculation [269, 198] could lead to more optimal multiplexing matrices that may differ from those when read noise dominates.

3.10 Concluding Remarks

Our experiments relied on some of the very first images from a C2B sensor. Issues such as fixed-pattern noise; slight variations in gain across buckets and across pixels; and other minor non-idealities do still exist. Nevertheless, we believe that our preliminary results support the claim that 3D data are acquired at near-sensor resolution.

We intentionally used raw, unprocessed intensities and the simplest possible approaches for demosaicing and reconstruction. There is no doubt that denoised images and more advanced reconstruction algorithms could improve reconstruction performance considerably. Our use of generic RGB demosaicing software is also clearly sub-optimal, as their algorithms do not take into account the actual correlations that exist across C2B pixels. A prudent approach would be to train an assorted-pixel algorithm on precisely such data.

Last but certainly not least, we are particularly excited about C2B cameras sparking new vision techniques that take full advantage of their advanced imaging capabilities.

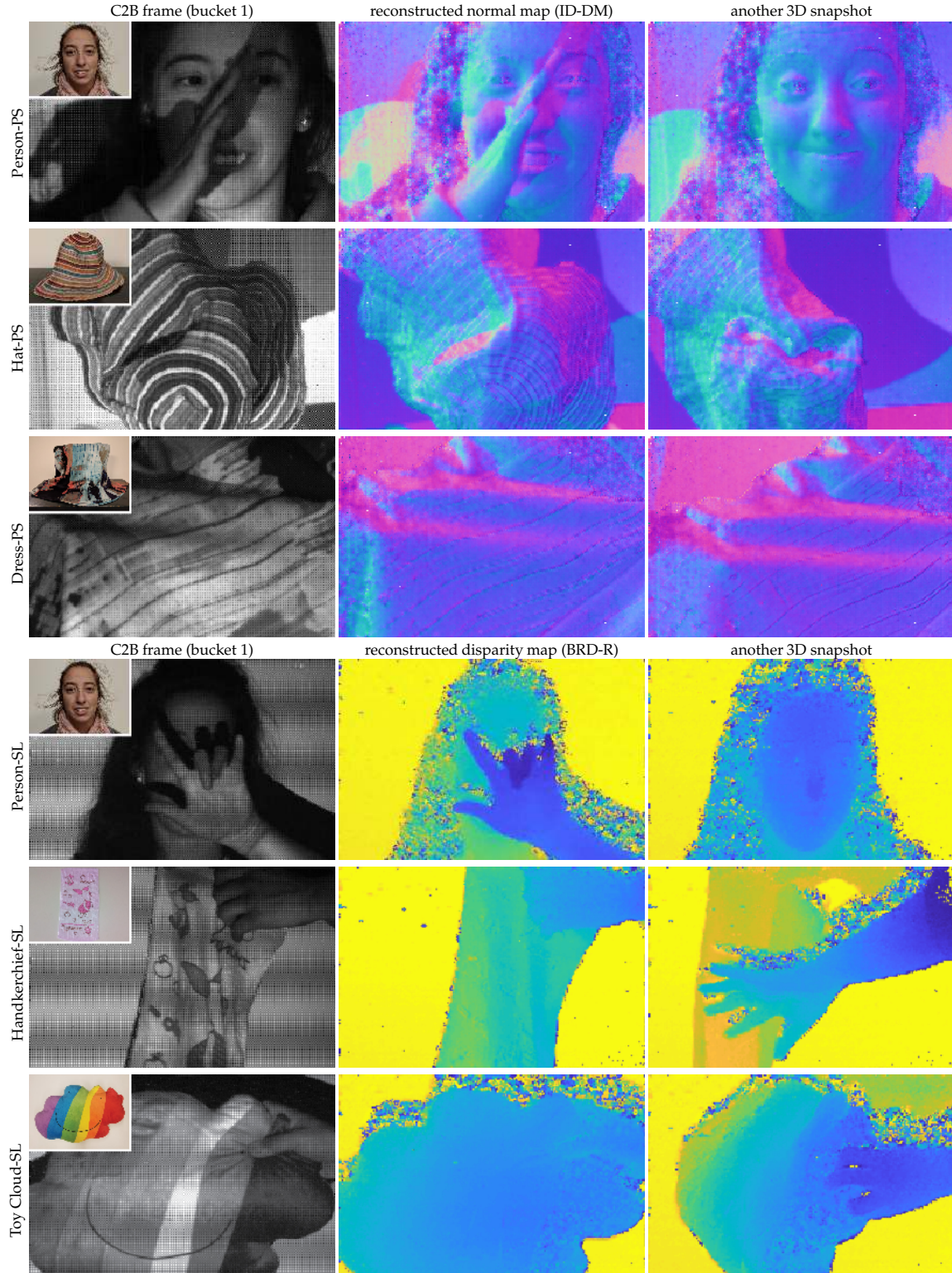


Figure 3.8.3: Live 3D acquisition experiments for photometric stereo (top) and structured light (bottom). Scenes were chosen to exhibit significant albedo, color, normal and/or depth variations, as well as discontinuities. For reference, color photos of these scenes are shown as insets in Column 1. Qualitatively, reconstructions appear to be consistent with the scenes' actual 3D geometry except in regions of low albedo (e.g., hair) or cast shadows.

4 Passive Ultra-Wideband Single-Photon Imaging

4.1 Introduction

A basic rule of thumb in high-speed imaging is that *speed needs light*: the faster a scene changes, the more light we need to image it accurately without excessive noise or motion blur. Over the decades, high-speed light sources [206], fast cameras [31, 280, 66], and depth sensors [165, 217] have made it possible to image dynamic phenomena occurring in ever-smaller time intervals with the help of actively-controlled light sources and *synchronization*: to collect enough light, the same picosecond- or nanosecond-scale event may be imaged millions of times by operating a camera and a source in lockstep, at MHz repetition rates or more.

Acquiring videos of ultrafast phenomena this way—from imaging light in flight [67] to fast biological processes [26]—is now quite common. Unfortunately, while these techniques do capture ultrafast events, they cannot simultaneously capture *slower* ones too: time wraps at the sync period, blurring out anything occurring over longer timespans.

But how do we image highly dynamic scenes—both slow and ultrafast—*passively*, without any light sources under our control, no synchronization, and not much light? Very little is known about this problem because existing models for passive low-light imaging [61, 151, 44, 171, 218] break down at timescales much shorter than the timespan between photon arrivals. As a result, ultrafast imaging in low light has remained beyond the reach of passive methods.

In this work we seek to bridge these two regimes, active and passive, by revisiting the need for synchronization when imaging ultrafast scenes in low light. Working from first principles, we develop a novel theory of passive single-photon imaging that is specifically designed to eliminate synchronization between a camera and the light sources in a scene. The only requirements are that (1) the camera’s pixels can detect and time-stamp individual photons and (2) their dead-time period does not impair detection of photons from one source significantly more than any other. In this imaging regime, each camera pixel time-stamps the photons it detects using an internal clock that follows the arrow of time, obviating the need for any external timing signals.

Our work is based on the observation that passive (sync-free) imaging is fundamentally *more* powerful than active imaging in such settings. Specifically, without the periodic timing signal from a light source, time never wraps at a sync period; ultrafast scenes can be imaged for arbitrarily long timespans; and flux variations that occur concurrently across 12 orders of magnitude in time (picoseconds to seconds)—and that involve many unknown sources—can be recorded with just one camera.

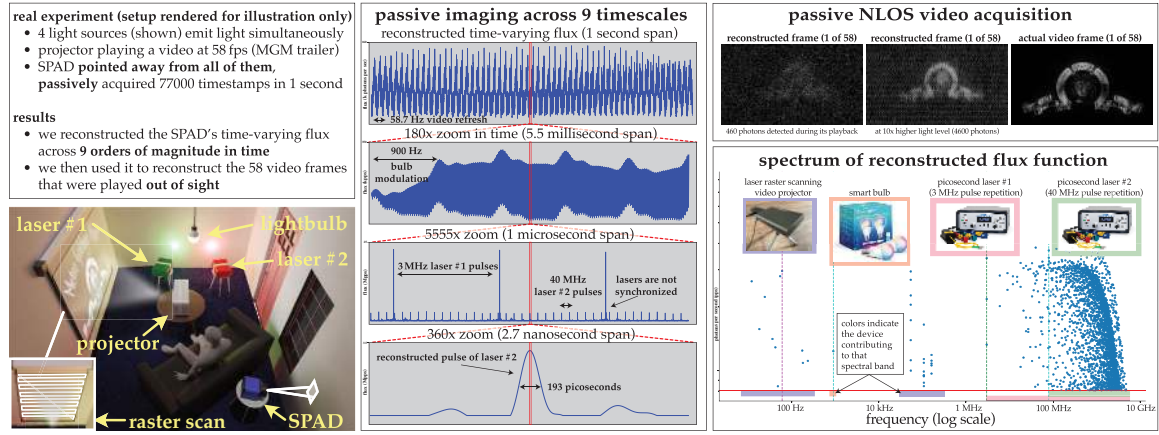


Figure 4.1.1: Passive ultra-wideband imaging with a single free-running SPAD pixel. **Left bottom & Center:** In a real, captured experiment, an unsynchronized single-photon avalanche diode (SPAD) passively records indirect light coming from multiple sources operating asynchronously from each other (unsynchronized picosecond lasers, projectors, etc.). See Figure 4.5.1 (row 2, middle) for actual scene. The incident flux exhibits simultaneous intensity variations with a bandwidth that spans roughly 9 orders of magnitude in frequency. **Bottom right:** Multiple concurrently-occurring phenomena in the flux function can be identified after acquisition: video flicker (58 Hz), the pulse-width modulation of an LED light bulb (900 Hz), a movie projected onto a nearby wall by a raster-scanning laser projector (up to 5 MHz), and two unsynchronized picosecond lasers (40 MHz–10 GHz). **Top right:** By reconstructing the time-varying flux function of the laser projector, video frames are reconstructed at 1280x720 resolution using roughly 450–4500 photons collected during each 1/58 s frame.

Because photon timestamps due to all light sources and all timescales are recorded concurrently, the choice of which timescale to show and which light source(s) to use for visual processing can be done *after* acquisition (Figure 4.1.1). Thus, just like light field cameras [163] enable post-capture refocusing in space, this new imaging regime enables post-capture refocusing *in time*—from transient to everyday timescales. We demonstrate this one-of-a-kind capability experimentally in Figure 4.5.1, where we use photon timestamp data captured by a free-running SPAD camera to play back video of a rapidly-spinning fan at both 1,000 and 250 billion frames per second. We call this novel regime *passive ultra-wideband imaging*.

The key challenge in this regime is how to reconstruct flux functions with an ultra-wide spectrum (DC to over 10 GHz) from a stream of photon timestamps that increase monotonically. To tackle it, we use our *flux probing theory* discussed in Chapter 5 that draws on results from stochastic calculus [50, 46] to relate the Fourier series decomposition of a time-varying flux function to the timestamp realizations of an underlying stochastic process [41, 123]. The mathematical underpinnings of our approach are grounded in statistics [23, 251, 65] and time series analysis [24, 279, 102, 133, 189, 247], and similar methods have explored flux function estimation in optical communications [140, 148, 108].

Our work ties together several lines of prior research on “extreme” imaging, both passive and active. In passive settings, several techniques have recently been proposed for estimating flux from photon data [139, 218, 95, 94]. These rely on a variety of flux constancy assumptions and, as a result, are not applicable to the ultra-wideband regime we consider. In active settings, single-photon imaging techniques have relied exclusively on sync-relative timestamps [126, 226, 238, 80, 221], where information about sub-MHz flux variations has already been lost. Aside from single-photon imaging, active ultrafast imaging techniques have also used heterodyning to measure flux

at one specific modulation frequency [86, 118, 119, 14, 79]. These techniques have neither the light efficiency nor the ultra-wide bandwidth we demonstrate in this chapter.

4.2 The Passive Ultra-Wideband Imaging Regime

Passive, sync-less imaging. We assume that the imaging system exerts no control over a scene’s appearance: the scene’s light source(s) can be natural, artificial, or both, and their number, operating principle, and time-varying properties are unknown and unconstrained (Figure 4.1.1). Importantly, we assume that no electronic timing signals, such as triggers or sync pulses, are received from any of them.

The time-varying flux function. Following standard radiometric conventions [29], we express incident light at a pixel as an unknown time-varying function $\varphi(t)$ that represents the pixel’s instantaneous flux at time $t \geq 0$. Our goal is to acquire a continuous representation of the flux function over a possibly unbounded acquisition interval $[0, t_{\text{exp}}]$ (milliseconds, seconds or much longer). In the following we assume that $\varphi(t)$ is expressed in units of photons per second, is continuous, and has finite spectral support bounded by frequencies f_{min} and f_{max} .

Ultra-wideband flux. We seek to reconstruct flux functions that have ultra-wide bandwidth, *i.e.*, whose frequency content spans the entire range from constant flux ($f_{\text{min}} = 0$ Hz) to extreme time-of-flight timescales ($f_{\text{max}} \geq 10$ GHz) [79, 100]. Moreover, we assume that no prior information is available about the spectrum of $\varphi(t)$.

Photon arrival model. Our work applies to the single-photon imaging regime, where the timespan between consecutive photon arrivals is not negligible. In this setting, $\varphi(t)$ is the rate function of an *inhomogeneous Poisson process* governing photon arrivals [204, 148]. The mean value of $\varphi(t)$ represents the average flux received over the observation interval $[0, t]$ in units of photons per second and its inverse, denoted by T_{avg} , is the average timespan between consecutive photon arrivals [94].

Low-flux photon detection model. Modern SPADs can detect and time-stamp the arrival of individual photons with extremely high temporal precision (typically tens of picoseconds [280, 96]). SPADs are not perfect detectors, however, as they exhibit four main non-idealities: quantum efficiency, dead time, timestamp quantization and jitter. *Quantum efficiency* refers to the pixel’s probability of actually detecting a photon when it is in its active state. This probability can be well below 1 depending on wavelength; since it can be thought of as scaling the flux function, we assume it is absorbed in $\varphi(t)$. After a photon detection, SPADs are blind to subsequent photon arrivals for an interval known as the *dead time*. Dead time can skew photon detection statistics quite significantly when photons arrive closely enough in time to fall within a SPAD’s dead-time window with high probability [180, 194, 95]. For simplicity, we focus on low-flux imaging in this chapter, where consecutive arrivals are spaced much farther apart than the SPAD’s dead time.¹ In this case, detections are governed by the same stochastic process that describes photon arrivals [107], with rate function $\varphi(t)$ and average timespan T_{avg} between detections.

¹For example, T_{avg} was nearly six times our SPAD’s dead time in the experiment of Figure 4.1.1 (1.3 microseconds versus 231 nanoseconds).

imaging regime	passive inter-photon imaging (interdetection-limited)	active histogram-based imaging (sync- and quantization-limited)	passive ultra-wideband imaging (quantization-limited only)	
	$f_{\max} < 1/T_{\text{avg}}$	$f_{\min} \geq f_{\text{sync}}, f_{\max} \leq 1/(2Q)$	$f_{\max} \leq 1/(2Q)$	
light source(s)	one or more sources, no sync	one source only, periodic with period $1/f_{\text{sync}}$, sync required	one or more sources, no sync	
typical freq. range	low Hz to tens of kHz (plus DC)	low MHz to well above 10 GHz (plus DC)	DC to well above 10 GHz	
valid frequencies	all frequencies in range	all frequencies in $\phi(t)$ must be integer multiples of f_{sync}	all frequencies in range	
corrupting flux	any flux with frequencies $> 1/T_{\text{avg}}$	any flux with frequencies that are not integer multiples of f_{sync}	frequencies > 31 GHz	
input data	stream of absolute timestamps	stream of sync-relative timestamps	stream of absolute timestamps	
# distinct time bins	not applicable	thousands (typical)	billions to trillions (increases with t)	
# photons per bin	not applicable	a non-negative integer; Poisson-distributed, mean proportional to $\lfloor t f_{\text{sync}} \rfloor$	0 or 1, vast majority of bins have 0	
illustration of flux, detections & sync in each regime	<p>$\phi_1(t) = 5 + 5 \sin(2\pi t)$</p> <p>no sync (0 Hz)</p> <p>detections due to ϕ_1</p> <p>elapsed time t (sec)</p>	<p>$\phi_2(t) = 5 + 5 \sin(7 \times 6 \times 2\pi t)$</p> <p>sync (7 Hz)</p> <p>detections due to ϕ_2</p> <p>elapsed time t (sec)</p>	<p>$\phi_3(t) = 5 + 5 \sin(5 \times 8 \times 2\pi t)$</p> <p>sync (5 Hz)</p> <p>detections due to ϕ_3</p> <p>elapsed time t (sec)</p>	<p>$\phi_4(t) = \phi_1(t) + \phi_2(t) + \phi_3(t)$</p> <p>no sync (0 Hz)</p> <p>detections due to ϕ_4</p> <p>elapsed time t (sec)</p>
measurement model of each regime	<p>inter-photon flux</p> <p>estimated flux</p> <p>elapsed time t (sec)</p>	<p>photon-count histograms</p> <p>7 Hz 5 Hz</p> <p>corruptions $\phi_1 + \phi_3$</p> <p>ϕ_2 photons</p>	<p>photon-count histograms</p> <p>5 Hz 7 Hz</p> <p>corruptions $\phi_1 + \phi_2$</p> <p>ϕ_3 photons</p>	<p>frequency spectrum</p> <p>amplitude</p> <p>0 Hz 1 Hz</p> <p>detection threshold</p> <p>40 Hz 42 Hz</p> <p>frequency (Hz)</p>

Table 4.1: Low-flux imaging with photon timestamps. **Left column:** Passive imaging has so far assumed that photons can be detected at a rate (much) greater than the highest flux frequency. **Middle column:** Active techniques are designed to handle the opposite case, where photon detections occur at a rate (much) lower than the flux function’s frequencies. If a light source’s frequency is not a multiple of f_{sync} (e.g., 5 Hz for φ_2 photons and 7 Hz for φ_3 photons in third row above), its photons will land in the wrong time bin. This contributes to noise instead of signal, i.e., the histogram will be “flattened” (compare the two red and two green histograms, respectively, in third row). Moreover, even when f_{sync} is well-matched to one light source, other sources emitting at non-multiples of f_{sync} will corrupt the histogram (third row, gray histograms). **Right column:** Our approach inherits the most important **features** of both regimes, without their **limitations**.

Non-negligible dead time. When dead-time intervals become comparable to T_{avg} (or longer), photon detections are not Poisson because the detection of a photon may impact the detection of subsequent ones. We show in Section 5.5 that our stochastic calculus framework covers this case as well, enabling acquisition of $\varphi(t)$ by slightly amending the equations and algorithm of the negligible dead-time case.

Timestamp model. Photon timestamps are subject to *quantization* from the time-to-digital conversion process and *jitter*, i.e., instabilities in timing electronics. Both can be as low as a few picoseconds for SPADs in the visible range [202]. To simplify our analysis, we assume without loss of generality that timestamp resolution and timestamp accuracy are identical, so that the timestamps’ bin size Q accounts for jitter as well.²

The stream of absolute detection timestamps. Since no external timing signal is available to serve as a reference, we assume that photon detection timestamps follow the arrow of time, increasing monotonically according to the SPAD’s internal clock. This results in a stream of timestamps $\mathcal{T} = (\tau_1, \dots, \tau_{N(t)})$, where τ_i is the elapsed time from the beginning of acquisition until the i -th photon detection, and $N(t)$ counts the total photons detected up to time t . We refer to timestamps τ_i as *absolute timestamps*. Absolute timestamps can be acquired by operating SPAD pixels in their “passive free-running” mode [95].

²When timestamp accuracy is worse than timestamp resolution, the timestamps’ effective number of bits is reduced [39]. Our system’s timestamps, for example, are quantized to 4 picoseconds but the standard deviation of jitter is 16 picoseconds, so we conservatively use $Q = 16$ for performance modeling purposes. Explicit treatment of jitter is beyond the scope of this thesis (e.g., jitter can actually *improve* timing resolution [190, 191, 186]).

4.2.1 Imaging with photon timestamps

The particular imaging regime outlined above generalizes two broad classes of single-photon imaging research, both of which use photon timestamps as their main input. We distinguish between the two by considering the relation between (a) the rate of photon detections and (b) the maximum reconstructible frequency in each case (Table 4.1).

Passive inter-photon imaging. Recent work in passive single-photon imaging has proposed treating the timespan between consecutive detections as a noisy sample of the scene’s flux [95, 94]. This implicitly assumes that flux does not vary in that timespan, which makes the rate of photon detections a (loose) upper bound on f_{\max} . As a result, passive low-flux imaging with SPADs has so far been restricted to slow speeds, with f_{\max} on the order of tens of kHz [218].

Active histogram-based imaging. Approaches that employ synchronized light sources [107, 131] occupy the other extreme of the frequency range. Their basic principle is to time-stamp detections relative to a sync signal of a known frequency f_{sync} , so that all timestamps wrap to the same brief interval $[0, 1/f_{\text{sync}}]$ regardless of the actual timespan between them. This forces photons to accumulate in a relatively small number of time bins—typically a few thousand—and yields a photon-count histogram [43] that is a noisy sampling of the scaled and time-wrapped flux function, $\lfloor tf_{\text{sync}} \rfloor \varphi(t - \lfloor tf_{\text{sync}} \rfloor / f_{\text{sync}})$. The maximum reconstructible frequency in this case is governed by the Nyquist theorem, not the photon interdetection time: a bin size of 16 picoseconds, for example, theoretically enables flux acquisition with f_{\max} equal to 31.25 GHz.³

Although this general approach achieves extremely high imaging speeds [171], its reliance on relative timestamps comes with a major constraint: the incident flux must also be a periodic function with a period equal to $1/f_{\text{sync}}$, to ensure that $\varphi(t)$ and its time-wrapped counterpart are identical. This can be trivially satisfied when the only light in the scene comes from a precisely-synchronized source (a pulsed laser, light-emitting diode, *etc.*), but flux variations due to other causes cannot be reconstructed. This includes variations caused by scene motion; light sources that emit at frequencies lower than f_{sync} ; and sources that emit at higher frequencies that are not integer multiples of f_{sync} . Photons from such sources result in histogram artifacts in the form of additional photon noise [77], beat signals [100], or both.

Role of the sync frequency f_{sync} . Sync frequencies are typically in the low-MHz range in single-photon imaging applications that involve pulsed sources [226, 85, 107, 285]. This choice balances improved signal-to-noise ratio (a faster sync means more laser pulses, more photons detected in each histogram bin, and fewer time bins for them to accumulate in) against the likelihood of photons being missed due to dead time [180, 77, 194], or photons arriving “too late” because time has wrapped already [179, 171]. At such MHz sync frequencies, the memory and compute cost of reconstructing histograms from timestamps can be significant, prompting several recent schemes for just-in-time processing of (sync-relative) photon timestamps [221, 80, 130].

The passive ultra-wideband imaging regime. Intuitively, the regime we tackle in this chapter can be understood as the limit case of photon histogramming, where the sync frequency is reduced *all the way to zero*. Specifically, as f_{sync} decreases, the interval $[0, 1/f_{\text{sync}}]$ increases; more histogram

³As a reference, 31.25 GHz is the -3 dB cut-off frequency of a Gaussian pulse with a full-width-at-half-max of 36 picoseconds.

bins are needed to span it; fewer photons land into each bin; the time-wrapped flux function is able to represent variations that take place over longer timespans; and the space of reconstructible frequencies (*i.e.*, the integer multiples of f_{sync}) expands. In the limit when f_{sync} is exactly zero, there is no sync at all, timestamps become absolute and every photon lands into its own unique time bin. Crucially, *all frequencies from DC up to the Nyquist limit—and from any light source—become potentially reconstructible.*

Mathematical modeling of this limit case, however, is non-trivial because the concept of a histogram breaks down: photons never accumulate, and the contents—0 or 1—of any given bin provide almost no information about the flux function.⁴ On the computational side, the entire acquisition interval $[0, t]$ is effectively partitioned into time bins at the SPAD’s timing resolution, so acquisitions of a second or more can potentially involve trillions of time bins (most of which are empty). Fortunately, both these challenges can be overcome by formulating flux reconstruction in terms of the *photon counting process*, introduced in Section 5.4, which is not degenerate even when $f_{\text{sync}} = 0$. In particular, once we have constructed the flux function according to Algorithm 4.1, we can render the flux at any timescale we want, by simply sampling the reconstructed flux function at timesteps matched to the respective timescale.

4.3 Probing Flux Functions

Our approach establishes a direct mathematical link between (1) the stream of absolute timestamps detected at a pixel—however few or far apart they may be—and (2) the flux function that produced them. This link allows us to “probe” the Fourier spectrum of an unknown flux function across the entire DC-to-GHz range for frequencies that have statistically-significant support in the timestamp data. We introduce our flux probing theory below and address flux reconstruction in Section 4.4. For the sake of generality, we consider timestamps to be continuous-valued random variables and model quantization as part of our theory.

4.3.1 The photon counting process

Even though a single absolute timestamp provides (almost) no information about the underlying flux function, the stream of absolute timestamps as a whole contains considerable information about it. The specific relation between the two comes from stochastic calculus [50]. Specifically, in the continuous-time domain, a stream \mathcal{T} of real-valued absolute timestamps provides a noisy “reconstruction” of the integral of $\varphi(t)$ (Figure 4.3.1):

$$\underbrace{N(t)}_{\text{counting process}} = \underbrace{\int_0^t \varphi(u) du}_{\text{flux integral up to time } t} + \underbrace{M(t)}_{\text{martingale noise}}. \quad (4.3.1)$$

⁴More formally, the Poisson-distributed random variable associated with any given time bin has a mean that goes to zero as $t \rightarrow \infty$ [232].

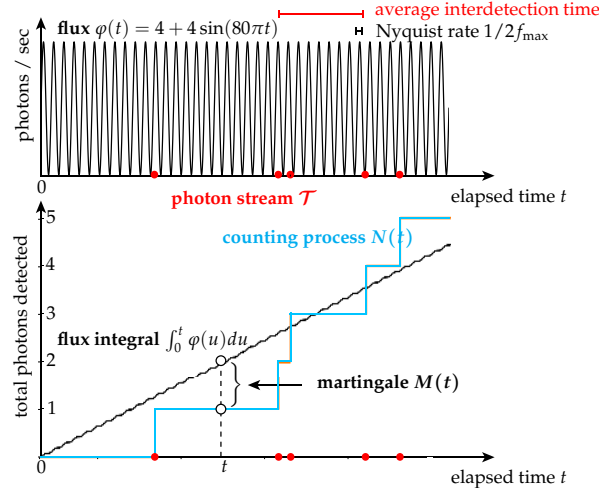


Figure 4.3.1: Relation between the stream of absolute timestamps, the counting process, the flux function and its integral. Timestamps are from a computational simulation of the inhomogeneous Poisson process with the flux function $\varphi(t)$ indicated above.

The function $N(t)$ in Eq. (4.3.1) counts the photons received up to time t and is completely determined by \mathcal{T} ; formally, it is a *counting process* [23, 50]. Viewed from the perspective of histogram-based single-photon imaging (Table 4.1, middle), $N(t)$ is the continuous-time analog of the *cumulative photon-count histogram* over the interval $[0, t]$, for $f_{\text{sync}} = 0$. The function $M(t)$ in Eq. (4.3.1) is a continuous-time random process called a *martingale* [254] that can be thought of as a form of additive zero-mean noise.⁵

As can be seen from the example of Figure 4.3.1, a single random realization of the counting process (cyan curve) is a highly discontinuous function that, on first inspection, bears no resemblance to the flux integral it is supposed to approximate in Eq. (4.3.1). These discontinuities introduce dense, spurious frequencies in the Fourier-domain representation of $N(t)$ that do not exist in the actual flux integral.

4.3.2 Flux probing theory

Our theoretical results use tools from stochastic calculus to address two basic questions. First, what is the highest possible frequency f_{\max} that can be recovered by a passive single-photon imaging system that outputs quantized absolute timestamps? Second, for frequencies within the attainable bandwidth, how can we derive a noise model that allows spurious frequencies to be efficiently detected and discarded, and the accuracy of real frequencies to be quantified as a function of the acquired timestamp stream? We summarize the key results below.

The basic premise of flux probing is that incident flux at a pixel cannot be sampled directly because MHz or GHz frequencies correspond to intensity fluctuations that are orders of magnitude faster than the rate at which individual photons arrive. In this photon-limited regime, the flux $\varphi(t)$ is the rate function of an inhomogeneous Poisson process that governs photon arrivals [148, 204] and must be inferred by passively detecting photons that arrive at the pixel.

⁵One example of a martingale is an unbiased random walk. Like $N(t)$ in Eq. (4.3.1), many other increasing stochastic processes can be expressed as the sum of a deterministic increasing function and a martingale [50]. See Section 5.2 for the formal definition of a martingale.

To achieve this, flux probing relies on SPAD cameras whose pixels can time-stamp individual photon detections with picosecond precision relative to an internal real-time clock.

The probing operation. Proposition 4.1 tells us that we can always probe the flux function to recover a (noisy) measurement of its inner product with practically any other function. Moreover, probing is efficient to compute from the timestamp stream and can be thought of as a continuous-time and sync-free generalization of compressive acquisition schemes for conventional photon-counting histograms [221, 80, 130]. In particular, let \mathcal{T} be the stream of real-valued absolute timestamps up to time t and let $p(t)$ be an arbitrary known and bounded function:

Proposition 4.1 (Flux Probing Equation)

The inner product of the probing function $p(t)$ and the unknown flux function $\varphi(t)$ over the time interval $[0, t]$ satisfies the relation

$$p(\mathcal{T}) = \langle p, \varphi \rangle + M_p(t) \quad (4.3.2)$$

where $p(\mathcal{T})$ are “probing measurements” which sum the values of the probing function at the absolute timestamps

$$p(\mathcal{T}) \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}} p(\tau) , \quad (4.3.3)$$

$M_p(t)$ is a martingale, and the inner product is defined as $\int_0^t p(u)\varphi(u) du$.

Fundamental limit on bandwidth. We prove in Appendix B.2.3 that probing flux frequency f with $f > 1/2Q$ yields aliased measurements that “wrap around” the frequency spectrum and are identical to—and indistinguishable from—lower-frequency measurements:

Proposition 4.2

Given timing resolution Q , the maximum recoverable frequency is $f_{\max} = \frac{1}{2Q}$.

Intuitively, Proposition 4.2 says that flux frequencies above $1/2Q$ are unrecoverable regardless of whether we detect a few photons or a million.

Noise model. Our model accounts for the inhomogeneous Poisson nature of photon detections and treats the general case of real-value timestamps. The model is valid for arbitrary flux levels within the low-flux regime and, as we show in Section 4.6.1, it remains valid for low-count acquisitions (*e.g.*, as few as ten photons). More specifically, Proposition 4.3 tells us that the noise in probing measurements has a distribution that can be estimated from the timestamp stream through another probing operation. Thus, probing gives the means both to observe a flux function and to quantify the uncertainty of that observation:

Proposition 4.3 (Distribution of Probing Measurements)

The probing measurements $p(\mathcal{T})$ are approximately normally distributed with mean $\langle p, \varphi \rangle$ and variance $\langle p^2, \varphi \rangle$.

Fourier probing noise. Corollaries 4.1 and 4.2 allow us to quantify the accuracy by which specific flux frequencies can be estimated from a given timestamp stream:

Corollary 4.1 (Distribution of Fourier Probing)

The Fourier probing measurements $p_f(\mathcal{T}) = \sum_{t \in \mathcal{T}} \exp(-j2\pi ft)$ approximately follow a complex normal distribution with mean and covariance matrix

$$\mu = \begin{bmatrix} \langle \cos(2\pi ft), \varphi(t) \rangle & \langle -\sin(2\pi ft), \varphi(t) \rangle \end{bmatrix} \quad (4.3.4)$$

$$\Sigma = \begin{bmatrix} \langle \cos^2(2\pi ft), \varphi(t) \rangle & 0 \\ 0 & \langle \sin^2(2\pi ft), \varphi(t) \rangle \end{bmatrix}. \quad (4.3.5)$$

Corollary 4.2 (Distribution of Fourier Probing Energy)

The normalized energy of the Fourier basis probing measurements

$$p_f^{\mathcal{E}}(\mathcal{T}) \stackrel{\text{def}}{=} \text{Re} \left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{1,1}}} \right]^2 + \text{Im} \left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{2,2}}} \right]^2 \quad (4.3.6)$$

follows a non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter $\mu_1^2/\Sigma_{1,1} + \mu_2^2/\Sigma_{2,2}$.

We show in Appendix B.2.8 that unbiased estimators of the parameters of the above distributions can be obtained via probing.

Frequency detection. Given the estimated distribution of the Fourier Probing Energy, we derive the constant false alarm rate (CFAR) detector [213] (see Appendix B.2.9) to identify and remove noisy frequencies based on a desired probability of false alarm α . False alarms occur when we keep f and $E[|p_f(\mathcal{T})|] = 0$; we remove f if $p_f^{\mathcal{E}}(\mathcal{T})$ is lower than $\text{CDF}_{\chi^2}^{-1}(1 - \alpha)$, derived from Corollary 4.2. Specifically, we detect frequencies for which

$$|p_f(\mathcal{T})|^2 \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \frac{N(t)}{2t^2}. \quad (4.3.7)$$

We can interpret the threshold in Eq. (4.3.7) as the significant level for rejecting the null hypothesis that frequency f is a noisy frequency. Note that (1) for a fixed α , the probability of detecting a frequency is proportional to the total number of photons detected, and (2) for flux functions dominated by a particular frequency such that $|p_{f_i}(\mathcal{T})|^2$ is large, $N(t)$ also tends to become proportionally larger, reducing the probability of detecting other frequencies.

Implications for passive single-photon imaging. Our flux probing theory implies that rather than being a hindrance, sync-less imaging with absolute timestamps confers an extreme band-

width advantage to SPAD cameras: systems with 16-picosecond resolution, such as our own, can *simultaneously* acquire flux variations that span the entire DC-to-31 GHz range of frequencies, and that are due to *any number of unknown light sources* operating independently. This bandwidth is orders of magnitude broader than intensity cameras—SPADs or otherwise—were thought capable of acquiring directly [79, 100, 119, 14], *i.e.*, without resorting to homodyne [79] or heterodyne [100, 119, 14] detection schemes. While Proposition 4.2 describes *reconstructability* (*i.e.*, frequencies above the limit are unreconstructible), Proposition 4.3 and Eq. (4.3.7) provide insights about the *accuracy* and *detectability* of flux variations at different frequencies. Lastly, although we have not verified the theorized DC-to-31 GHz bandwidth experimentally due to unavailability of lasers that are fast enough, we show several real-world demonstrations of simultaneous DC-to-16.9 GHz imaging under very challenging low-flux conditions (see Figure 4.1.1, Sections 4.5.2-5.6.1). These validate our noise models in Eqs. (4.3.4)-(4.3.6) and (partially) confirm our theoretical bound.

4.4 Flux Function Reconstruction

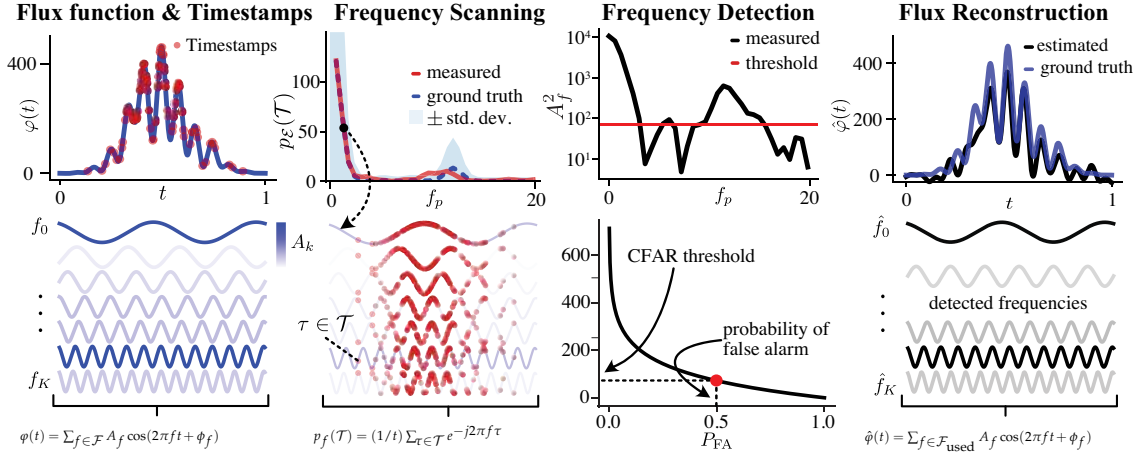


Figure 4.4.1: Visual illustration of the algorithm. *Flux function:* A flux function of finite spectral support can be expressed as a sum of sinusoids and produces a stream of absolute timestamps. *Frequency scanning:* We probe the flux function using a Fourier basis and measure the response at each frequency. *Frequency detection:* For each of the probed frequencies, we detect whether it contributes to the flux function if its corresponding amplitude is greater than a threshold (top) which is selected to achieve the desired probability of false alarm. (bottom). *Flux reconstruction:* Finally, we reconstruct a continuous-time flux function from the amplitudes and phases of detected frequencies.

Our probing theory leads directly to an algorithm that reconstructs flux function by first frequency-scanning the entire DC-to-GHz bandwidth with a step size of Δf to recover the Fourier transform of the flux function (Figure 4.4.1). This algorithm is similar to the Fourier-domain histogramming technique used in conventional active settings [130], but also differs in two key respects: (1) it provides a principled way to estimate frequency uncertainty in an acquired timestamp stream, and (2) it enables tractable operation in a regime involving potentially billions of candidate frequencies—*e.g.*, a 1 Hz-resolution scan of DC-to-20 GHz—by rejecting spurious frequencies and reducing storage requirements. The frequency detection step uses the CFAR detector of Section 4.3.2, where we set α so that the expected number of false alarms is less than 1. Figure 4.4.1 includes a visual depiction of the flux reconstruction algorithm and Algorithm 4.1 provides a complete description in

pseudocode.

4.4.1 Choosing the step size Δf for flux probing

We discuss how to set the frequency step size Δf for frequency scanning in procedure FLUXREC (Algorithm 4.1). We assume that the flux function has a known bandwidth $f \in [0, f_{\max}]$ and was observed for an interval $[0, t]$. Since a finite exposure time corresponds to windowing the flux function, the measured Fourier spectrum is the result of convolving $\varphi(t)$ with a sinc function whose full width at half maximum (FWHM) is approximately $1.2/t_{\text{exp}}$ Hz [176]. We therefore scan the desired range of frequencies with $\Delta f = 0.6/t$ Hz in order to ensure that we “hit” the main lobe of every sinc function contributed by each frequency. We use this scheme for the entire frequency range except near DC, where the Fourier-domain sinc lobe corresponding to the interval $[0, t]$ can interfere with detection of low frequencies when the DC component of flux is strong. In such cases, we use $\Delta f = 1/t$ instead. Note that our scheme is designed to balance resolution and computational efficiency, and that finer-grained frequency scanning is also possible (*e.g.*, as employed in astronomical imaging [53]).

Algorithm 4.1: FLUXREC—flux probing with negligible dead time

Input: $\mathcal{T}, t, f_{\max}, \alpha$
Output: $\hat{\varphi}(t)$
Frequency scanning
 $\Delta f \leftarrow 0.6/t$
 $\mathcal{F} \leftarrow$ frequencies from Δf to f_{\max} with step Δf
for $f \in \mathcal{F}$:
 $p_f(\mathcal{T}) \leftarrow \frac{1}{t} \sum_{\tau \in \mathcal{T}} e^{-j2\pi f\tau}$
Frequency detection
 $\mathcal{F}_{\text{used}} \leftarrow \emptyset$
 $A_0 \leftarrow N(\mathcal{T})/t$
for $f \in \mathcal{F}$:
 $A_f \leftarrow 2|p_f(\mathcal{T})|, \phi_f \leftarrow \angle p_f(\mathcal{T})$
 Reject f using CFAR (Eq.(4.3.7))
 if f is not rejected:
 $\mathcal{F}_{\text{used}} \leftarrow \mathcal{F}_{\text{used}} \cup \{f\}$
Flux reconstruction
 $\hat{\varphi}(t) \leftarrow \sum_{f \in \mathcal{F}_{\text{used}}} A_f \cos(2\pi ft + \phi_f)$
return $\hat{\varphi}(t)$

4.5 Experiments

We validate our flux probing theory for passive ultra-wideband imaging experimentally with (1) passive ultra-wideband imaging of both 1D intensity signals and 2D video signals ranging from DC to 16.9 GHz and (2) passive non-line-of-sight (NLOS) video via MHz-rate flux function reconstruction. These results are summarized in Figure 4.5.1.

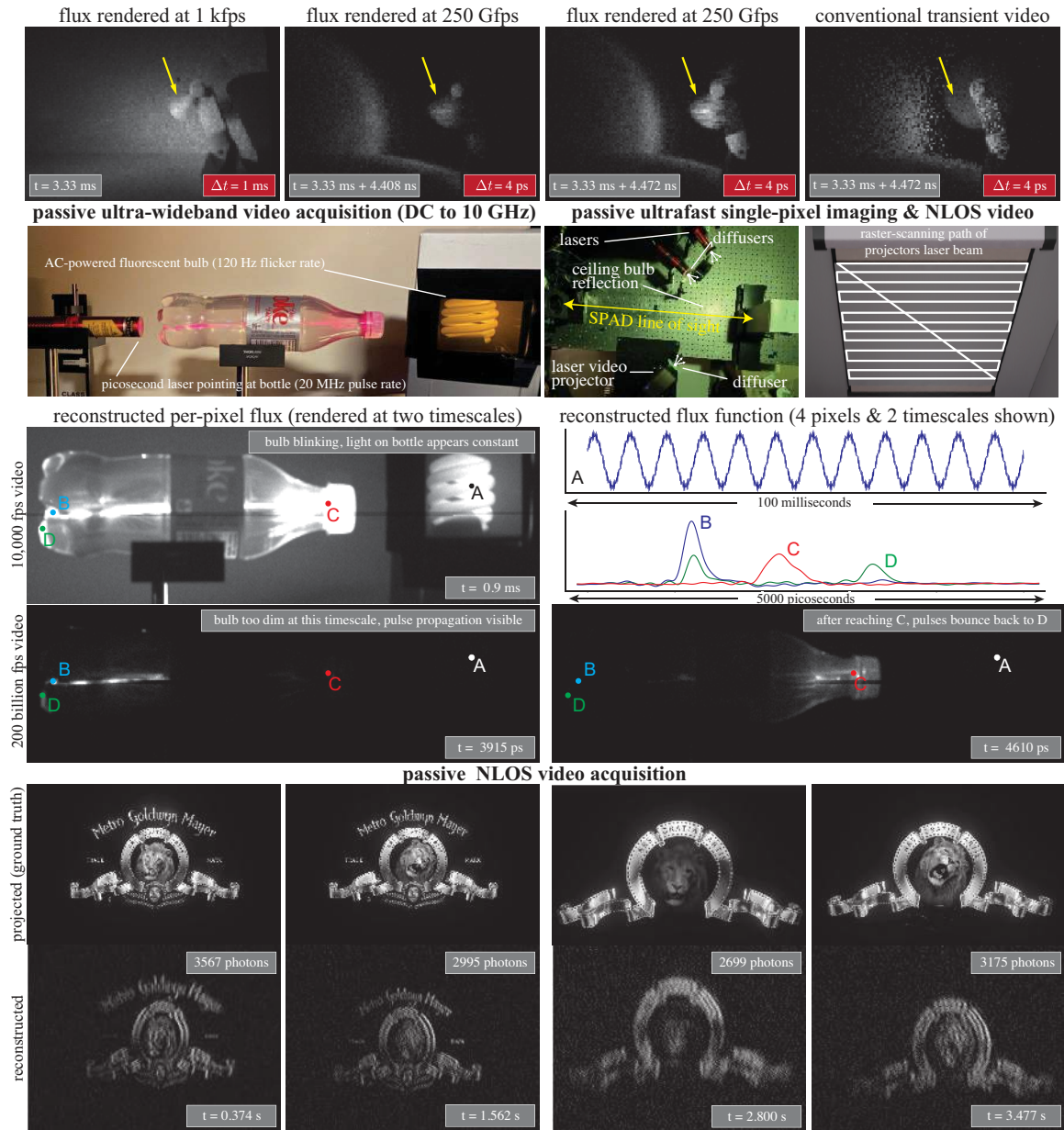


Figure 4.5.1: Passive ultra-wideband imaging experiments. **Row 1:** Simultaneous recovery of spinning fan and light propagation (see text for experiment setup). Rendering the flux function at 1 kfps reveals motion of the fan (yellow arrow), but light propagation is invisible; at 250 Gfps light propagation is visible, and the fan freezes. Conventional histogramming (right) synced to the laser fails to recover the (unsynced) fan rotation. **Row 2:** Experimental setup for ultra-wideband video acquisition (left) and NLOS video imaging in a scene with multiple light sources (right). **Rows 3–4:** Flux function images at two timescales (left, bottom right) and for different points in the scene illuminated by a pulsed laser and flickering light bulb (top right). The three peaks at B, C, and D correspond to a light pulse entering the bottle (B) propagating to the cap (C) and reflecting back (D). **Rows 5–6:** Passive NLOS acquisition using a raster-scanning laser projector with ground truth and reconstructed frames.

4.5.1 Hardware

The following sensors and light sources are used in our experiments:

- *Free-running SPAD*: We use a single-pixel PDM Series Fast-Gated SPAD from Micro Photon Devices, operated in asynchronous mode. The SPAD has a $50 \times 50 \mu\text{m}$ active area, a timing jitter of 68 ps, a 231 ns dead time and a quantum efficiency of 18% at 660 nm.
- *Time-to-digital converter (TDC)*: The SPAD's output is converted into a stream of timestamps by a PicoQuant PicoHarp300 system. Except where noted, the PicoHarp is operated in asynchronous mode and is not connected to any laser sources. The PicoHarp's timing resolution is $Q = 4$ ps and its overall timing jitter is 16 ps.
- *Galvo mirrors*: For experiments where 2D imaging was required, we used a pair of galvo mirrors from Thorlabs (Model GVS012) to acquire measurements over a 2D field of view with our single-pixel SPAD.
- *Laser scanning video projector*: We used an AnyBeam MEMS raster scanning laser projector (model HD301M1-H2-01) for some of our imaging experiments. Its nominal frame rate is 60 Hz, its resolution is 1280×720 and its nominal brightness is 30 ANSI lumens. Video was supplied via an HDMI port.
- *Miscellaneous bulbs*: We used three lamps for additional illumination: (1) a Kasa Smart Wi-Fi LED Multi-Color Light Bulb (800 lumens), (2) an off-the-shelf compact fluorescent lamp, and (3) our laboratory's ceiling lights, consisting of twelve T8 replacement LED lamps (SYLVANIA OCTRON, FO28/835XP/XL/SS/ECO3, 2600 lumens, driven by SYLVANIA QUICKTRONIC high-efficiency ballasts).
- *Spinning fan*: We used a handheld, battery-operated spinning fan operating at a frequency of approximately 54 Hz.
- *Color filters*: For acquisition of color ultra-wideband video, we used red, green, and blue transparent color correction lighting gel filters from SAKOLLA.
- *Avalanche Photodiode*: Thorlabs Model DET10A2.
- *Collimated Laser Diode*: Thorlabs Model CPS532-C2.
- *Ultrafast low-power pulsed lasers*: We used two Alphas Lasers picosecond lasers for most of our passive ultra-wideband imaging experiments, both having a wavelength of approximately 665 nm (models PLD670-100 and PLD670-50). The lasers have a user-selectable power setting, a user-selectable pulse repetition rate between 1 MHz to 100 MHz, and a nominal pulse FWHM of 83 ps for the power setting we employed.
- *Ultrafast high-power pulsed laser*: We also used a class-4 picosecond laser with a wavelength of 532 nm (NKT Photonics Katana 05HP). The repetition rate was 10 MHz and its FWHM was 58 ps under the experimental conditions we employed, resulting in a flux function with frequency support up to 17.24 GHz.
- *Experimentally-verified top imaging speed*: The upper bound is limited by the speed of our fastest light source—*i.e.*, the NKT Photonics Katana 05HP—not our method or our SPAD.

4.5.2 1D passive ultra-wideband imaging of Figure 4.1.1 scene

We demonstrate recovering signals with frequencies spanning roughly 9 orders of magnitude from DC to 10 GHz (Figure 4.1.1). We place a single-pixel SPAD in the scene to capture flux variations

from (1) pulse-width modulation of a light bulb (900 Hz), (2) backscattered light from a raster-scanning laser projector (60 Hz–10 MHz), and (3) two unsynchronized picosecond lasers (40 MHz–10 GHz). Remarkably, the flux function is reconstructed from only 77,000 photon timestamps.⁶ We recover time-varying flux across billions of frequencies from this minuscule set of photons.

Scene: The scene is shown in Figure 4.5.1 (row 2, second image from right). The scene illumination consisted of the laser projector, our smart light bulb and our two low-powered picosecond laser sources. At any given instant, all light sources behaved as diffuse point emitters either by design (*i.e.*, smart bulb) or because of a diffuser placed in front of them. The SPAD was focused on a single point on a planar target as shown in the photo.

Illumination: We used the projector to play a five-second MGM video clip at a nominal frame rate of 60 frames per second. A Thorlabs ground-glass diffuser was placed in front of the laser projector so that light is diffused from each projector pixel as would happen with light bouncing off of a line-of-sight (LOS) wall. A similar diffuser was placed in front of the pulsed laser sources, which were operating at 3 MHz and 40 MHz, respectively. The smart bulb was experimentally determined to flicker at approximately 900 Hz. The smart bulb was located two meters above the optical table shown in the photo.

Acquisition: We acquired a single timestamp stream over a one-second acquisition interval. A total of 770k photons were detected. To make flux reconstruction even more challenging, we applied Bernoulli thinning with $p = 0.1$ to the captured timestamps, resulting in a single stream of 77k timestamps in total. This stream was used for probing and for the results shown in Figure 4.1.1.

Frequency scanning and detection: For frequencies between 0 Hz and 3 MHz we probed with a step of 1 Hz, taking into account the 77k timestamps spanning the full one-second acquisition interval. For frequencies between 3 MHz and 10 GHz we probed with a step of 6 Hz using only the approximately 7.7k timestamps acquired in the first 100 ms of the acquisition. This was done for computational efficiency.⁷ For each probed range, we set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$, where #freqs is the number of probed frequencies. This ensures that only one false alarm is returned by the CFAR detector on average.

Flux function visualizations: To plot the reconstructed flux function at the four timescales shown in Figure 4.1.1, we sampled it over four different temporal windows—1 s, 5.5 ms, 1 μ s, and 2.7 ns. For temporal windows less than one second, we centered them at $t = 500.983$ ms. This time instant was chosen to highlight the asynchronous relation of the two laser pulses. The flux function was evaluated at 10k regular samples over each temporal window. These samples were then low-pass filtered with frequency cutoffs at 5 kHz, 100 kHz, 5 GHz, and 1.8 THz, respectively, to avoid aliasing in each plot.

Figure 4.5.2 shows an enlarged and more detailed view of the flux spectrum shown in Figure 4.1.1, where the detected frequencies and the CFAR thresholds employed are shown, for both the 0-to-3 MHz and the 3 MHz-to-10 GHz bands.

⁶Conventional camera pixels collect a few thousand photons to return a single estimate of light intensity.

⁷Probing the 3 MHz-to-10 GHz bandwidth with 6 Hz steps took 100 minutes on a 383 GB Dual 14-Core CPU running at 2.2GHz and a 24 GB Titan RTX GPU. Note that the computational cost increases quadratically with exposure time because it increases the number of timestamps processed and also reduces the step size employed in frequency scanning.

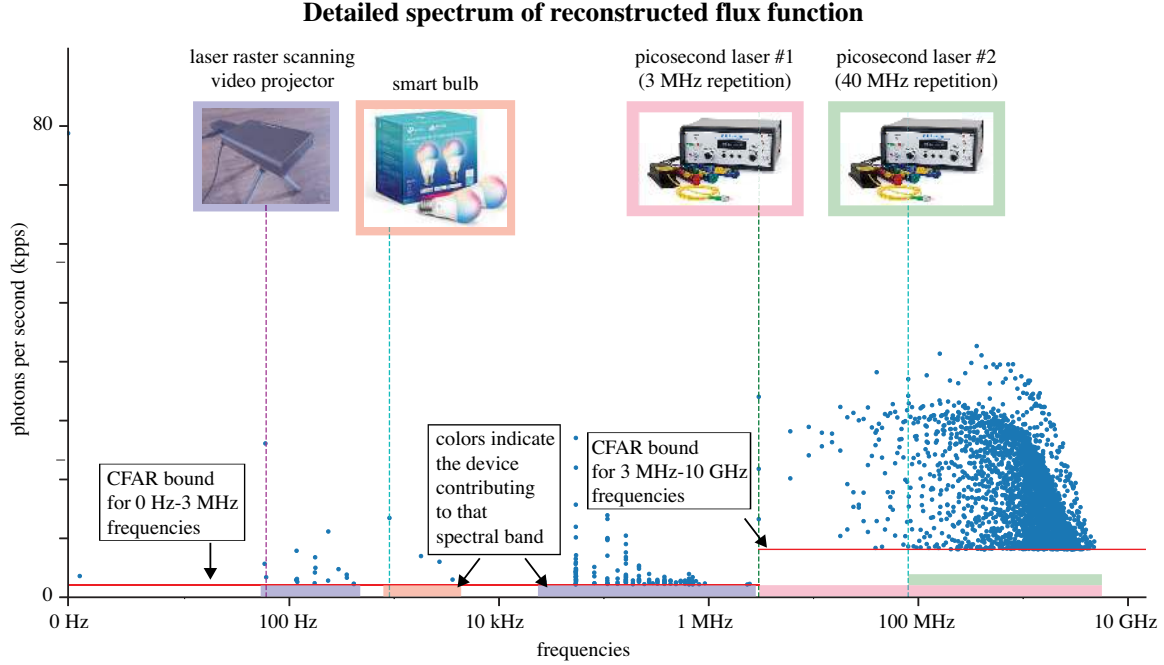


Figure 4.5.2: Detailed spectrum of the reconstructed flux function for experiment in Figure 4.1.1.

A microscope in time. We demonstrate the ultra-wideband capability of our flux reconstruction through an interactive visualization of the 1D flux function reconstructed by the experiment in Figure 4.1.1. This was made possible by performing an analytical integration and efficient sampling of the function $\hat{\phi}(t)$ defined in Figure 4.4.1 using OpenCL, enabling real-time zoom-in and zoom-out in time. To create the interactive plot, the function is sampled at 1024 regular intervals for the time window currently being displayed. Snapshots from this visualization are shown in Figures 4.5.3 and 4.5.4.

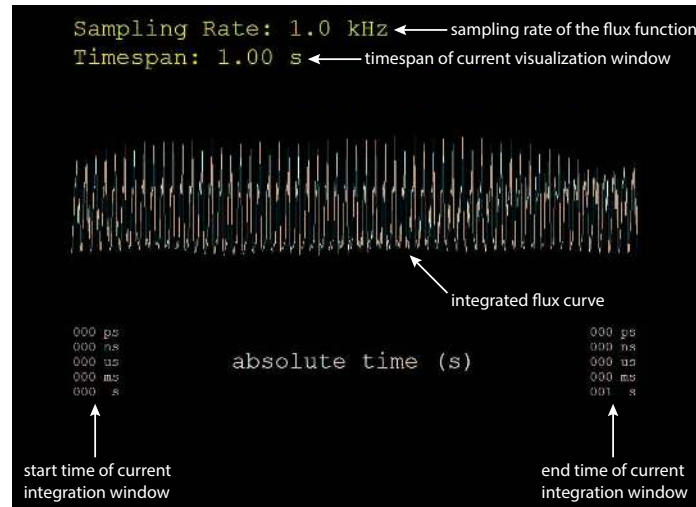


Figure 4.5.3: Visualization of interactive flux function using OpenCL.

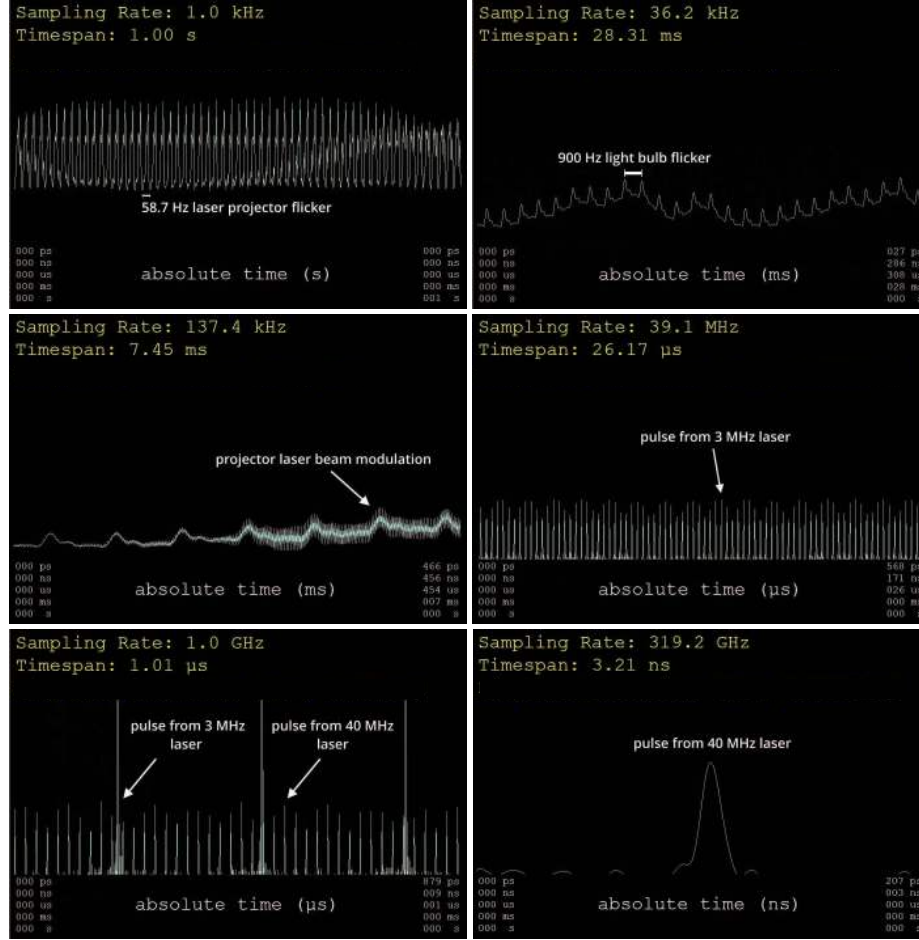


Figure 4.5.4: Visualization of interactive flux function using OpenCL. **Top left:** At the span of one second, we can see the 58.7 Hz refresh rate of the laser projector. **Top right:** Over tens of milliseconds, we can see the 900 Hz flicker of the light bulb. **Middle left:** Zooming further, at a couple of milliseconds, we can see the time-varying modulation of the projector’s laser beam. **Middle right:** At the scale of microseconds, we can observe the multiple pulses of the 3 MHz laser. **Bottom left:** Zooming even further, we can see the pulses from the asynchronous 3 MHz and 40 MHz lasers. **Bottom right:** At the finest timescale, over a couple of nanoseconds, we can resolve an individual laser pulse of the 40 MHz laser.

4.5.3 1D passive ultra-wideband imaging over room-size distances

In this experiment we demonstrate passive reconstruction of multiple concurrent high-frequency flux variations from DC-to-16.9 GHz over room-sized distances by employing a brighter and faster laser.

Scene: Our SPAD was focused at a spot on one of our laboratory’s diffusely-painted walls, approximately one meter away from the SPAD. The spot, as well as the entire laboratory, was illuminated simultaneously by two emitters:

- a very bright, diffusely-reflecting laser spot approximately four meters away from where the SPAD was focused; this laser spot was created by aiming our high-power pulsed laser to a second diffusely-painted wall in the laboratory, and
- our laboratory’s ceiling lights, the closest of which was approximately two meters away from where the SPAD was focused; these lights exhibited a very faint high-frequency flicker due to

their electronic high-efficiency ballasts.

Acquisition: We used an exposure of two seconds for timestamp stream acquisition. The total number of photons detected was 940k.

Frequency scanning and detection: We probed frequencies from 0 Hz to 100 Hz with a step size of 0.5 Hz to avoid detection of frequencies corresponding to the sinc lobe centered at DC. In addition, we probed frequencies from 100 Hz to 17 GHz with a step size of 0.3 Hz. As in the experiment of Figure 4.1.1, we set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$ for each probed range, where #freqs is the number of probed frequencies.

The recovered spectrum of the flux function is shown in Figure 4.5.5. It is evident from the spectrum that we can passively reconstruct flux variations well above 10 GHz with our SPAD: the maximum frequency detected was 16.9 GHz. Moreover, despite the fact that the 85.1 kHz flicker emitted our laboratory’s LED lighting was very faint, it was still recoverable despite our laser’s simultaneous emissions. Lastly, note that the spectrum contains a single outlier at 85 Hz, as its amplitude was just above the bound set by the CFAR detector.

Flux function visualization: We plot the reconstructed flux function at two timescales—a 60 μ s temporal window (Figure 4.5.6, left) and a 2.5 ns window (Figure 4.5.6, right). The former window captures the component of the flux function due to the flickering ceiling lamps; the latter shows the reconstruction of a single laser pulse whose estimated FWHM is 58.9 ps. To create these plots for each window, we evaluate the flux function at 1k regular samples, with frequency cutoffs at 8.3 MHz and 200 GHz, respectively, to avoid aliasing.

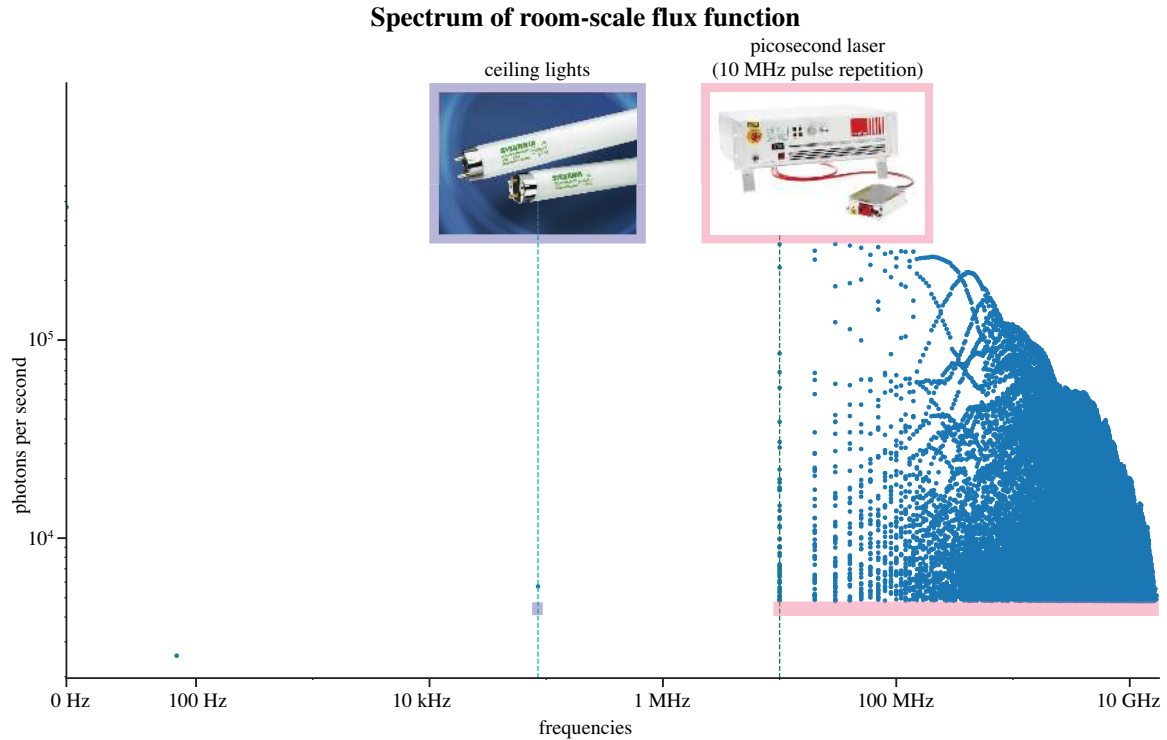


Figure 4.5.5: Spectrum of the reconstructed flux function from our room-scale passive ultra-wideband imaging experiment. Our method successfully reconstructs flux due to the flickering LED ceiling lamps (85.1 kHz), the dominant frequency of the pulsed laser (10 MHz), as well as the laser’s higher harmonics (up to 16.9 GHz).

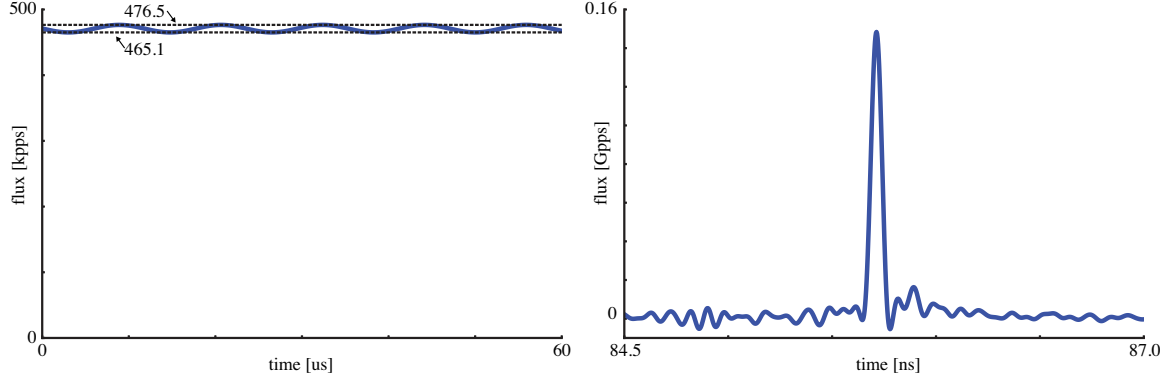


Figure 4.5.6: Flux function visualization for our room-scale experiment. **Left:** Flux component due to the LED ceiling lamps’ faint (2.4%) flicker at 85.1 kHz. **Right:** Flux component due to a picosecond laser pulse (58.9 ps FWHM).

4.5.4 Passive reconstruction of laser pulse trains

Here we demonstrate passive unsynchronized reconstruction of picosecond-scale laser pulse trains with high accuracy.

Scene: Our scene consists of a planar target, approximately one meter away from the SPAD, illuminated by a diffused low-power pulsed laser operating at 20 MHz. The SPAD is focused on a single point on the planar target. We use the same setup as in Figure 4.1.1, but with only one laser source turned on.

Active histogram-based imaging: To acquire a conventional photon-count histogram that captures the shape of the laser pulse, we synchronized the SPAD to the laser and operated the PicoHarp as a Time-Correlated Single-Photon Counter. We used a one second exposure for this procedure, detecting 267k photons in total.

Flux probing: We used the first 2657 raw (*i.e.*, absolute-time) timestamps from the PicoHarp as a timestamp stream for the purpose of probing. This corresponds to an acquisition time of approximately 10 milliseconds.

Frequency scanning and detection: We probed frequencies from 0 Hz to 11 GHz with a step of 60 Hz. This frequency upper bound was motivated by the simulations in Section 4.6.2 which indicated that a maximum frequency of 11 GHz is sufficient to resolve pulses with a FWHM close to our laser’s 83 ps FWHM. As in the experiments described previously, we set the probability α of false alarm to $\frac{1}{\#freqs}$, where $\#freqs$ is the number of probed frequencies.

Pulse train visualization: We sample the flux function at 1 million equidistant points across one 50-nanosecond period of the laser. This corresponds to a sampling rate of 20 THz. Figure 4.5.7 (top) depicts the reconstructed laser pulse train obtained from our method. We note that reconstruction of such a train is not possible using a conventional photon-count histogramming technique because their reliance on relative times effectively time-wraps the entire pulse train into a single pulse. Figure 4.5.7 (bottom-left) shows a zoomed-in view of an individual pulse from the pulse train, reconstructed by flux probing (estimated FWHM: 76 ps, peak location: 4.656 ns) along with the reconstruction obtained from histogramming (estimated FWHM: 72 ps, peak location: 4.648 ns). Figure 4.5.7 (bottom-right) shows a zoomed-in view of these reconstructions. Overall, the shapes

are very similar even though probing uses $100\times$ fewer photons. Additionally, we observe a slight 8 ps shift between the pulses reconstructed by probing versus histogramming. This shift is well below the 16 ps jitter of our TCSPC and thus not significant.

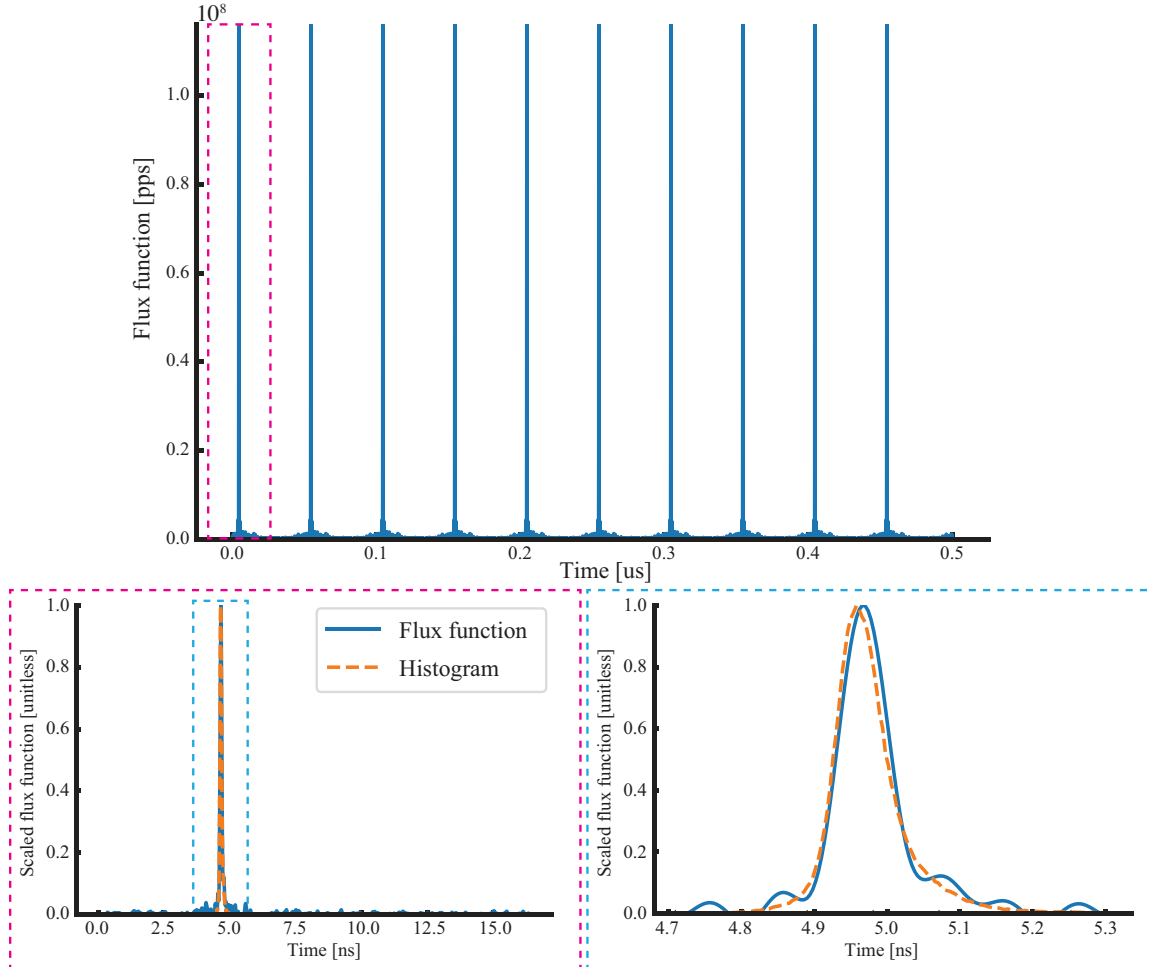


Figure 4.5.7: Passive reconstruction of picosecond laser pulse trains. **Top:** Reconstructed laser pulse train. **Bottom left:** One reconstructed pulse from the train, along with the single pulse reconstructed by active photon histogramming. **Bottom right:** Zoomed-in view of the laser pulses shown on the left.

4.5.5 2D fan experiment: freezing time at all timescales

We demonstrate ultra-wideband video (Figure 4.5.1, top) by raster scanning a scene in which a pulsed laser, with 20 MHz repetition and 80 ps FWHM, is diffused to illuminate a fan spinning at 54 Hz. We detect frequencies from DC to 10 GHz and render flux functions at 1 kfps and 250 Gfps, showing both the fan blades rotating and the propagation of the laser pulse. In contrast, conventional approaches reconstruct the scene at only one of the aforementioned frame rates, temporally blurring either slow or fast events. Furthermore, our method can render the flux at whatever timescale, essentially *freezing time at all timescales*. Note that because we only had access to a single pixel SPAD, the timestamps were collected by scanning across the field of view of the SPAD. To temporally align the flux functions *between pixels*, we use synchronization signals from both the laser and the fan. We emphasize that no synchronization signals were used to reconstruct the flux

functions themselves—we are demonstrating a new capability of reconstructing the appearance of the scene *as it appeared during each laser pulse*—this is distinct from the use of synchronization and histogramming to estimate the *average appearance* of the flux function over time [171, 250, 67]. We also emphasize that the images are rendered by integrating the flux function over the exposure of each frame. As such, these images exhibit not only high dynamic range, but their intensities are also expressed in physical units of photons, thereby ensuring radiometric calibration by nature.

Scene: The scene itself is shown in Figure 4.5.8. It consisted of a battery-operated spinning fan and two white cardboard panels placed at a right angle. The scene was approximately 400 mm \times 250 mm in size and approximately one meter away from the SPAD.

Illumination: We operated the laser at a pulse repetition rate of 20 MHz for this experiment. A Thorlabs ground-glass diffuser was placed in front of the laser so that the whole scene was illuminated by a diffuse ultrafast point source.

Scanning and timestamp stream collection: Since we only have a single-pixel SPAD available, we used the galvo mirrors to scan the field of view with a spatial resolution of 3 mm, resulting in an image with a resolution of 151 by 91 pixels. We used an exposure time of $t_{\text{exp}} = 0.1$ s for timestamp acquisition at each galvo position. Photon counts ranged from approximately 13k photons at the brightest pixels (*i.e.*, on the bottom left point of the fan) to 5k photons at pixels on the background white plane, to far fewer photons elsewhere in the scene.

Temporal pixel alignment: In order to simulate concurrent imaging, the flux functions of all pixels must be computationally shifted relative to each other to create the effect of a common absolute clock. We use the laser sync to provide temporal alignment of the flux functions within the laser repetition period. However, to ensure alignment of the pixels' absolute times, we need to also have a correspondence between pulses, *i.e.*, that the i -th pulse of pixel a corresponds to the j -th pulse of pixel b . For that, we use a collimated DC laser diode that is interrupted by the fan blades at each rotation. This provides a timing signal that is sufficient for such a temporal pixel alignment. In particular, we consider a pixel's pulse 0 to be the pulse closest to the time instant when the fan interrupts the DC laser beam. To detect the on-off light signal from the DC laser, we use the avalanche photodiode (APD) connected to an oscilloscope that generates temporal markers at each on-off transition, which are then provided to the TCSPC. After reconstructing the flux function at each pixel independently, we use those markers to align all the flux functions onto a common absolute timeline.

Frequency scanning and detection: For the sake of efficiency, the frequency-scanning step of the algorithm in Figure 4.4.1 was performed for only one pixel in the scene, and the set of flux frequencies detected for that pixel were used to probe flux at all other pixels as well. Specifically, we selected the brightest pixel on the fan blades and scanned frequencies from 0 Hz to 10 GHz with a step size of 6 Hz. As usual, we set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$, where \#freqs is the number of probed frequencies.

We used the probing measurements at the brightest pixel on the fan blade to determine the fan's spinning frequency, f_{fan} , and the laser's dominant frequency, f_{laser} . Note that since the fan's two blades block the APD twice per full rotation, flux probing detected the frequency $f_{\text{fan}} = 108$ Hz, which was twice the fan's actual rotation. Using f_{fan} and f_{laser} , we expanded the set of probed

frequencies to also include (1) higher harmonics of the laser’s dominant frequency f_{laser} (*i.e.*, integer multiples of f_{laser} up to 15 GHz); (2) higher harmonics of the fan’s spinning frequency (*i.e.*, integer multiples of f_{fan} up to 100 kHz); and (3) linear combinations of f_{fan} , f_{laser} and their harmonics, computed by convolving the harmonics of f_{fan} and f_{laser} .

Rendering ultra-wideband video: The reconstructed flux functions are continuous and include variations from millisecond to picosecond timescales. To render discrete video frames at the finest resolution, we integrate the flux function over consecutive intervals of four picoseconds, corresponding to a frame rate of *250 billion frames per second*. At the coarsest timescale, frames are computed by integrating the flux function over one-millisecond intervals.

Acquired ultra-wideband videos: Since the flux functions reconstructed at each pixel are ultra-wideband, it is possible to “zoom in” and “zoom out” in time across vastly different timescales:

- *1 kfps*: The fan blades are spinning and perform five full rotations in 0.1 seconds. At this timescale, the fan appears to be illuminated by temporally- and spatially-constant illumination from the laser. Figure 4.5.9 shows 9 images at this timescale. Note the shadow of the fan’s blade on the background plane (middle row).
- *250 Gfps*: At the finest timescale, each laser pulse can be seen traversing the scene individually, with one laser pulse sweeping through the scene in one nanosecond. The fan blades appear to be frozen, and the pulse appears to propagate over them. Note that we can render such a video sequence for each individual laser pulse, as the fan blades slowly change position. Figure 4.5.10 shows 9 frames at this timescale, where a single laser pulse traverses the scene while the fan blades appear frozen. Figure 4.5.11 shows the results obtained from histogramming the relative timestamps as obtained from the TCSPC. Our method provides sharper transients and can also recover the blades of the fan, whereas conventional photon histogramming blurs them. Figure 4.5.12 shows a different laser pulse propagating through the scene 2.2 milliseconds later, hence the different positions of the fan’s blades.

Color ultra-wideband video experiment: We also capture a color ultra-wideband video of the scene using color filters. Specifically, we used red, green, and blue color filters, scanning the scene sequentially by placing one color filter in front of the SPAD at a time. To account for the time-consuming scan, we operated the fan with an AC power supply. Because the AC power leads were reversed relative to the fan’s internal battery, the fan spun counter-clockwise for this experiment. To reconstruct the flux function, we repeated the same procedures outlined in the previous experiment. Due to the imperfect response of the color filters, we applied manual scaling to the red, green, and blue channels. In particular, we multiplied the red channel by 1, the green channel by 0.75, and the blue channel by 0.4.

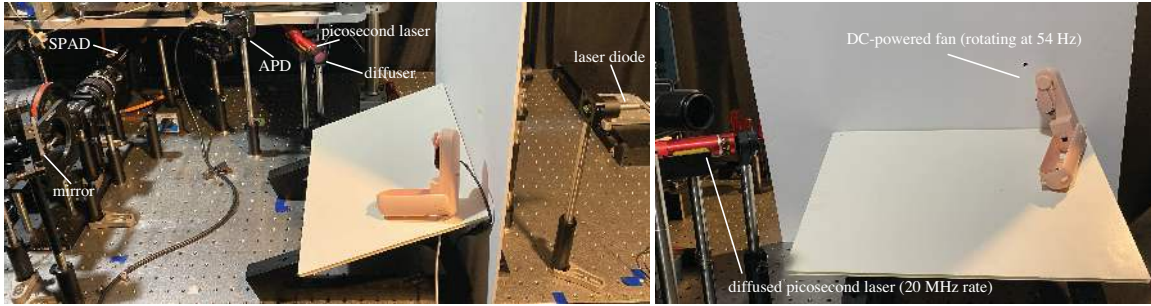


Figure 4.5.8: Setup for the spinning fan experiment. Left: Synchronization setup. A collimated laser diode illuminates the fan from the back providing an on-off signal to the APD. Right: Photo of the experimental setup captured from the SPAD's viewpoint.



Figure 4.5.9: Images show various stages of a fan spinning, rendered at 1 kfps.



Figure 4.5.10: Images show a picosecond laser pulse propagating over two planes and a fan. Note how the fan blades appear as frozen. The effective frame rate is 250 Gfps.

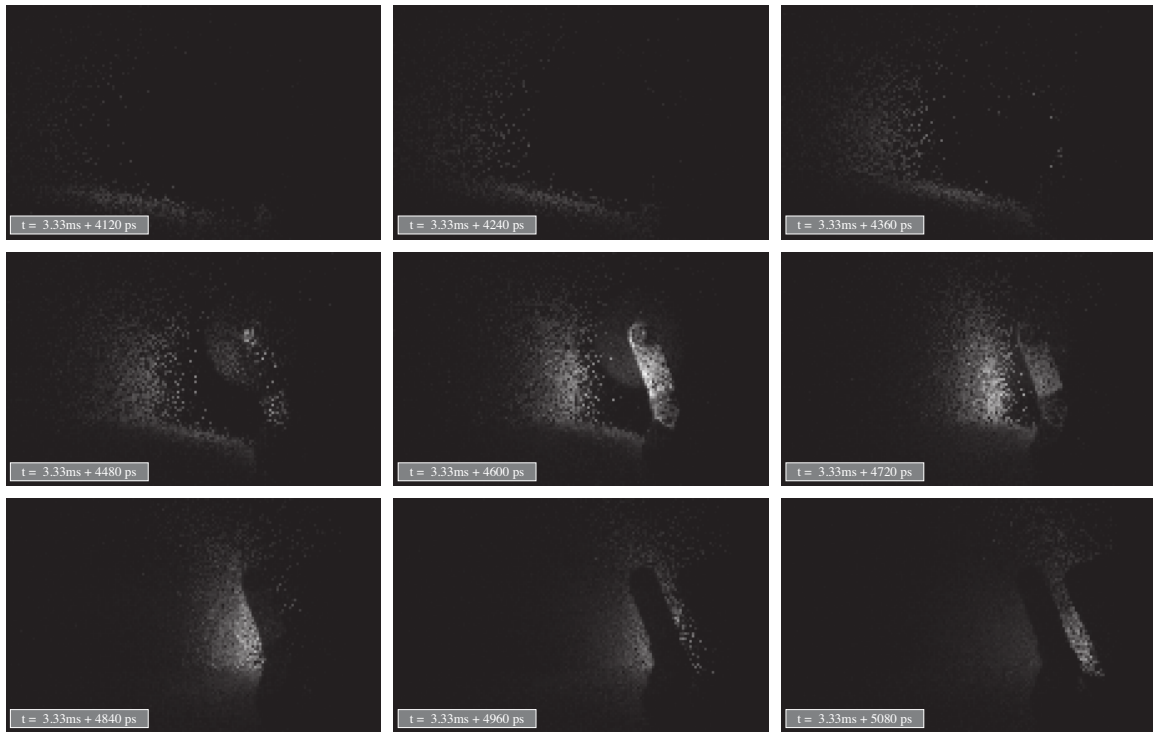


Figure 4.5.11: Images show a picosecond laser pulse propagating over two planes and a fan obtained from histogramming using the frequency of the laser. Note how the fan blades appear blurred.



Figure 4.5.12: Images show a picosecond laser pulse propagating over two planes and a fan. The fan blades are at a different position demonstrating the *unique* capability of our approach to *freeze time at all timescales*.

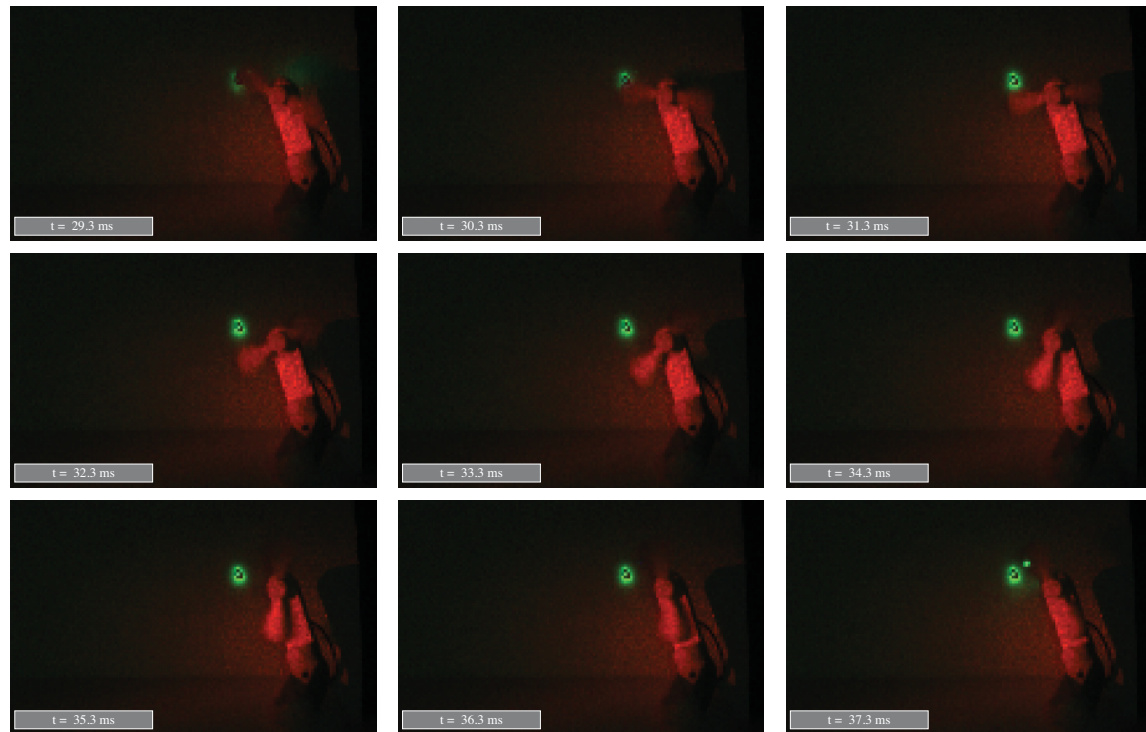


Figure 4.5.13: Color images of a fan spinning rendered at 1 kfps. Note how we can accurately recover the reflection of the green laser from the fan blade to the back plane (top left) and the moment at which the edge of the fan blade blocks the green laser beam (bottom right).



Figure 4.5.14: Color images depicting the propagation of a picosecond laser pulse across two planes and a fan. At this timescale, only the red pulsed laser contributes to the illumination of the scene.

4.5.6 2D Coca-Cola bottle experiment

We show another video example in Figure 4.5.1 (row 2), where the same picosecond laser illuminates a Coca-Cola bottle filled with water and a small amount of milk to scatter the light. Within the same scene, a compact fluorescent lightbulb (CFL) flickers at 120 Hz. We render videos at 10 kfps and 200 Gfps to visualize the CFL flicker and light pulses propagating through the bottle (Figure 4.5.1, rows 3–4). For the same reasons outlined before in Section 4.5.5, we use synchronization signals from the laser and the bulb.

Scene: The scene itself is shown in Figure 4.5.1 (row 2, left). It consisted of a Coca-Cola bottle filled with water and a small amount of milk, as well as a CFL light bulb connected to an AC outlet. The scene was approximately 392 mm × 121 mm in size and approximately one meter away from the SPAD.

Illumination: We operated the laser at a 20 MHz pulse repetition rate for this experiment. Due to its 60 Hz AC power supply, the CFL bulb flickered at 120 Hz. Both light sources were turned on during acquisition.

Scanning and timestamp stream collection: As in the ultra-wideband video acquisition experiment of Section 4.5.5, we used the galvo mirrors to scan the field of view with a spatial resolution of 3 mm, resulting in an image with a resolution of 392 by 121 pixels. We used an exposure time of $t_{\text{exp}} = 0.1$ s for timestamp stream acquisition at each galvo position. Photon counts ranged from approximately 37k photons at the brightest pixels (*i.e.*, near-specular transmission from the bottle’s surface to the SPAD) to 17k photons at pixels on the CFL bulb, to far fewer photons elsewhere in the scene.

Temporal pixel alignment: Since it was not possible to acquire timestamps simultaneously from all pixels in the field of view, we required an external mechanism to align the acquired flux functions across the field of view. To do this, we used synchronization signals from the laser and the bulb to align the estimated flux functions at picosecond and millisecond timescales. Specifically, we used the sync signal of the laser and an oscilloscope connected to the AC to produce timing markers that were detected by the PicoHarp. After reconstructing the flux function at each pixel independently, we used those markers to align all the flux functions onto a common absolute timeline.

Frequency scanning and detection: For the sake of efficiency, the frequency-scanning step of the algorithm in Figure 4.4.1 was performed for only two pixels in the scene and the flux frequencies detected for these pixels were then used to probe the flux frequencies at all other pixels. Specifically, we selected (1) a pixel on the CFL bulb and scanned frequencies from 0 Hz to 10 kHz with a step size of 10 Hz and (2) a pixel on the Coca-Cola bottle and scanned frequencies from 10 kHz to 10 GHz with a step size of 6 Hz. As usual, we set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$.

Rendering ultra-wideband video: The reconstructed flux functions are continuous and include variations from millisecond to picosecond timescales. To render discrete video frames at the finest resolution, we integrate the flux function over consecutive intervals of 5 ps, corresponding to a frame rate of 200 billion frames per second. Figure 4.5.15 shows 10 frames rendered at this timescale, where a single laser pulse traverses the scene. To improve visibility, we scale each video frame’s intensity by a factor of two.

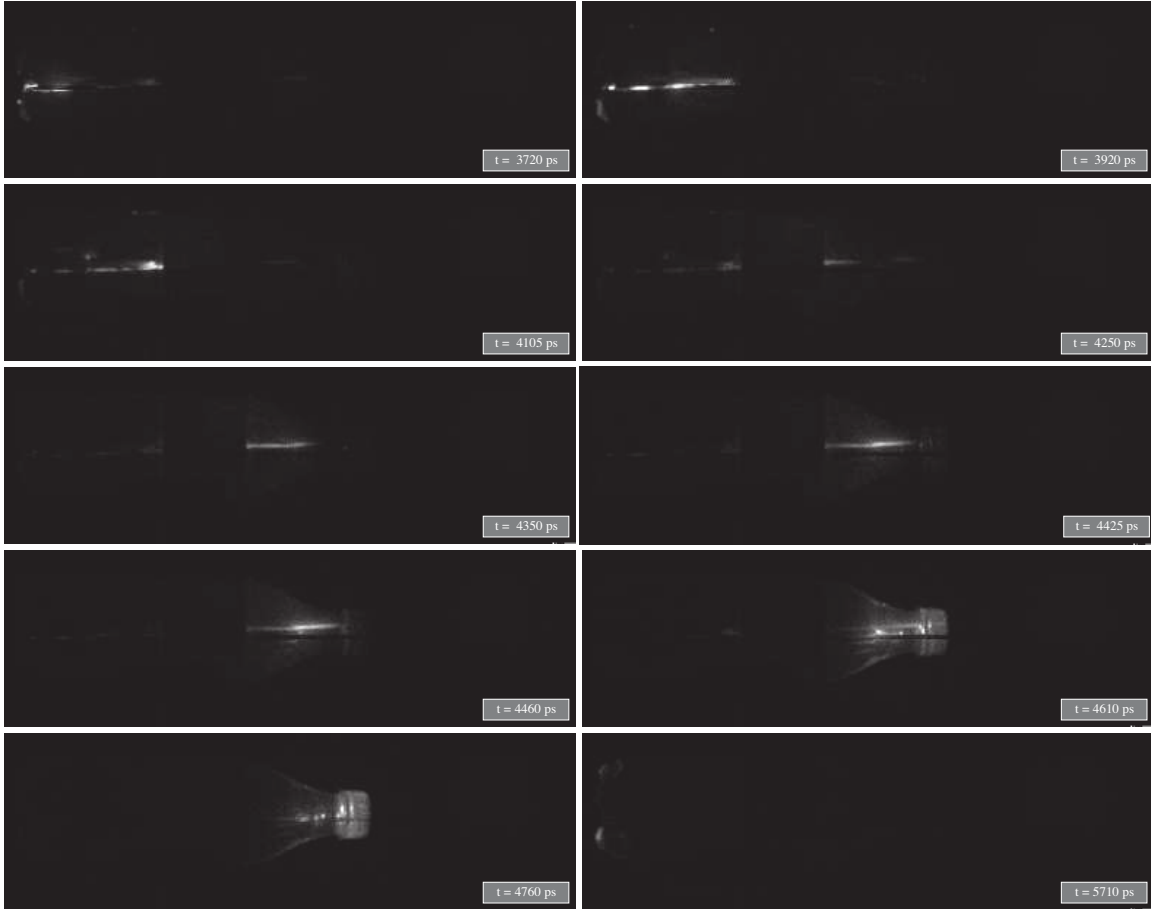


Figure 4.5.15: Passive transient imaging. Images show a picosecond laser pulse propagating inside a coca-cola bottle filled with water and a small amount of milk. Note how we reconstruct the light bouncing off the cap of the bottle and propagating all the way back to where it entered (last row right).

4.5.7 Recovering passive non-line-of-sight (NLOS) videos

We demonstrate passive NLOS video reconstruction using light measured indirectly from a raster-scanning laser projector (see illustration in Figure 4.1.1 and photo in Figure 4.5.1, row 2). The SPAD observes a *single* point on a diffuse box during the projector beam’s raster scan, collecting light that bounced twice (*i.e.*, diffuser→diffuse box→SPAD). This configuration is analogous to dual photography [219]. By reconstructing the 1D flux function over a one-second span, we recover the video being played. We show results for the multi-illumination setting of Figure 4.1.1 and for a projector-only setting. In the latter case, we recover fine details of each video frame (Figure 4.5.1, rows 5–6) even though only 3000 photons were collected on average during a frame’s $1/58$ s raster scan.

Scene. The scene is shown in Figure 4.5.1 (row 2, second image from right) except that the two lasers were turned off. A Thorlabs ground-glass diffuser is placed in front of the laser projector so that light is diffused from each projector pixel as would happen with light bouncing off of a line-of-sight (LOS) wall. The SPAD was focused on a single point on a planar target as shown in

the photo.

Illumination. We used the projector to play a five-second MGM video clip in a loop at a nominal frame rate of 60 fps. The projector was the only light source in this experiment.

Acquisition. We acquired a single timestamp stream over an exposure time $t_{\text{exp}} = 5$ seconds. A total of 959k photons were detected.

Frequency scanning and detection. In order to improve the computational efficiency of our method, we split the acquired timestamps into five one-second streams, each corresponding to a one-second exposure. The resulting streams contained 214k, 180k, 162k, 190k, and 213k timestamps, respectively. For each clip, we probed frequencies up to 40 MHz with a step of 1 Hz. We initially set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$, as in our other experiments. We observed, however, that setting the probability of false alarm to a slightly higher value, namely $\alpha = 0.3\%$, yielded a higher-quality reconstruction. We believe this is related to the fact that as the number of photons increase the CFAR bound also increases, making weaker flux frequencies harder to detect (see the discussion in Section 4.3.2).

Reconstructing individual video frames from the reconstructed flux function. We used the reconstructed flux function to reconstruct the video frames as seen from the projector’s point of view. This required converting the high-frequency 1D flux signal from the SPAD into the 2D image that the projector actually projected by raster-scanning onto the diffuser. We briefly describe this procedure below.

First, we used the detected frequencies to estimate the projector’s actual refresh rate. This was found to be 58.7 Hz, *i.e.*, slightly below the nominal video rate. This refresh rate defines the start time and the end time of each frame. Given these times, we regularly sampled the flux function at 1.177M (1280×920) points. While the resolution of the laser projector is 1280×720 , we sample with a resolution of 1280×920 to compensate for the time taken by the scanning laser to travel from the end of the raster scan to the beginning. This provides the flux of each pixel of each frame, which can then be remapped onto a 2D image plane. Second, we observed that the projector’s refresh rate drifted over time, causing the video frames to go slightly out of sync. Due to the raster scanning pattern of the laser projector, this drift also causes the odd and even rows to be slightly misaligned. To compensate for this, we manually searched around the detected frequency for an “average” refresh rate which we found to be 58.70893, which produced stable frames in the entire acquisition interval. Moreover, only the odd rows of the reconstructed video frames are shown in order to improve visual clarity.

To calculate the average number of photons per frame in each video clip, we divide the total number of collected timestamps by 60.

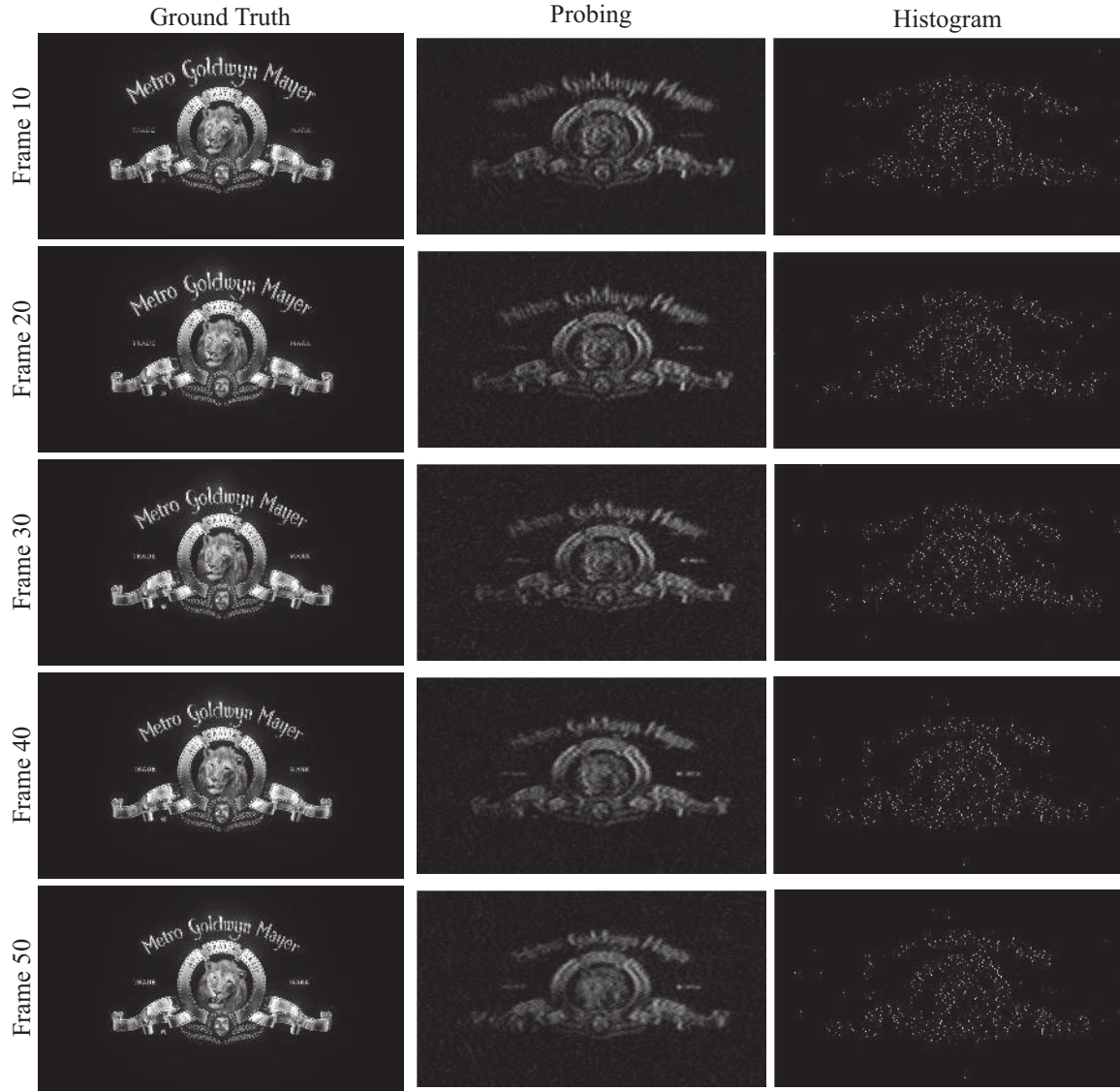


Figure 4.5.16: NLOS video reconstruction. Example reconstructed frames from probing and timestamp binning for the video sequence in $[1, 2)$ seconds.

NLOS video reconstruction results. Figures 4.5.16 compares the results of the above NLOS video reconstruction procedure to the video frames actually projected by the projector, along with the frames estimated by the timestamp binning procedure outlined in Section 4.6.3, for three different video frames. Despite the low number of detected photons, the imaging speeds involved, and the video frames' complexity, flux probing yields very good reconstruction results when compared to the ground truth. In comparison, timestamp binning, which does not reconstruct the flux function, produces poor results and makes clear that the photons detected per frame are not sufficient to obtain a high-quality image. Indeed, the contrast of frames reconstructed by timestamp binning is limited to just one photon because the time it takes for the raster scan and laser beam to move from one pixel to another is less than the SPAD's dead time. As a result, no more than one photon can ever be detected per pixel in a single video frame.

The above results point to the somewhat surprising property of our approach: even though the flux

function varies considerably from video frame to video frame in this experiment, the timestamps produced in *all* frames of the video contribute to the reconstruction quality of a given frame's flux function, not just the timestamps acquired in that particular frame.

4.6 Simulations

4.6.1 Normal distribution of probing measurements

As the number of timestamps increases, probing measurements tend to approximate a normal distribution (Proposition 4.3). This is confirmed by the simulations shown in Figure 4.6.1, where we simulate the distribution of Fourier probing measurements for various numbers of timestamps and probed frequencies, for a given flux function.

More specifically, we generate a stream of 3.5 million timestamps and thin them into separate timestamp stream realizations of either 2, 5, 10, or 50 timestamps that span the entire acquisition interval. We then use these timestamp streams to probe various frequencies, and we estimate the distribution of the resulting probing measurements. Figure 4.6.1 overlays their empirical distribution and the probability density function for a Gaussian centered at the ground-truth mean and variance (calculated by probing the ground-truth flux function). For probing measurements with at least 10 timestamps, the joint and marginal distributions of the complex Fourier probing measurements show close agreement with the normal distribution.

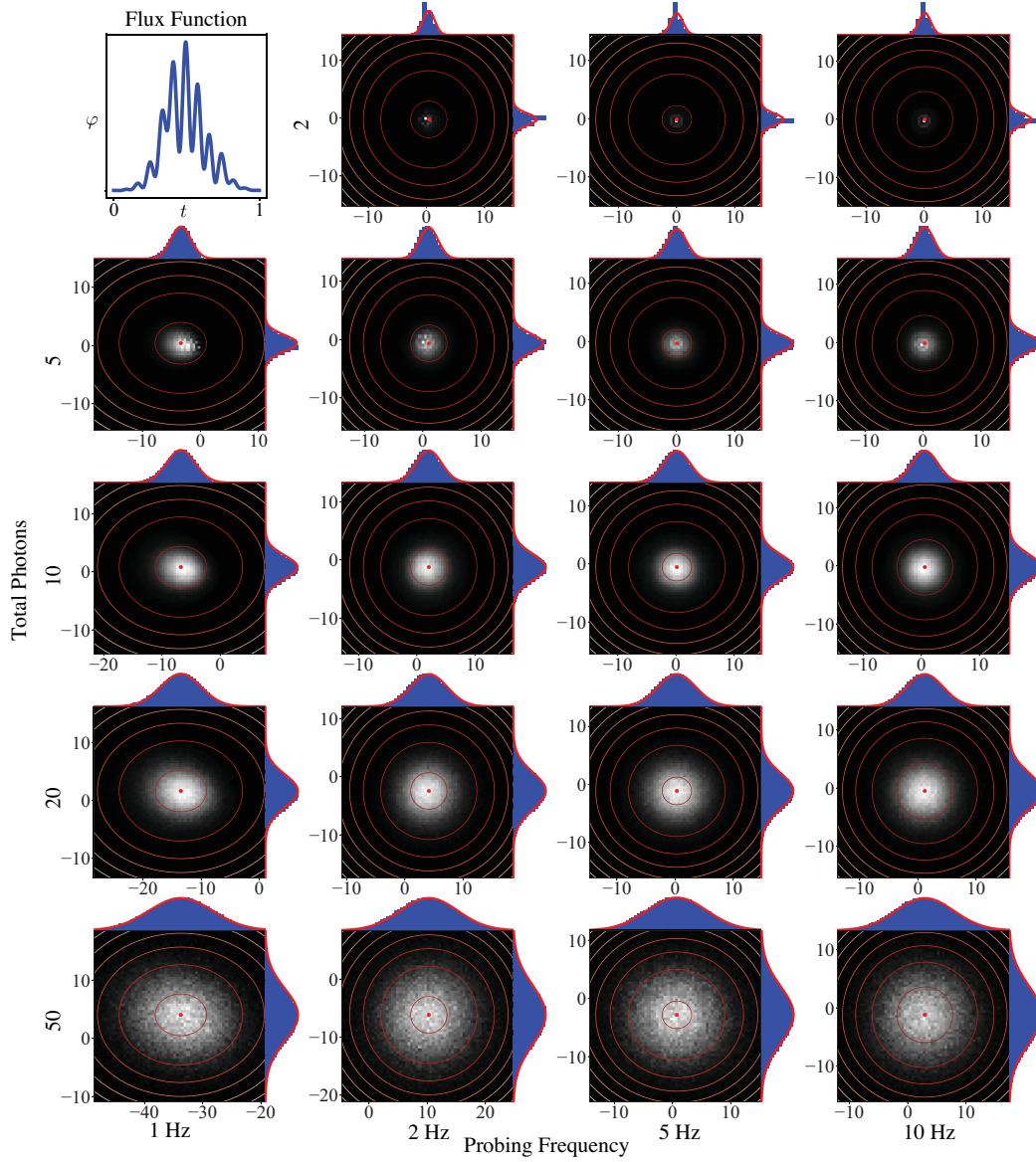


Figure 4.6.1: Distribution of probing measurements. We generate a timestamp stream according to the flux function shown in the upper left and use a thinning procedure to obtain a collection of Fourier probing measurements for various frequencies (columns) and timestamp stream realizations containing either 2, 5, 10, or 50 timestamps (rows). The empirical joint and marginal distributions of the real and imaginary measurements are shown, along with contour plots of a normal distribution whose mean and variance is calculated using the ground-truth flux function.

4.6.2 Passive reconstruction of laser pulse trains

We consider reconstruction of a flux function corresponding to 20 MHz pulse trains from an ultra-fast laser with frequency support of 12.5 GHz and 125 GHz, respectively, *i.e.*, up to the theoretically-attainable limit of a jitter-less SPAD with 4 ps timestamp quantization (Proposition 4.2). Figure 4.6.2 compares the result of three methods: (1) conventional photon-count histogramming which requires synchronization, (2) our sync-less reconstruction algorithm in Figure 4.4.1, and (3) “oracle-based” flux probing, which probes the ground-truth set of frequencies instead of relying on frequency scanning. As can be seen, probing can recover pulses up to the theoretical limit from just 50 timestamps and, despite being passive, outperforms histogramming considerably as photon counts increase.

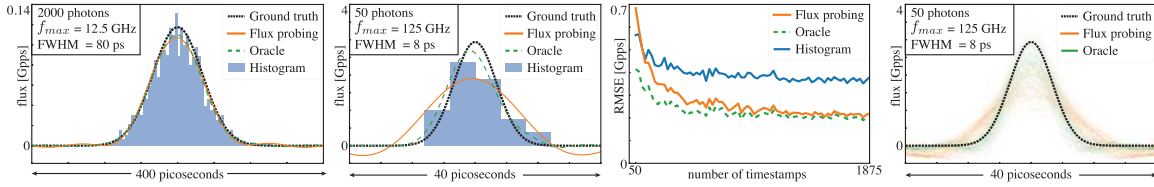


Figure 4.6.2: Simulation-based comparisons. **Left to right:** Laser pulse reconstructed by three methods from 2000 timestamps produced by a 20 MHz pulse train. Reconstruction from a 20 MHz train of much shorter pulses, using just 50 timestamps. Pulse reconstruction error as a function of the number of timestamps given as input. Reconstructed pulses from 100 realizations of a fifty-photon timestamp stream.

Model of laser pulse. We approximate the lobe of the laser pulse as a Gaussian distribution whose standard deviation is $\frac{\text{FWHM}}{2.355}$. This is approximately the standard deviation of a Gaussian with the desired FWHM. We use a repetition rate of 20 MHz, so the simulated pulse train has a period of fifty nanoseconds. We position the Gaussian pulse so that its mean is at the period’s center, *i.e.*, at $t = 25$ nanoseconds, and truncate it at $[0, 50 \text{ ns}]$ so that consecutive pulses in the train do not overlap. The frequency support of pulses with a FWHM of eight and eighty picoseconds is approximately 12.5 GHz and 125 GHz, respectively. In particular, frequencies outside these ranges have amplitude less than 3% of the peak amplitude.

Timestamp stream generation. We use Çınlar’s method [45] to generate individual photon timestamps from the pulse train. This method is preferable over thinning [117] because the narrow temporal support of the laser pulse train makes thinning computationally inefficient.

After timestamp generation, we apply a dead time of 250 ns to remove timestamps that arrive during the SPAD’s inactive period. The remaining timestamps are quantized to a resolution of four picoseconds, *i.e.*, the nominal quantization of our experimental system’s time-to-digital converter.

For the 80 ps laser pulse train simulation, we use an exposure time $t_{\text{exp}} = 10$ ms and set the average flux to be 200 kpps.

For the 8 ps pulse train simulation, we set the average flux to 500 kpps and vary t_{exp} in the range $[0.1 \text{ ms}, 3.75 \text{ ms}]$. This range of exposure times results in timestamp streams with lengths between 50 and 1875 timestamps on average. These streams were used for the comparisons shown in Figure 4.6.2 (column 3).

Estimation methods. We compare three methods for flux function estimation:

- **Flux probing:** We probe up to 12.5 GHz with a step size of 100 Hz for the 80 ps laser pulse simulations. We set the CFAR detector's probability of false alarm to be just one frequency on average out of the resulting samples, *i.e.*, $\alpha = \frac{100 \text{ Hz}}{12.5 \text{ GHz}}$. For the 8 ps laser pulse simulations, we probe up to 125 GHz with a step size of $\frac{1}{t_{\text{exp}}}$, where t_{exp} ranges from 0.1 ms to 3.75 ms. We again set the probability of false alarm so that only one false alarm occurs on average, *i.e.*, $\alpha = \frac{1}{t_{\text{exp}}} \frac{1}{125 \text{ GHz}}$.
- **Oracle-based probing:** We regularly sample the ranges $[0, 12.5 \text{ GHz}]$ and $[0, 125 \text{ GHz}]$ with a sampling rate of 20 MHz. This corresponds to the frequency support of the 80 ps and 8 ps ground-truth laser pulse trains, respectively. We then probe these sets of frequencies and use the probing results to reconstruct the laser pulses.
- **Active histogram-based imaging:** We assume the sync frequency is $f_{\text{sync}} = 20 \text{ MHz}$ and map each timestamp to the relative time interval $[0, 50 \text{ ns}]$. These relative timestamps are then binned using Matplotlib's `hist()` routine, with a bin width of 4 ps. To convert raw photon counts to a flux function expressed in photons per second, we multiply each bin's contents by the total number of pulses emitted during the exposure time as well as by the bin width, *i.e.*, we multiply by the scalar $(4 \text{ ps} \times 20 \text{ MHz} \times t_{\text{exp}})^{-1}$.

To produce the plot shown in Figure 4.6.2 (third from left), we randomly generated twenty timestamp streams for 74 exposure times regularly sampled from the range $[0.1 \text{ ms}, 3.75 \text{ ms}]$. For each method, we report the root mean squared error (RMSE) between the estimated and ground-truth flux functions. For a fairer comparison, we sample the flux function returned by each method at the center of each histogram bin.

As can be seen in that plot, oracle-based probing always outperforms both histogramming and probing. Probing, on the other hand, outperforms histogramming except when the total number of timestamps is very small (below 100). In that regime, many probed frequencies fall below the CFAR threshold and therefore are not detected. This results in a worse flux function approximation on average, despite the very coarse reconstruction returned by histogramming (*e.g.*, see Figure 4.6.2, second from left).

4.6.3 Recovering passive non-line-of-sight (NLOS) videos

Since we have no access to the actual ground truth for the non-line-of-sight (NLOS) video, we performed simulations of NLOS video reconstruction to provide quantitative results.

Model of flux function. We take 30 frames of video of the MGM lion (Figure 4.6.4). Each frame is 720 by 1280 pixels. We vectorize the video to create a 27.648M-dimensional $(720 \times 1280 \times 30)$ vector φ , which is our flux function. We treat each element of φ as the flux during a small, 36.17 ns time window that corresponds to the time allocated to each individual projector pixel. We then rescale the vector to achieve an average flux of 2 Mpps.

Timestamp stream generation. We use thinning [117] to generate a single stream of timestamps corresponding to an exposure time of one second and the flux function defined above. To simulate a lower flux level of N pps with $N < 2\,000\,000$, we use Bernoulli thinning with $p = \frac{N}{2\,000\,000}$. We

take dead time and quantization into account as described in Section 4.6.2.

Estimation methods. We compare three methods for flux function estimation:

- **Flux probing:** We probe up to 7.5 MHz with a step size of 1 Hz. We set the probability of false alarm α to be 0.3% to match the value used in the real-world experiment of Section 4.5.7. We then regularly sample the reconstructed flux function to obtain 27.648M samples spanning the one-second acquisition interval.
- **Oracle-based probing:** We computed the fast Fourier transform of the vector φ and kept the 105735 most significant frequencies according to their amplitude. This corresponds to 0.3% of total frequencies, *i.e.*, the expected number of frequencies that will pass the CFAR detection test. We probe using these frequencies and regularly sample the reconstructed flux function to obtain 27.648M samples spanning the one-second acquisition interval.
- **Timestamp binning:** Each timestamp is mapped to an element of vector φ . Because of the SPAD's dead time, at most one photon can be detected for each element of φ , *i.e.*, each pixel in the video will contain either zero or one photon. The resulting flux function is therefore an indicator function that describes whether a photon was detected or not.

Each estimated flux function is then reshaped into a $720 \times 1280 \times 30$ video and its values rescaled to the range $[0, 255]$.

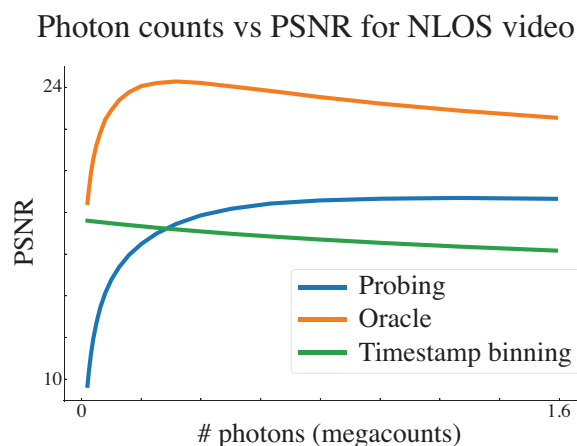


Figure 4.6.3: Plot showing number of photons vs PSNR for the three estimation methods.

Reconstructions. We show the PSNR plot for flux levels in Figure 4.6.3. In general, flux probing and oracle-based probing enable recovery of video frames with better contrast than timestamp binning, whose frames are strictly binary. The PSNR plot shows that oracle-based probing performs the best as expected. Note that while timestamp binning has higher PSNR than flux probing for flux levels below 300 kpps, Figure 4.6.4 shows that the lion remains visible in frames reconstructed by flux probing whereas this is not the case for frames reconstructed by timestamp binning.

Lastly, the performance of all three methods drops as flux levels increase to the point where dead time becomes significant. This is evident in the reconstructed frames shown in Figure 4.6.5, which have lower contrast and a saturated appearance.

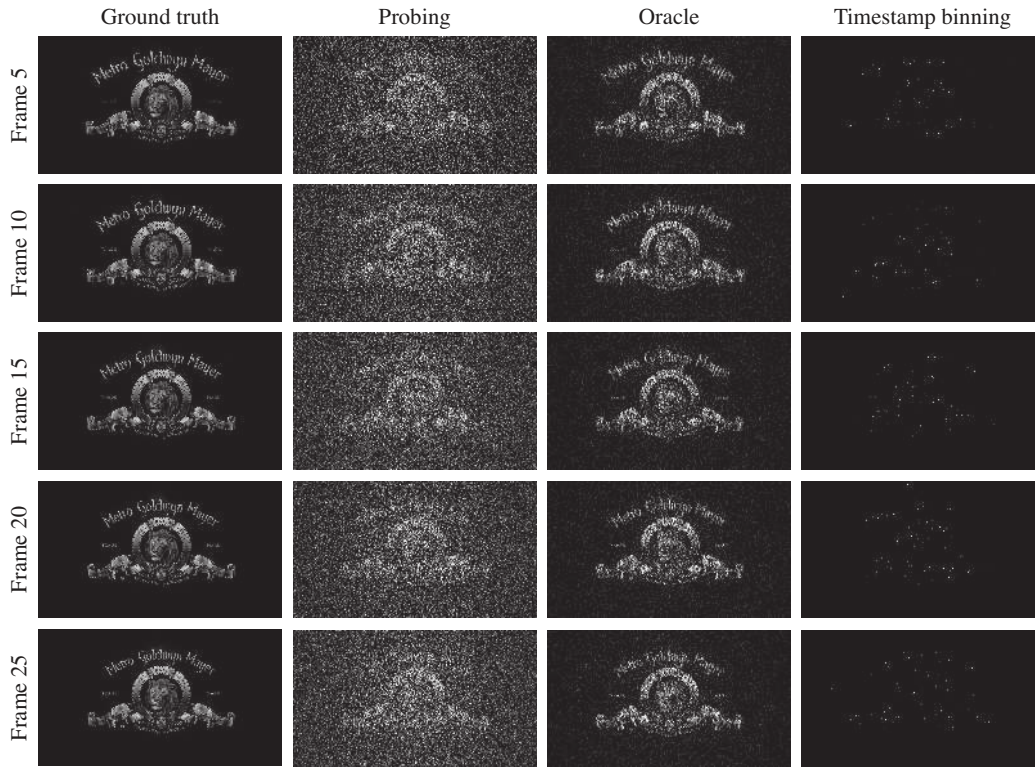


Figure 4.6.4: Simulation comparison for NLOS video reconstructions using 30k photons.

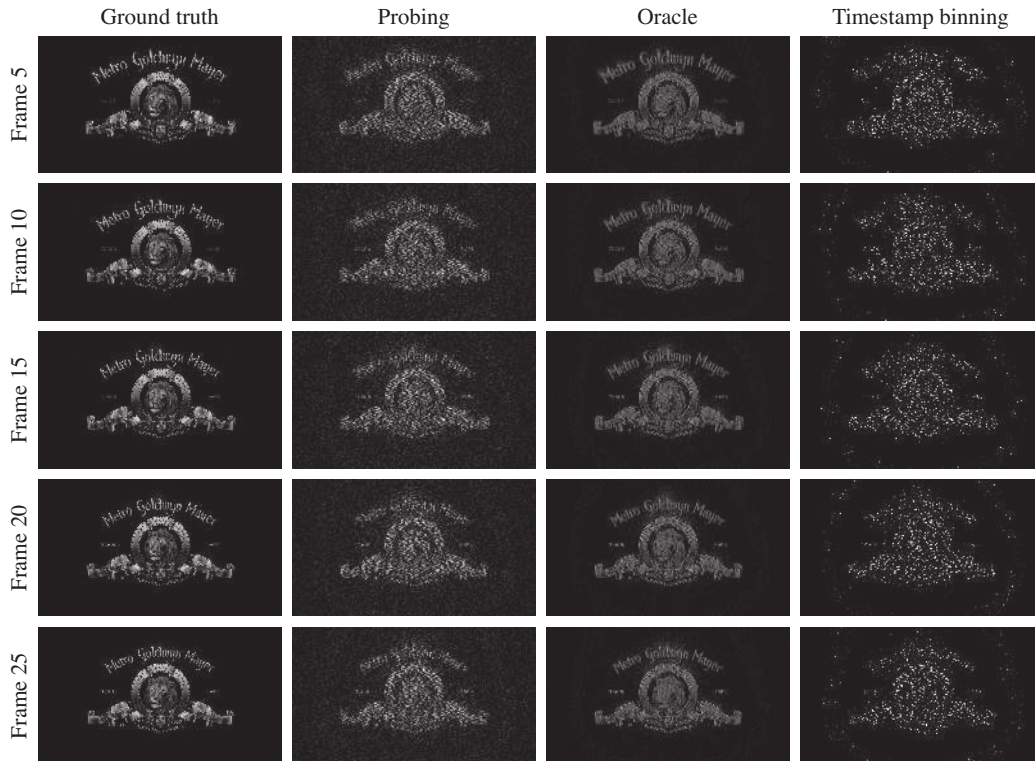


Figure 4.6.5: Simulation comparison for NLOS video reconstructions using 1.6M photons.

4.7 Limitations

The most significant limitation of the algorithm presented in UWB imaging at the moment is its poor computational efficiency. When we are purely agnostic to the frequency content of our scene, we potentially need to sweep billions of frequencies to reconstruct the flux function, of which we only keep about 2 000–10 000 frequencies. Furthermore, the current approach of summing across all timestamps and frequencies is very inefficient, with a time complexity that is quadratic in the exposure time used. Even milliseconds of data can therefore take minutes or hours to probe. In comparison, histogramming or counting photons is linear in the exposure time, and therefore, is much more commonly used in SPAD devices [10, 181], albeit without the asynchronous and passive capabilities and fidelity offered by passive UWB imaging. Improving the computational efficiency of flux probing would be a significant step to making passive UWB imaging broadly applicable and practical.

Additionally, many everyday signals cannot be recovered using current passive UWB imaging. Due to the use of Fourier basis functions to estimate the underlying flux and the dead time of the SPAD pixel, discontinuities in the flux as well as short, transient, non-periodic events cannot be recovered. In the first case, discontinuities lead to the Gibbs phenomenon, introducing ringing artefacts due to the need for infinite spectral support to represent the discontinuity. As a result, even simple scenes such as ceiling lights turning on suddenly in a room or a person getting blocked by a car may be poorly reconstructed. In the second case, short, transient, non-periodic events can be missed due to the lack of photons associated with them. For example, an event lasting 1 microsecond contributes only four photons when the dead time is 250 ns, which is a tiny signal for recovering the flux. Furthermore, even the question of how many photons we need to recover the flux is left unanswered by our theory.

While our flux probing theory can be applied generally to a broad set of basis functions, deciding on the form of the basis functions a priori, whether it is Fourier, wavelet, or polynomial, assumes that the underlying flux lies within the span of our chosen basis functions. This may not always be the case, so it would be useful to consider extending our framework, perhaps by leveraging modern machine learning techniques [231, 249, 115], to understand how the photon stream can be used to guide the choice of basis functions or even to eschew the need for basis functions by learning the flux function directly from the photon stream.

From the practical side, since we only had access to a single-pixel SPAD, our 2D experiments needed to be captured by raster scanning the scene. Due to raster scanning taking hours to capture data, we were limited to scenes with repeatable motion, so we could capture a cohesive video. As 2D SPAD arrays start emerging with higher resolution [152, 252], it would be great to capture ultra-wideband videos for more general scenes without periodic motion.

4.8 Concluding Remarks

The sheer amount of data involved in probing timestamp streams cannot be ignored: even a single pixel can output tens of thousands of timestamps per second in low light, and our ultrawidebandwidth results require probing billions of frequencies. Sketching [221] and Non-Uniform FFT [21]

may offer ways forward but major challenges remain. That said, we believe that passive acquisition and processing of timestamp streams from free-running SPADs opens new directions in dynamic imaging: completely unsynchronized, single-shot observations of ultrafast phenomena with multiple light sources across different timescales; passive depth imaging using uncooperative, environmental light sources; compressive ultrafast video recording from sparse timestamps; temporal “microscopes” that allow monitoring intensity fluctuations across timescales spanning the nine-plus orders of magnitude (*i.e.*, DC to 31 GHz) theoretically captured by SPADs; and more. We are thus looking forward to more advances on these remarkable sensors.

5

Flux Probing Theory for Non-Negligible Dead Time

5.1 Introduction

In this chapter, we generalize our flux probing theory to handle dead time. In the preceding chapter, we did not emphasize non-negligible dead time for the sake of simplicity since most high-level ideas, illustrations and algorithms can be communicated via the easier-to-grasp setting of negligible dead time and independent detections. Here, we consider how dead time induces probabilistic dependencies in the timestamp stream and how we can incorporate concepts from stochastic calculus to alter the flux probing operation to account for these dependencies. Additionally, we expand upon concepts that were briefly touched upon in the preceding chapter such as the counting process.

In order to describe the theory rigorously, we begin by introducing a few preliminary mathematical definitions from [46] that will be helpful for some of our core results.

5.2 Preliminary definitions

Definition 5.1 (σ -algebra (Definition 1.1.1 in [46]))

Given a space Ω , a σ -algebra \mathcal{F} is a collection of subsets of Ω that contains the empty set \emptyset , is closed under complement and countable intersections and, by extension, countable unions.

Σ -algebras describe the event space, wherein an event is a set of outcomes we may be interested in. For example, "ten photons arrived in one microsecond" as well as "no photon arrived during the time interval $[s, t]$ " are both examples of events. More formally, Ω is the set of all possible photon arrivals and each element of \mathcal{F} is a set in Ω . In many cases, we are only interested in events up to a certain time t . In this case, we can model this using filtrations:

Definition 5.2 (Filtration (Definition 3.1.2 in [46]))

A **filtration** $\mathbf{F} = (\mathbf{F}_t, t \geq 0)$ is a family of σ -algebras \mathbf{F}_t which is increasing, i.e., $\forall s < t, \mathbf{F}_s \subseteq \mathbf{F}_t$.

A filtration \mathbf{F} restricts the event space to events regarding timestamp sequences up to any given time t . By increasing t , we increase the event space and can then model the time evolution of photon arrivals.

Definition 5.3 (Adapted Process (Definition 3.2.14 in [46]))

A stochastic process X is **\mathbf{F} -adapted** if $\forall t \geq 0$, the random variable $X(t)$ is \mathbf{F}_t -measurable.

Intuitively, F_t -measurable means that the value of $X(t)$ is determined entirely by the information available at time t , *i.e.*, the current value of $X(t)$ does not depend on the future. In our case, since the number of photon timestamps collected up to time t is known, the counting process $N(t)$ is F -adapted with F_t being the event space up to time t .

Definition 5.4 (F-Martingale (Definition 5.1.1 in [46]))

An F -adapted process $M = (M(t), t \geq 0)$ is an **F-martingale** if

1. the expectation of the absolute value of $M(t)$, denoted as $E[|M(t)|]$, is finite for all $t \geq 0$ and
2. the conditional expectation of $M(t)$ given the σ -algebra F_s , $E[M(t)|F_s]$, equals $M(s)$ almost surely for every pair (s, t) such that $s < t$.

Intuitively, a martingale can be thought of as the stochastic process equivalent of zero-mean additive noise: given the value of the martingale at time s , its expected value at any later time $t > s$, is its value at time s .

5.3 Photon statistics

Consider a stream of photons arriving at the sensor where each photon arrives at some time τ_i . The set $\mathcal{T} = \{\tau_i\}$ defines a stochastic process. Formally speaking, a stochastic process is just a family of random variables over an index set (usually time); in this case, each photon arrival time is a random variable. Strictly speaking, we need a way to assign probabilities to different realization of photon arrivals. For example, given a bright sunny scene, it is much less likely for a sensor to observe no photons as opposed to many. After all, cameras would not be very useful in everyday scenes if this were not the case. Fortunately, it turns out that the arrival of photons is then governed by an *inhomogeneous Poisson process* with a rate parameter equal to the photon flux $\varphi(t)$.¹ Given this, we can determine the statistical properties of the photon count $N(s, t)$ in any interval $[s, t]$:

Definition 5.5 (Inhomogeneous Poisson Process [232])

Given an inhomogeneous Poisson process with photon flux function $\varphi(t)$, $N(s, t)$ is a random variable with the following properties:

1. $P(N(s, s) = 0) = 1$.
2. $N(s, t) \sim \text{Poisson}\left(\int_s^t \varphi(u) du\right)$.
3. N has independent increments so $P(N(s, t) = k | N(r, s)) = P(N(s, t) = k)$.

For convenience, we also define the *photon flux integral* $\Phi(s, t)$ [photons]:

¹Sub-Poissonian and super-Poissonian light sources do exist, but they are seldom modelled for cameras. Thermal radiation is an instance of super-Poissonian light. $\lambda(t)$ is more commonly used in the statistics literature, but we reserve λ to denote spectral wavelengths.

Definition 5.6 (Photon Flux Integral)

$$\Phi(s, t) = \int_s^t \varphi(u) du.$$

The fact that photons form an inhomogeneous Poisson process allows us to derive many useful properties for photon arrival statistics. For one, the photon count N follows a Poisson distribution, and we can characterize both its mean and variance in terms of the photon flux integral Φ :

$$E[N(s, t)] = \text{var}[N(s, t)] = \Phi(s, t). \quad (5.3.1)$$

Additionally, due to the independent increments of Poisson processes, we can now compute the conditional probability density function of a j^{th} photon arriving at time τ_j given previous photon arrivals $\tau_1, \dots, \tau_{j-1}$ [232]:

$$\text{Prob}(\tau_j = t \mid \tau_1, \dots, \tau_{j-1}) = \varphi(t) e^{-\int_{\tau_{j-1}}^t \varphi(u) du}. \quad (5.3.2)$$

5.4 Mathematical Treatment of the Counting Process

While it is convenient to represent the stream of photon arrival times as a discrete sequence of random variables, another equivalent representation is to consider a continuous-time process $N(t)$, formally called a *counting process* that counts the photons received up to time t and is completely determined by \mathcal{T} :

Definition 5.7 (Counting Process [232])

The counting process $N(t)$ of a point process X is a nonnegative integer-valued stochastic process with the property:

$$N(t) = \sum_{x \in X} I_{x \leq t}, \quad (5.4.1)$$

where $I_{x \leq y}$ is the characteristic function:

$$I_{x \leq y} = \begin{cases} 1 & x \leq y \\ 0 & \text{otherwise} \end{cases} \quad (5.4.2)$$

We can observe that there is one-to-one mapping between a sequence of photon arrival times and the counting process. It turns out that one key advantage of the counting process is that its relation to the photon flux can be made explicit through the following decomposition:

Proposition 5.1

Let $N(t)$ be the counting process of an inhomogeneous Poisson process with flux function $\varphi(t)$

and let \mathbf{F} be a filtration such that $N(t)$ is \mathbf{F} -adapted. The process

$$M(t) = N(t) - \int_0^t \varphi(u) du, \quad t \geq 0 \quad (5.4.3)$$

is an \mathbf{F} -martingale.

Proposition 5.1 tells us that the counting process approximates the integral of the photon flux up to a continuous-time random process called a *martingale*. One important consequence of Proposition 5.1 is that we can use the \mathbf{F} -martingale $M(t)$ as a stochastic integrator of a bounded deterministic function p to yield another \mathbf{F} -martingale:

Definition 5.8 (Definition 8.3.2.1(i) in [98])

Let $M(t)$ be an \mathbf{F} -martingale. For any bounded deterministic function p , the process $M_p(t)$ given by

$$M_p(t) = - \int_0^t p(u) dM(u) \quad (5.4.4)$$

is a \mathbf{F} -martingale.

Combining Proposition 5.1 and Definition 5.8 gives us our main result of the flux probing theory: the flux probing equation, that we restate for convenience:

Proposition 4.1 (Flux Probing Equation)

The inner product of the probing function $p(t)$ and the unknown flux function $\varphi(t)$ over the time interval $[0, t]$ satisfies the relation

$$p(\mathcal{T}) = \langle p, \varphi \rangle + M_p(t),$$

where $p(\mathcal{T})$ are “probing measurements” which sum the values of the probing function at the absolute timestamps

$$p(\mathcal{T}) \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}} p(\tau) ,$$

$M_p(t)$ is a martingale, and the inner product is defined as $\int_0^t p(u) \varphi(u) du$.

5.5 Flux Probing under Non-Negligible Dead Time

When dead time is not negligible, photon detections are no longer Poisson [194] but Eqs. (4.3.1) and (4.3.2) still apply because they do not rely on the independence of photon detections. In that case, the timestamp stream is described by a more general point process; the random process $N(t)$ in Eq. (4.3.1) is still a counting process and the noise is still a martingale; Eq. (4.3.2) still holds but the inner product is over the known timespan that the SPAD is active, not over the full interval $[0, t_{\text{exp}}]$; and Algorithm 4.1 can be readily modified to convert inner products over the SPAD’s active time to inner products over $[0, t_{\text{exp}}]$.

5.5.1 Generalized flux probing

We introduce the generalization of our flux probing equation to the case of non-negligible dead time. While the photon detections no longer form a Poisson process when the dead time is not negligible, we can still define its corresponding counting process. We analyze the statistical properties of the resulting counting process through a *marked point process*. Suppose that in addition to the set of photon arrival timestamps \mathcal{T} themselves, each timestamp τ_n is assigned a binary mark m_n , representing the two possible outcomes, *i.e.*, detected (1) and not-detected (0). The joint sequence (τ_n, m_n) is called a *marked point process* [32] and specifically a 2-variate point process. For now, we remain agnostic to how marks are assigned and let $N^{(1)}(t)$ be the counting process of all timestamps with $m_n = 1$. We can see that $N^{(1)}$ corresponds to the counting process we compute from the stream of photon timestamp detections. We define the history $\mathbf{F}_{\tau_{j-1}}$ to be the sequence $(\tau_n, m_n)_{n < j}$. Now, we can define some basic properties of the marked point process starting with the conditional probability of the j^{th} timestamp arriving at time t given the history $\mathbf{F}_{\tau_{j-1}}$:

$$P_j(\tau_j = t | \mathbf{F}_{\tau_{j-1}}) = \varphi(t) e^{-\int_{\tau_{j-1}}^t \varphi(u) du}. \quad (5.5.1)$$

Equation (5.5.1) follows from the Poisson arrival of τ_j [232]. Furthermore, the conditional probability P_j of the j^{th} timestamp arriving at time t **and** being assigned mark $m_j = 1$ is

$$P_j(\tau_j = t, m_j = 1 | \mathbf{F}_{\tau_{j-1}}) = \mathbf{1}_{t > \tau_i + D} \varphi(t) e^{-\int_{\tau_{j-1}}^t \varphi(u) du}, \quad (5.5.2)$$

where τ_i is the last timestamp before τ_j for which $m_i = 1$, and $\mathbf{1}_{t > \tau_i + D}$ is the indicator function for t being outside the dead time D (which occurs when $t > \tau_i + D$). Equation (5.5.2) follows from Lemma B.8 in Appendix B.3.

Let us define the following stochastic process $\Phi(t)$, we call $\Phi(t)$ the *flux process*:

$$\Phi(t) = \mathbf{1}_{t > \tau_i + D} \varphi(t). \quad (5.5.3)$$

We can think of the flux process as being equal to the flux function when the SPAD is active and 0 otherwise. Intuitively, we can think of the photon detections as coming from a Poisson process parametrized by a stochastic flux process.² Now, let us define Ω_{active} to be the set of active SPAD times in the interval $[0, t_{\text{exp}}]$. Then

$$\int_0^{t_{\text{exp}}} \Phi(u) du = \int_0^{t_{\text{exp}}} \mathbf{1}_{t > \tau_i + D} \varphi(t) du = \int_{\Omega_{\text{active}}} \varphi(t) du. \quad (5.5.4)$$

We can now incorporate dead time D into our probing equation. Our key result is as follows:

Proposition 5.2 (Generalized Flux Probing Equation)

The inner product of $p(t)$ and the unknown flux function $\varphi(t)$ over the time interval $[0, t_{\text{exp}}]$ can be expressed as:

$$\langle p, \varphi \rangle_{\text{active}} = p(\mathcal{T}) + M(t_{\text{exp}}), \quad (5.5.5)$$

²The approach taken here can be thought of as a generalization of [194] to the case of flux probing.

where $\langle p, \varphi \rangle_{\text{active}} = \int_{\Omega_{\text{active}}} p(u) \varphi(u) du$ and Ω_{active} is the subset of $[0, t_{\text{exp}}]$ where the SPAD is active.

5.5.2 Generalized algorithm

In the case when T_{avg} is comparable to dead time, the probing operation requires solving a complex least squares problem. Use of complex least squares is necessary because when dead time is not negligible, the union of time intervals when the SPAD is active is not equal to $[0, t_{\text{exp}}]$, and thus the Fourier basis is not orthogonal over that union. We refer to this method as *Flux probing with non-negligible dead time* (Algorithm 5.1). Note that this algorithm employs a frequency-scanning step size equal to $1/t_{\text{exp}}$ instead of $0.6/t_{\text{exp}}$ used in Algorithm 4.1. This ensures that the resulting system of linear equations in the dead-time compensation step is invertible.

The main idea of Algorithm 5.1 is to identify frequencies through CFAR detection, followed by the solution of a complex linear system of equations to obtain the amplitudes and phases compensated for the SPAD's active period. The formulation of this linear system is discussed in Section 5.5.3.

Algorithm 5.1: FLUXRECDEADTIME—flux probing with non-negligible dead time

Input: $\mathcal{T}, t_{\text{exp}}, f_{\text{max}}, \alpha, D$
Output: $\hat{\varphi}(t)$
Frequency scanning.
 $\Delta f \leftarrow 1/t_{\text{exp}}$
 $\mathcal{F} \leftarrow$ frequencies from Δf to f_{max} with step Δf
for $f \in \mathcal{F}$:
 $p_f(\mathcal{T}) \leftarrow \frac{1}{t_{\text{exp}}} \sum_{\tau \in \mathcal{T}} e^{-j2\pi f \tau}$
Frequency detection.
 $\mathcal{F}_{\text{used}} \leftarrow \emptyset$
 $A_0 \leftarrow N(\mathcal{T})/t$
for $f \in \mathcal{F}$:
 $A_f \leftarrow 2|p_f(\mathcal{T})|, \phi_f \leftarrow \angle p_f(\mathcal{T})$
 Reject f using CFAR (Eq.(4.3.7))
 if f is not rejected:
 $\mathcal{F}_{\text{used}} \leftarrow \mathcal{F}_{\text{used}} \cup \{f\}$
Dead-time compensation.
Solve Eq. (5.5.9) to obtain C_f
 $A_f \leftarrow |C_f|, \phi_f \leftarrow \angle C_f$
Flux reconstruction.
 $\hat{\varphi}(t) \leftarrow \sum_{f \in \mathcal{F}_{\text{used}}} A_f \cos(2\pi f t + \phi_f)$
return $\hat{\varphi}(t)$

5.5.3 Dead-time compensation step of Algorithm 5.1

In this section we assume that the flux function is a linear combination of Fourier basis functions:

$$\varphi(t) = \sum_{f \in \mathcal{F}_{\text{used}}} A_f \cos(2\pi f t + \phi_f) = \sum_{f \in \mathcal{F}_{\text{used}} \cup -\mathcal{F}_{\text{used}}} C_f p_f(t), \quad (5.5.6)$$

where we denote the complex amplitude of the Fourier basis as C_f and use $-\mathcal{F}_{\text{used}}$ to denote the frequencies in $\mathcal{F}_{\text{used}}$ multiplied by -1 .

The presence of dead time distorts our probing measurements, *i.e.*, $\langle p, \varphi \rangle_{\Omega_{\text{active}}} \neq \langle p, \varphi \rangle$ in general. However, it is still possible to recover the amplitudes $\langle p, \varphi \rangle$. Namely, we have:

$$\langle p_l, p_k \rangle_{\Omega_{\text{active}}} = \sum_{i=1}^{N(t_{\text{exp}})} \int_{\tau_{i-1}+D}^{\tau_i} p_l(u) p_k(u) du, \quad (5.5.7)$$

where $\tau_0 = -D$. We can compute $\langle p_l, p_k \rangle_{\Omega_{\text{active}}}$ either analytically or numerically. With some abuse of notation, substituting Eq. (5.5.7) into Eq. (5.5.5) gives us:

$$p_l(\mathcal{T}) = \sum_{k=1}^K C_k \langle p_l, p_k \rangle_{\Omega_{\text{active}}} - \underbrace{M(t_{\text{exp}})}_{=-e_l} \quad (5.5.8a)$$

$$= \begin{bmatrix} \langle p_l, p_1 \rangle_{\Omega_{\text{active}}} \\ \langle p_l, p_2 \rangle_{\Omega_{\text{active}}} \\ \vdots \\ \langle p_l, p_K \rangle_{\Omega_{\text{active}}} \end{bmatrix}^T \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_K \end{bmatrix} + e_l, \quad (5.5.8b)$$

where K is the cardinality of $\mathcal{F}_{\text{used}} \cup -\mathcal{F}_{\text{used}}$. Combining the K equations associated with each probing function p_l yields a linear system of equations for recovering the complex Fourier amplitudes C_k :

$$\begin{bmatrix} p_1(\mathcal{T}) \\ p_2(\mathcal{T}) \\ \vdots \\ p_K(\mathcal{T}) \end{bmatrix} = \begin{bmatrix} \langle p_1, p_1 \rangle_{\Omega_{\text{active}}} & \langle p_1, p_2 \rangle_{\Omega_{\text{active}}} & \cdots & \langle p_1, p_K \rangle_{\Omega_{\text{active}}} \\ \langle p_2, p_1 \rangle_{\Omega_{\text{active}}} & \langle p_2, p_2 \rangle_{\Omega_{\text{active}}} & \cdots & \langle p_2, p_K \rangle_{\Omega_{\text{active}}} \\ \vdots & \vdots & \ddots & \vdots \\ \langle p_K, p_1 \rangle_{\Omega_{\text{active}}} & \langle p_K, p_2 \rangle_{\Omega_{\text{active}}} & \cdots & \langle p_K, p_K \rangle_{\Omega_{\text{active}}} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_K \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_K \end{bmatrix}. \quad (5.5.9)$$

Therefore, we can solve for these amplitudes by solving the above least squares problem in the complex domain.

5.6 Experiments

5.6.1 Probing with SPAD arrays and non-negligible dead time

We apply our method to data from off-the-shelf 2D SPAD sensors. To demonstrate this, we compare against Seets et al. [218] who recover high-speed video with a 32×32 SPAD array. Due to its architecture, the sensor specifically outputs just one timestamp per 20 microseconds for each pixel,

resulting in a dead time too long to ignore even at relatively low flux levels. Therefore, we use Algorithm 5.1 in order to recover the flux at each pixel.

Scene. We applied our approach to the dataset in [218] that was graciously provided by its authors. The maximum number of timestamps acquired per pixel is 100k.

Frequency scanning and detection. We sweep frequencies from 0 Hz to 100 kHz with a step size of 5 Hz. As usual, we set the probability α of false alarm to $\frac{1}{\text{\#freqs}}$.

Flux probing with non-negligible dead time. After detecting the frequencies, we recover the amplitude and phase of each Fourier coefficient using Algorithm 5.1.

Flux reconstruction results. We use flux probing with non-negligible dead time to recover the flux function at each pixel of the SPAD array. Figure 5.6.1 shows the recovered video frames from [218] (left) and our method (right) as well as the reconstruction flux for one pixel in the video frame. Seets et al [218] assume flux is piecewise constant and identify contiguous sets of timestamps with the same flux using change-point detection algorithms. Photon inter-arrivals are then averaged to obtain a *single* flux estimate per set. In contrast, we recover a time-varying flux function by probing, yields a periodic flux function truer to the rotating fan’s motion. In contrast, periodicity and high-frequency variations are lost by the method in [218]. Additionally, despite our processing using strictly pixel-wise reconstruction of flux functions, our method recovers sharper videos of the scene. Since our flux functions are continuous, we can also render video at higher frame rates as well. When playing back videos at different frame rates, we observe that the motion of the fan from [218] appears discontinuous, likely due to the piecewise constancy assumption, whereas in our video the motion is smooth, and the fan appears sharp as well.

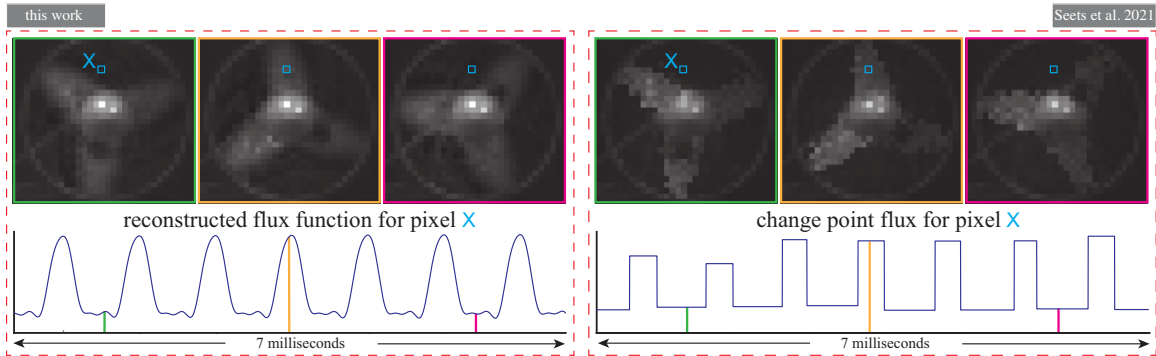


Figure 5.6.1: We use SPAD array data from Seets et al. [218] to reconstruct per-pixel flux (left). The common assumption of piecewise constant flux (right) does not hold even for this simple scene of a rotating fan.

Figure 5.6.2 shows the difference in the recovered frame when the dead time is not correctly compensated. We notice that by correctly compensating for dead time, we alter the magnitude of the flux function significantly, increasing the dynamic range of the recovered video frames. This experiment validates that our flux probing theory can correctly account for dead time and that we can apply our method broadly across different sensor data.

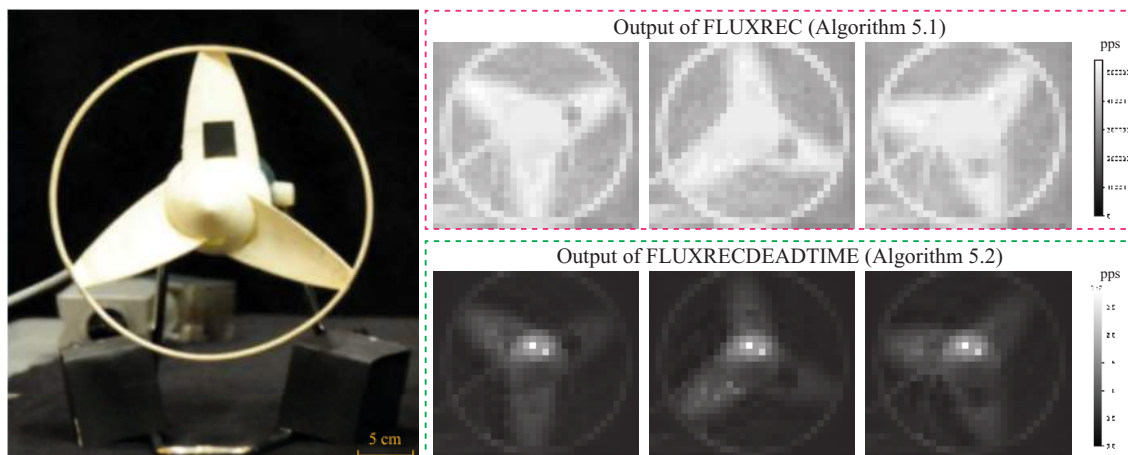


Figure 5.6.2: Probing with SPAD arrays. We apply our approach to the dataset from Seets et al. [218], by probing under the assumption of negligible dead time (top right) and non-negligible dead time (bottom right).

5.7 Simulations

Here we show some simulations that validate the generalization of flux probing to non-negligible dead time.

5.7.1 Flux probing under non-negligible dead time

To validate our theory, we perform simulations in the regime where the average timestamp detections T_{avg} is comparable to dead time.

Flux function. We define the flux function as DC plus a linear combination of N sinusoidal functions (N varies from 1 to 50) with frequencies uniformly sampled without replacement in the interval [5 kHz, 100 kHz] in steps of 100 Hz. We vary the DC, *i.e.*, average photons per second from 100 kpps to 100 Mpps. The amplitude of each sinusoid is set to the average photons per second divided by N .

Timestamp stream generation. We scale the flux function for a desired photon timestamp count and simulate timestamps using the scaled flux function and thinning [117]. We use an exposure time of 10 ms. We take dead time and quantization into account as described in Section 4.6.2.

Estimation methods. We compare two methods for flux function estimation, Algorithms 4.1 in Section 4.4 and 5.1 in Section 5.5.2.

Results. Figure 5.7.1 depicts the flux functions reconstructed using flux probing and flux probing with non-negligible dead time for various photon levels and number of frequency components. We observe that although the shape of the flux function reconstructed by both methods is similar, the flux function’s amplitude is underestimated when ignoring dead time. This phenomenon has a significant impact in the dynamic range of recovered flux function as can be seen in the real-world experiment of Figure 5.6.2. Finally, we highlight the fact that it is possible to reconstruct flux functions in settings where only 4% of the photons are detected.

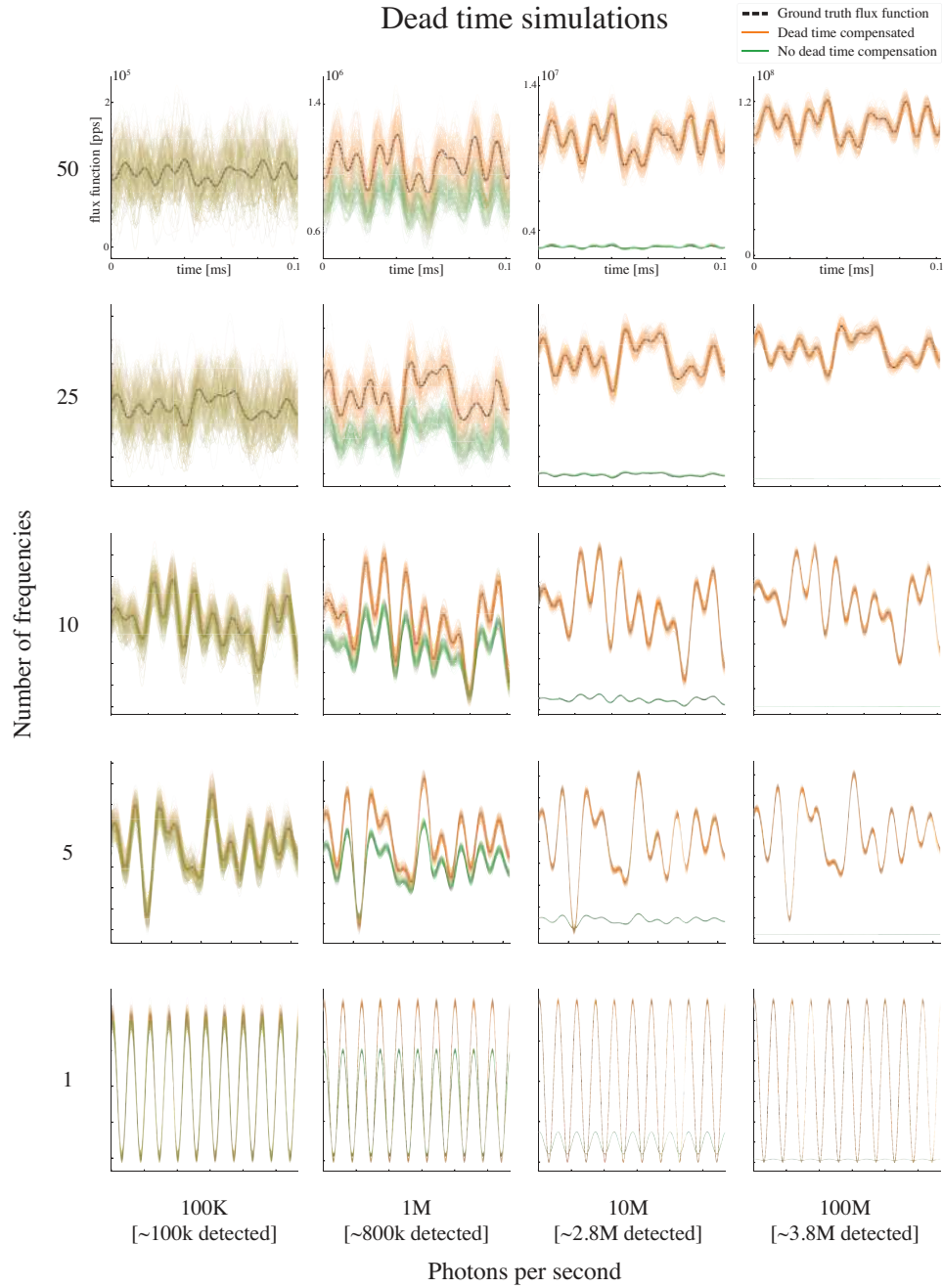


Figure 5.7.1: Our theory generalizes to the case where T_{avg} is comparable to dead time. We simulate various flux functions and demonstrate that flux probing with non-negligible dead time allows us to accurately recover the flux function even in the case when 96% of the photons are not detected (last column).

5.8 Limitations

Even though our generalized flux probing theory allows us to compensate for dead time in the photon arrivals when computing probing measurements, doing so requires solving a linear system of equations involving an N -by- N matrix. As mentioned in the preceding chapter, we potentially probe billions of frequencies, so N is on the order of billions. This means that the N -by- N matrix contains around a quintillion (10^{18}) elements.³ Putting aside the issue of computing and storing the matrix, even solving the linear system of equation is intractable, for example, Strassen’s algorithm used for matrix multiplication and inversion has a computational complexity of $\mathcal{O}(n^{2.807})$ [236]. Therefore, compensating for dead time is something that, although we know how to do it in theory, still remains future work that aims to make our generalized flux probing theory practical and applicable in a broad range of scenes.

5.9 Conclusion

Our generalized flux probing theory highlights both the utility of stochastic calculus in the analysis of photon timestamp streams as well as the need for accurate modelling of photon detections. So far, we have barely scratched the surface of things we can do with stochastic calculus, as we have been focused on just extracting flux from the photon timestamp stream. However, other more advanced techniques and results in stochastic calculus may enable new ways to process photon timestamp streams and allow for additional information to be inferred directly from the timestamp stream beyond just flux information.

³A back-of-the-envelope calculation with double precision, *i.e.* each element is 8 bytes, gives us a required memory capacity of 8 exabytes.

6 Opportunistic Single-Photon Time of Flight

6.1 Introduction

Our environments are increasingly illuminated by faint but extremely fast light signals. Devices such as lidars on robots, cars, and drones [203, 120, 124], proximity sensors and time-of-flight (ToF) cameras in smartphones [134, 239, 36], along with wireless light-based communication systems [243, 178, 146], all emit light fluctuating at frequencies from MHz to GHz. This light is generally considered useful only to the emitting device, due to the precise synchronization and calibration required to detect it. ToF cameras, for instance, use a dedicated light source that emits periodic signals in sync with their sensor, enabling accurate measurement of the time delay between emission and detection [114, 17]. To nearby cameras, however, these signals blend into the ambient background, contributing to noise or, worse, interfering with their own measurements [155, 76].

In this work we ask: can such fast “ambient” light signals be automatically detected and leveraged for passive 3D vision? As a first step toward answering this question, we show that this is indeed possible to do with a time-resolving single-photon camera, as long as some of the ambient light comes from *pulsed lasers*—a light source technology used in smartphone flash lidars [188, 134] and many advanced beam-scanning lidar systems [187].

Our focus is on scenarios where several independent and unknown pulsed lasers are present in a scene, each acting as a point light source. This occurs, for example, when many flash lidar devices are active, or when beams from different lasers illuminate distinct points in an environment. In such settings, each laser can be thought of as an ultra-fast strobe light, periodically flooding the scene with nanosecond- or sub-nanosecond pulses a few million times a second [171].

To passively harvest this light, we develop an incident flux model suited for cases where an unknown number of lasers emit pulses asynchronously from beyond the camera’s line of sight, possibly amid other sources of strong ambient light. We then combine this model with passive ultra-wideband imaging from Chapter 4 to automatically detect the presence of these lasers; characterize their pulse emissions; separate their image contributions; and—if two or more lasers are detected—localize them in 3D while simultaneously recovering a depth map of the camera’s field of view. The only requirements are that (1) the single-photon camera can passively timestamp photon arrivals concurrently and asynchronously across all its pixels, and (2) both the camera and the lasers remain stationary for the (sub-second) exposure duration. The resulting approach is *entirely passive*: it requires no light emission of its own, relying instead on detecting and repurposing light from other lasers operating nearby. We call this novel form of 3D imaging *opportunistic single-photon ToF*.

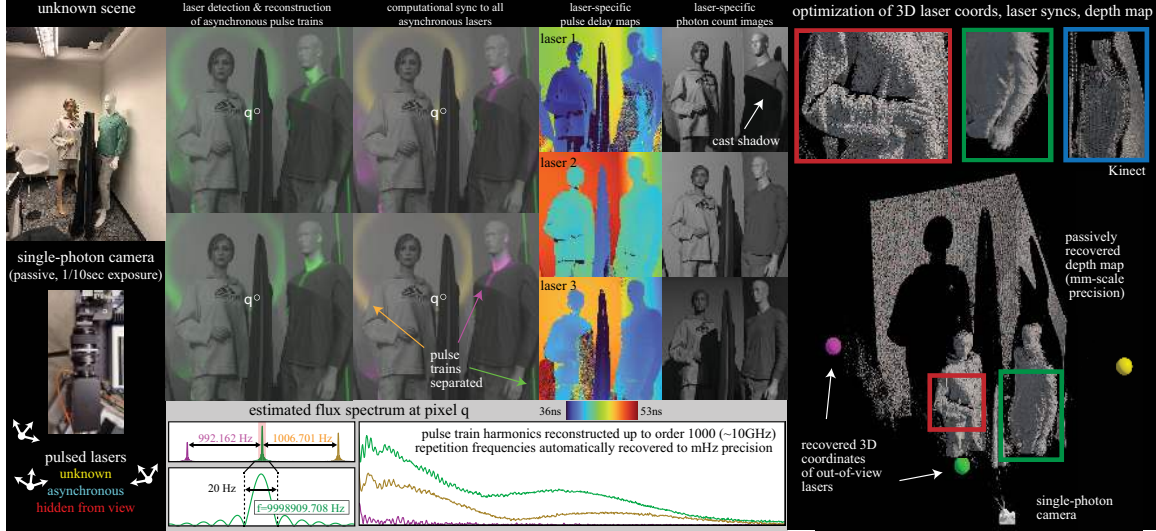


Figure 6.1.1: Harvesting scattered photons from unknown ultrafast sources for 3D sensing. **Left:** In this real experiment, three diffused lasers act as independent, unknown flash lidar devices, concurrently and asynchronously flood-illuminating a room with trains of picosecond-scale flashes from different positions. A SPAD camera, with no prior information about the scene’s light sources and no direct view of them, passively time-stamps photons with its internal clock. **Middle:** By collecting approximately 10K timestamps per pixel, the individual lasers are automatically detected; their individual pulse repetition frequencies are resolved to mHz-scale precision; their pulse spectra reconstructed to GHz harmonics; their asynchronous pulse trains separated; and the relative delay of pulses from different lasers to each pixel computed to sub-nanosecond precision. **Right:** Using this information, the lasers are localized in 3D and the scene’s depth map computed to mm-scale precision. See Section 6.7.7 for results in an even more challenging setting with high levels of ambient light ($\sim 5K$ laser photons per pixel, $\sim 170K$ ambient photons per pixel, ~ 0.01 signal-to-background ratio, SPAD dead-time limited regime).

Opportunistic harvesting of various signals has been practiced since at least the 1930s, when reflections from ambient BBC radio signals were used to detect incoming aircraft [12, 113]. Since then, serendipitous ambient signals have enabled passive imaging across diverse domains, including coherent optical [110, 42], acoustic [286, 212, 244, 158], sonar [38, 266, 37, 109], radar [73, 74, 75, 245, 144], Wi-Fi [5, 4, 6, 7] and microwave [287, 277, 138, 278]. Our work is motivated by similar principles but applied to the domain of incoherent optical ToF. In computer vision, high-speed cameras and motion enhancement techniques have been used to reveal hidden motions in the world [268, 255, 256], sense sound [49], and analyze mechanical vibrations [223, 48]. The visual signals we exploit here are considerably faster and much dimmer.

Our approach tackles this challenge by formulating and solving two independent subproblems: a *single-photon imaging* task of converting raw photon detections into laser-specific pulse-delay maps relative to the camera’s internal clock, and a purely *geometric* task of computing a depth map, 3D laser coordinates, and camera-to-laser clock offsets that are consistent with those pulse-delay maps.

On the geometric side, our work relates to techniques for radio wave device localization (*e.g.*, GPS [271, 15], wildlife tracking [13, 16, 141], cellphone localization [143, 195, 237, 9], indoor positioning [112, 128, 222, 54, 8]) which extract ToF information from ambient signals emitted by radio beacons. These techniques rely on a known beacon infrastructure, precise synchronization between beacons, and line-of-sight communication with them. In contrast, the laser “beacons” in our work are unknown, asynchronous, out of sight, and can change their position independently from one

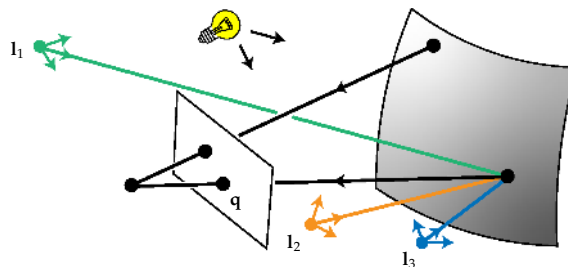


Figure 6.1.2: Viewing geometry of opportunistic ToF.

exposure to the next. This results in a geometric problem that is considerably more general than device localization, and whose main goal is 3D imaging.

In the area of single-photon imaging, systems that combine a pulsed laser with single-photon avalanche diode (SPAD) pixels are widely used for lidar [253, 125, 72], light-in-flight imaging [171], non-line-of-sight imaging [172, 270], and biomedical imaging [90, 116]. These methods fundamentally depend on precise hardware synchronization between laser and detector [77], and on extensive modeling—as well as full control [106]—of the laser device [193, 85, 192]. Our work departs from this paradigm by relinquishing control over the laser and modeling its operation on the fly from the ambient reflections it produces. We show that this leads to a new form of laser-camera synchronization that is *computational* rather than hardware based, and thus can be done after all photon data has been captured. A key advantage of this approach is that one camera can be (computationally) sync-locked to any number of lasers operating asynchronously, something impossible to achieve with a hardware sync. While we build on the ultra-wideband imaging framework from Chapter 4 to realize this capability for opportunistic ToF, that framework is agnostic to the characteristics of light sources in a scene and, crucially, is limited to cases where no strong ambient illumination (such as room lighting or sunlight) overpowers other sources.

Central to our approach is the observation that scattered light from pulsed lasers leaves a distinctive pattern in the temporal frequency domain of incident flux. This pattern—a *sinc comb*—is specific to each laser, invariant to the scene, and enables opportunistic, highly precise sync-locking even in the presence of far stronger light sources.

We use this core ability in our experimental system to demonstrate powerful new ultrafast imaging capabilities: (1) a first-of-its-kind visualization of asynchronously propagating light pulses from multiple lasers through the same scene; (2) passive estimation of a laser’s MHz-scale pulse repetition frequency with mHz precision from scattered light alone; (3) mm-scale 3D imaging over room-scale distances by passively harvesting photons from out-of-view lasers; and (4) achieving these results under challenging conditions of strong ambient light (signal-to-background ratio $\sim 1\%$) and SPAD dead time effects.

6.2 Opportunistic Time of Flight

Consider a camera that is imaging an unknown scene in the presence of an unknown number of pulsed lasers, and potentially other (non-ToF) ambient illumination. We model each laser as a point source that periodically emits pulses from a distinct, unknown 3D location (Figure 6.1.2). Both the

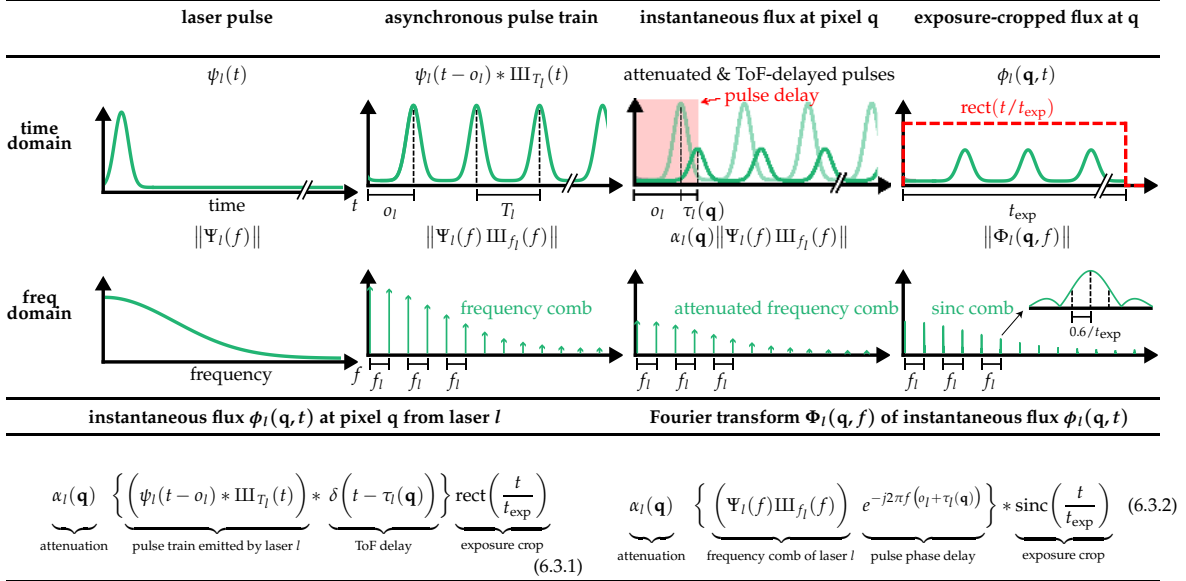


Figure 6.3.1: Incident flux model for pulsed laser light. **Top and middle row:** Our model’s individual components in the temporal and frequency domain, respectively. Note that the pulse delay $o_l + \tau_l(q)$ in top row (column 3) is the only ToF-related quantity that a camera can measure passively with its internal clock. The full width at half max (FWHM) of the sinc in middle row (column 4) is equal to $1.2/t_{\text{exp}}$. **Bottom row (Eqs. 6.3.1 and 6.3.2):** We use $\delta(t)$ to denote the impulse function, $\text{III}_{T_l}(t)$ the impulse train with spacing T_l , and $*$ for convolution.

camera and the lasers are assumed to be stationary during an exposure of duration t_{exp} but can move freely and independently between exposures. Our objective is to automatically detect these lasers and use their scattered light to reconstruct a pixelwise depth map of the camera’s 2D field of view.

Passive imaging conditions. We assume that neither the camera nor the individual lasers exchange timing signals of any kind, *i.e.*, they operate independently and asynchronously from each other. In this entirely passive setting, the camera must rely on its own internal clock to measure time. We use t to denote elapsed time according to the camera’s clock, with $0 \leq t \leq t_{\text{exp}}$.

Our focus is on the general case where the sources are outside the camera’s field of view and their operation—pulse profile, repetition frequency, sync signal, *etc.*—are all unknown. Under such conditions, the time of flight from individual sources to individual camera pixels cannot be measured as in conventional lidar; it must be estimated along with the 3D coordinates, pulse emission properties and sync timing of the lasers themselves.

6.3 Incident Flux Model for Pulsed Laser Light

The light signal transmitted from a pulsed laser to individual camera pixels carries rich information about the laser itself. This is because the laser’s periodic emission creates a distinctive sinc comb pattern in the temporal frequency domain of incident flux. This pattern is invariant to the scene, highly specific to the laser, and occurs at any pixel that receives non-negligible light from it.

More specifically, suppose that laser l emits a train of pulses over the interval $[0, t_{\text{exp}}]$ with repetition frequency f_l and period $T_l = 1/f_l$ (Figure 6.3.1, first row). Since the laser is not synchronized with the camera, these pulses are emitted with an unknown clock offset $t = o_l$. The pulses then propagate through the scene at the speed of light, becoming attenuated by various factors as they travel [29]—squared-distance fall-off, surface reflection, radiant intensity of the laser source, *etc.* Consequently, the incident flux at a camera pixel \mathbf{q} due to direct surface reflection will also be an attenuated and time-delayed pulse train.

Radiometrically, this train can be expressed as a time-varying function $\phi_l(\mathbf{q}, t)$ that describes instantaneous flux at time t in units of photons per second. This function depends on the temporal profile of the laser’s pulse $\psi_l(t)$; the speed-of-light propagation delay $\tau_l(\mathbf{q})$ from the laser to the pixel; the pulse’s attenuation $\alpha_l(\mathbf{q})$; and the laser’s repetition frequency and clock offset. Equations (6.3.1)-(6.3.2) in Figure 6.3.1 provide full expressions for the incident flux and its Fourier transform, respectively; we provide further insight about them below.

The laser frequency comb. A basic property of pulse trains is that their spectrum is non-zero only at the harmonics of the repetition frequency, *i.e.*, the integer multiples of f_l (Figure 6.3.1, second row). This set of frequencies is known as a *frequency comb* [58] and has three important ramifications in our context. First, with repetition frequencies between 100 kHz and 50 MHz typical for pulsed lasers used in lidar [135, 260, 83, 10], the frequency spectrum of incident flux is extremely sparse, with gaps of potentially millions of Hz between non-zero values. Second, the flux spectrum may have many thousands of non-negligible harmonics because laser pulses must have broad spectral support to allow precise ToF measurements.¹ Third, the frequency comb depends only on the repetition frequency, not the ToF delay at a pixel. Thus, any pixel receiving direct reflections from a laser provides essentially equivalent information about the laser’s presence and its repetition frequency.

Frequency-blurring impact of exposure. Cameras capture incident light over a finite duration, not indefinitely. Mathematically, restricting the incident flux to the interval $[0, t_{\text{exp}}]$ is equivalent to convolving its frequency-domain representation with an exposure-stretched sinc function. This convolution transforms the laser’s frequency comb into a comb of sinc functions, effectively blurring the flux spectrum around each harmonic. For example, an exposure time of $\frac{1}{100}$ seconds will “blur” each harmonic by approximately 60 Hz.

Multiple lasers and the asynchrony advantage. When many lasers emit light in the same scene, the incident flux at a pixel will be a superposition of their individual contributions (Figure 6.4.1). Crucially, lasers whose repetition frequencies are not precise multiples of each other will have largely distinct frequency spectra, as their harmonics will rarely fall within each other’s narrow sinc neighborhood (except for DC). Thus, asynchronous lasers leave a distinct “signature” on the spectrum of incident flux that makes their flux contribution potentially separable from each other’s. In practice, the degree of asynchrony needed for separability is naturally present in many lasers due to changes in temperature, power supply, and mechanical vibrations, *etc.* [228].

¹For instance, a one-nanosecond Gaussian pulse—equivalent to roughly 30 cm of light travel—has significant support up to 1 GHz.

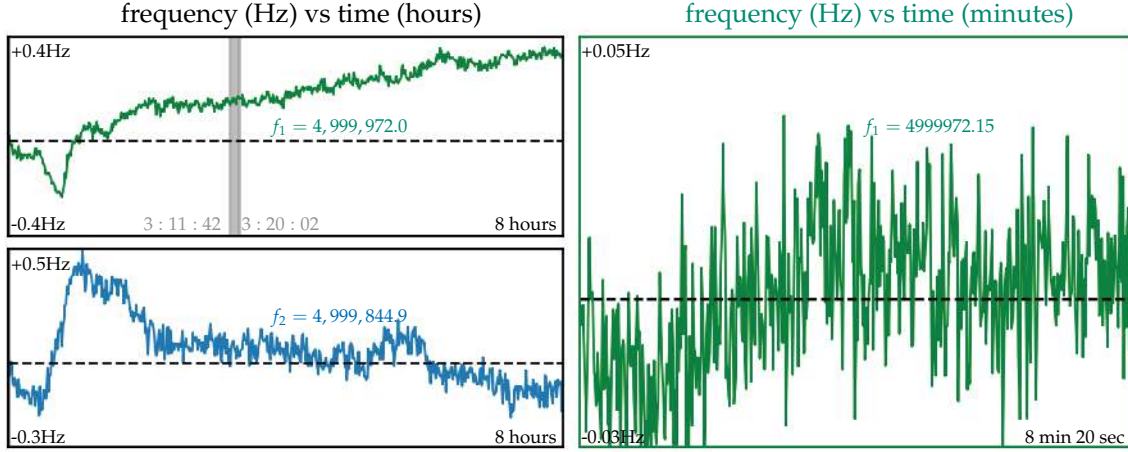


Figure 6.3.2: **Left:** Frequency estimates of two laser generators operating at 5 MHz over an 8-hour period. **Right:** Zoomed-in frequency estimates for a duration of 8 minutes and 20 seconds.

Figure 6.3.2 demonstrates the natural asynchrony between lasers using two lasers set to a nominal operating frequency of 5 MHz. By capturing markers from the lasers, we estimated their operating frequencies. Despite using identical laser drivers for each laser, we observe the operating frequencies of the two lasers differing by 128 Hz (Figure 6.3.2 (left)) immediately and exhibiting additional drift of nearly 1 Hz over 8 hours of constant operation. While Hz-scale relative stability over that interval is maintained, the absolute difference between the frequencies of even two identical devices can be significantly larger (over 100 Hz in our lasers).

Incident flux from other ambient sources. Unlike the broadband and sparse sinc combs produced by lasers, non-ToF ambient sources contribute negligibly to the incident flux at MHz to GHz frequencies. For instance, sunlight contributes only to DC at the sub-second timescales considered here and indoor lights, which flicker at frequencies ranging from Hz to tens of kHz [224, 82, 265], lack the sub-microsecond precision necessary to produce harmonics in the MHz to GHz range. Consequently, these sources introduce only random noise with near-zero amplitude in the Fourier domain of the incident flux, which we consider small enough to ignore.

6.4 Flux Probing Framework for Opportunistic ToF

Analyzing the flux received at a pixel from pulsed lasers is challenging due to its broad spectral support and weak amplitude—especially when mixed with much stronger contributions from ambient non-ToF sources. To address this problem for opportunistic ToF, we adopt and generalize the flux probing theory from Chapter 5. The approach leverages the capabilities of emerging single-photon avalanche diode (SPAD) cameras to passively reconstruct weak flux signals spanning frequencies from Hz to tens of GHz. Here we summarize key aspects relevant to this chapter.

Given a frequency f and an asynchronous stream of photon timestamps at a SPAD pixel, we formulate an unbiased estimator for the Fourier coefficient $\Phi(f)$ and use the same CFAR detection criterion as in Equation (4.3.7) to establish if f has significant support in the timestamp data:

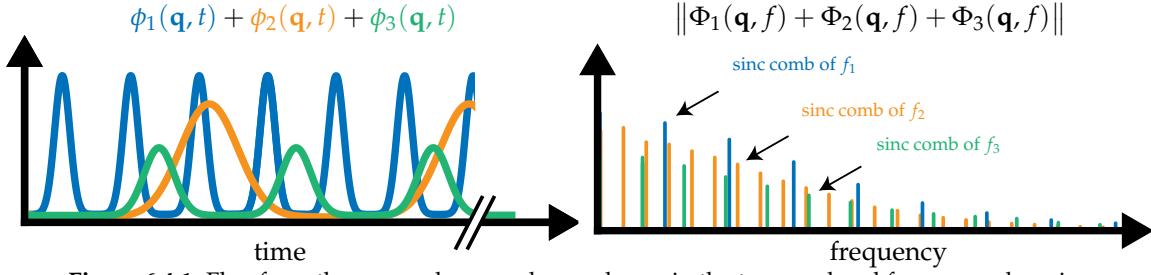


Figure 6.4.1: Flux from three asynchronous lasers shown in the temporal and frequency domains.

Probing flux frequency f : Given a stream of photon timestamps \mathcal{T} , the Fourier coefficient estimator

$$p_f(\mathcal{T}) = \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} \exp(-j2\pi ft), \quad (6.4.1)$$

approximately follows a complex normal distribution with mean $\Phi(f)$ and covariance matrix proportional to $|\mathcal{T}|$, the number of timestamps in \mathcal{T} .

CFAR frequency detector: We restate the CFAR frequency detector for convenience. To achieve a constant probability of false alarm α , frequency f is detected if

$$|p_f(\mathcal{T})|^2 \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \frac{|\mathcal{T}|}{2t_{\text{exp}}^2}, \quad (6.4.2)$$

where CDF_{χ^2} is the chi-squared cumulative distribution function.

To reconstruct the time-varying incident flux at a pixel, the SPAD's entire DC-to-GHz bandwidth is densely scanned for frequencies that pass the detection criterion.² Function $\phi_l(\mathbf{q}, t)$ is then expressed as a Fourier series using the estimated Fourier coefficients at those frequencies.

The CFAR detector and the dense frequency-scanning approach of Algorithm 4.1 used in passive ultra-wideband imaging have four limitations in this current setting.

Source-agnostic flux estimation. Opportunistic ToF requires detecting and separating the contribution of individual laser sources, not merely recovering their combined contribution to incident flux.

Photon noise. The 3D precision of opportunistic ToF implicitly depends on the GHz-scale harmonics of pulsed lasers because those harmonics allow the arrival of individual pulses to be localized more precisely in time. However, detection and reconstruction of a laser's GHz harmonics is inherently limited by photons coming from other light sources in a scene (including other lasers). Those photons raise the variance of the estimator $\hat{\Phi}(\mathbf{q}, f, \mathcal{T})$ for every frequency f , contributing noise to the entire spectrum of estimated flux and raising the noise floor of the CFAR detector. This makes GHz harmonics less likely to be detected since they naturally have lower amplitude (Figures 6.1.1 and 6.5.2).

²In Chapter 4, we proposed scanning the range [DC, 15 GHz] at 0.6 Hz based on the system's timestamp precision and exposure time used at the time.

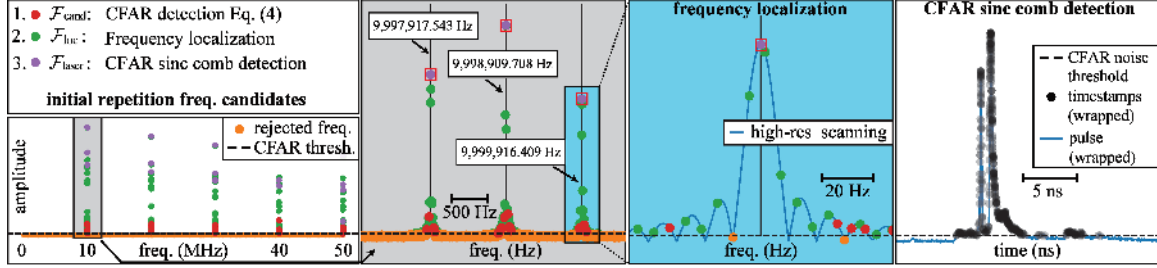


Figure 6.5.1: Overview of frequency detection. Our approach consists of four steps: (1) we apply constant false alarm rate (CFAR) detection to identify all candidate frequencies $\mathcal{F}_{\text{cand}}$ from zero to 50 MHz (red, green, blue dots in left and center left panels); (2) we obtain a refined set of frequencies \mathcal{F}_{loc} (green dots) by performing a high-resolution search around each candidate frequency (center left, center right) and precisely localizing frequencies whose MHz- to GHz-rate harmonics are above the CFAR threshold; (3) we identify all dominant frequencies \mathcal{F}_{rep} (purple dots) by reconstructing the time-domain flux using each frequency and its harmonics and checking if the flux has statistically significant support at the timestamp locations (right); (4) we output fundamental frequencies (red boxes) by removing all harmonics and side-lobe detections in the dominant frequencies.

SPAD dead time. SPADs become inactive for a nanosecond scale interval after each photon detection, known as the dead time. While dead time does not impact the precision of individual photon timestamps, it may cause photons to be missed, particularly as the timespan between consecutive photon arrivals approaches (or falls below) the dead time. Those missed photons reduce the terms included in the sum of Eq. (6.4.1), introducing a bias in the estimator. Importantly, when some light sources are much brighter than others (room lights, a strong laser, *etc.*), photon detections from a weaker laser may get so infrequent that the Fourier coefficients of their harmonics no longer pass the CFAR detector. This exacerbates the impact of photon noise.

Computational considerations. Densely scanning the GHz band is inefficient since only a tiny fraction of scanned frequencies will ever fall near a laser’s sinc comb (Figure 6.4.1).

6.5 Laser Discovery & Synchronization

We bypass the limitations of flux probing by exploiting the frequency-domain structure of incident flux from pulsed lasers (Section 6.3). Our method is based on a key observation: the laser’s repetition frequency defines the full set of non-zero Fourier frequencies in the flux spectrum, via its sinc comb. Therefore, by precisely identifying the repetition frequency, we can also determine the GHz harmonics of all pixels receiving the laser’s light, and vice versa.

Building on this observation, we (1) develop a method that efficiently searches the entire MHz-to-GHz band for candidate sinc combs—rather than isolated frequencies—and (2) formulate a sinc comb detection criterion to evaluate each candidate’s support in the timestamp data, enabling the detection of individual lasers. See Figure 6.5.1 for an illustration and Section 6.5.1 for pseudocode.

Initial repetition frequency candidates. We identify a set $\mathcal{F}_{\text{cand}}$ of candidate frequencies by densely scanning the expected range of repetition frequencies for lasers in a scene, including in the set only frequencies that pass the CFAR detector (Eq. 4.3.7).

Frequency localization by high-res scanning. Candidate frequencies may be on the side lobe of a sinc instead of its peak. To ensure each candidate represents a local extremum in the flux spectrum, we perform a high-resolution scan in the neighborhood of each candidate and adjust its frequency accordingly, to obtain a localized set \mathcal{F}_{loc} .

Frequency pruning by second harmonic detection. Noise may produce a large number of candidate frequencies. These frequencies do not define a sinc comb and can be pruned by examining their second harmonics. Specifically, we first localize the second harmonics of all candidates in \mathcal{F}_{loc} and then prune any candidates whose localized second harmonics fail the CFAR detector.

mHz frequency localization by harmonic hopping. Higher-order harmonics are extremely sensitive to small changes in a laser’s repetition frequency. Conversely, by precisely localizing the highest detectable harmonic of a candidate frequency, we can achieve frequency localization at a much finer scale—potentially thousands of times more precise than direct localization. Starting with a candidate in \mathcal{F}_{loc} we “hop” to its next higher harmonic, localize it, apply the CFAR detector, and repeat recursively until the detector fails. If the order- n harmonic is the highest detected and is localized to frequency f , we replace the candidate frequency in \mathcal{F}_{loc} with f/n .

Pulse train reconstruction by harmonic probing. Each candidate frequency f still in \mathcal{F}_{loc} corresponds to a potential laser signal whose harmonics extend to the GHz band, but whose individual harmonics may not be detectable. To leverage the signal those harmonics provide, we reconstruct the pulse train of a putative laser emitting pulses at frequency f by (1) estimating the Fourier coefficients of all harmonics of f up to the SPAD’s band limit using Eq. (4.3.2), and (2) expressing the train $\hat{\phi}_f(\mathbf{q}, \mathcal{T})$ as a Fourier series.

CFAR sinc comb detection and synchronization. Including all harmonics of a candidate frequency f that does not form a sinc comb will add only noise. To capture this intuition, we formulate a novel CFAR detection criterion in the time domain to test whether a frequency’s reconstructed pulse train is synchronous with any photon detections. Specifically, a candidate frequency f is detected if its pulse train $\hat{\phi}_f(\mathbf{q}, \mathcal{T})$ exceeds a noise threshold for at least one timestamp in \mathcal{T} . Specifically, we derive a distribution for the reconstructed pulse train when using harmonic probing with frequencies that do not align with a laser pulse train:

Proposition 6.1 (Distribution of Periodic Flux Estimator)

If $\Phi(\mathbf{q}, f) = 0$ and $t \in t_k$, then $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+T}{t_{\text{exp}}} + \langle p, \phi \rangle, \langle p^2, \phi \rangle\right)$ where $p(t) = \frac{2}{t_{\text{exp}}} \sum_{n=1}^N \cos(2\pi n f(t - t_k))$.

Proposition 6.1 tells us that the flux reconstructed from such harmonics follows a normal distribution, with its mean and variance determined by a probing function equal to the sum of cosines and the flux. As a corollary, we can derive analytical expressions for both the mean and variance under negligible dead time:

Corollary 6.1 (Periodic Flux Estimator under Negligible Dead Time)

If $\Phi(\mathbf{q}, f) = 0$, $t \in t_k$, and dead time is negligible, then $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+|\mathcal{T}|}{t_{\text{exp}}}, \frac{2N|\mathcal{T}|}{t_{\text{exp}}^2}\right)$.

Corollary 6.1 tells us that under negligible dead time, the mean and the variance of our estimator is uniquely determined by the number of photons, the number of harmonics, and the exposure time. Intuitively, if the set of harmonics does not align with the pulse train frequencies, summing over more harmonics adds more noise and increases the variance of the result. Leveraging this insight, we can identify pulse repetition frequencies that reconstruct pulse trains because the distribution of the corresponding photon timestamps is statistically distinct from ambient noise. Mathematically, this can be expressed as:

Definition 6.1 (CFAR sinc comb Detector)

To achieve a constant probability of false alarm p , a candidate frequency f is detected if:

$$\mathcal{T}_{\text{dom}} = \{t \in \mathcal{T} \mid \hat{\phi}_f(t) > \Delta\} \neq \emptyset, \quad (6.5.1)$$

where $\Delta = \frac{2N+|\mathcal{T}|}{t_{\text{exp}}} + \text{CDF}_{\mathcal{N}}^{-1}(1-p) \frac{\sqrt{2N|\mathcal{T}|}}{t_{\text{exp}}}$, the CFAR sinc detector threshold, N is the number of harmonics, and t_{exp} is the exposure time.

The number of detected frequencies is the number of pulsed lasers in the scene.

Boosting SNR by local timestamp aggregation. All the above steps involve a sequence of probing operations and thus critically depend on the SNR of individual Fourier coefficients computed by Eq. (6.4.1). To boost it, all probing operations described above are applied to the *union of timestamps* detected in small N -sized neighborhoods of pixels. The intuition here is that all pixels in a neighborhood are likely to receive light from the same laser and thus they all contribute to the same Fourier coefficients with correlated phases. This effectively treats each neighborhood as a single “superpixel,” and results in an SNR boost whose upper bound is \sqrt{N} from Poisson noise. We find that aggregation is particularly important in low-SBR conditions, where the timestamps at individual pixels are overwhelmed by background photons and blinded by dead-time effects (Figure 6.5.2).

Pulse-delay map estimation. Upon detection of a pulsed laser with repetition frequency f , the pulse train arriving from that laser at each pixel \mathbf{q} (Figure 6.3.1, third column) is reconstructed independently. Specifically, we apply the harmonic probing procedure to that pixel’s timestamps without aggregation to ensure that photons with different ToFs do not affect the phase of the pulse train’s Fourier coefficients. Once the train $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$ is reconstructed, we compute the pulse delay $\phi_f + \tau_f(\mathbf{q})$ by finding the global maximum of its period-wrapped counterpart, $\hat{\phi}_f(\mathbf{q}, (t \bmod 1/f), \mathcal{T})$.

6.5.1 Method pseudocode

We include the pseudocode of our frequency detection and pulse delay map estimation method discussed above, matching the name of each pseudocode to the corresponding paragraph header.

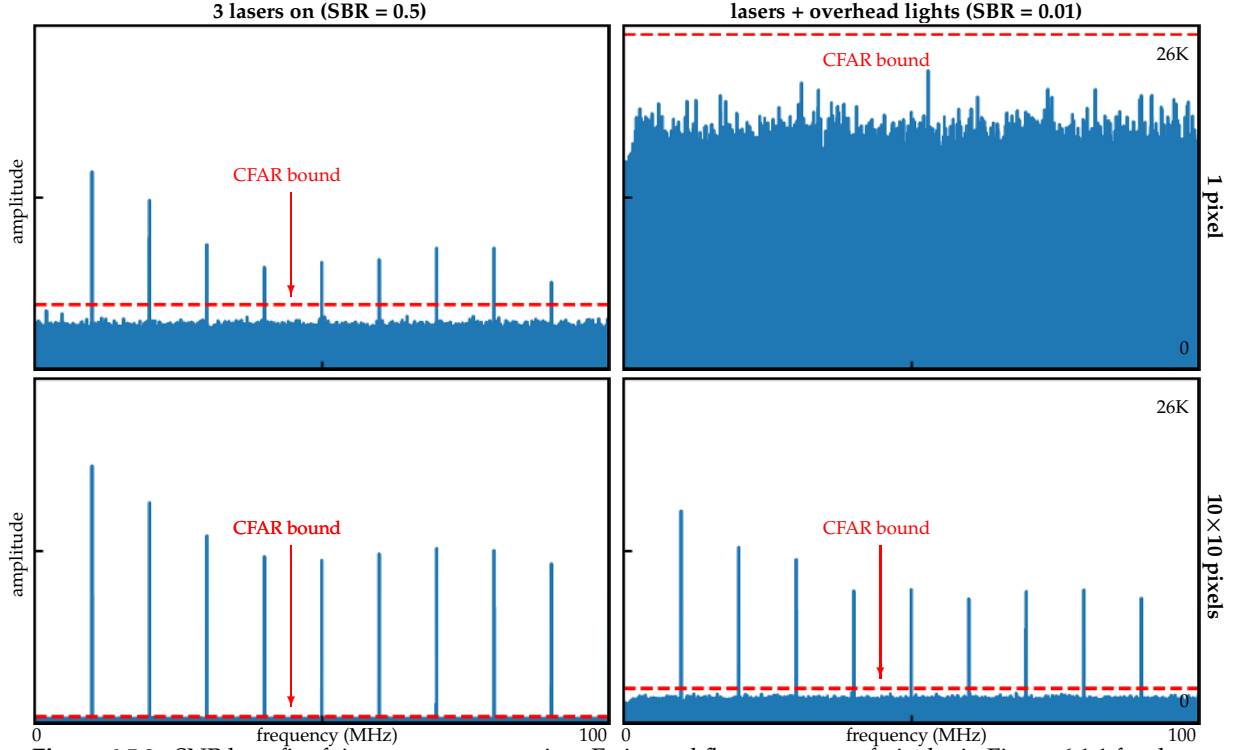


Figure 6.5.2: SNR benefit of timestamp aggregation. Estimated flux spectrum of pixel \mathbf{q} in Figure 6.1.1 for the range [DC, 100 MHz]. With overhead lights on and no notch filter on the SPAD to block them, no laser can be detected by probing a single pixel's timestamps.

Pseudocode 6.1: Initial repetition frequency candidates

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp}
Output: $\mathcal{F}_{\text{cand}}$ - set of candidate repetition frequencies

```

 $\mathcal{F}_{\text{cand}} \leftarrow \emptyset$ 
# sweep frequencies up to 50 MHz with step size  $0.6/t_{\text{exp}}$ 
 $\Delta_f \leftarrow 0.6/t_{\text{exp}}$ 
 $\mathcal{F} \leftarrow \{f \mid f < 50 \text{ MHz and } f = n\Delta_f \text{ for } n \in \mathbb{N}\}$ 
 $p \leftarrow \frac{1}{|\mathcal{F}|}$ 
for  $f$  in  $\mathcal{F}$ :
    # Equation (6.3.1) in Section 6.3
     $\hat{\Phi}(\mathbf{q}, f, \mathcal{T}) \leftarrow \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} \exp(-j2\pi ft)$ 
    # Equation (6.3.2) in Section 6.3
    if  $\|\hat{\Phi}(\mathbf{q}, f, \mathcal{T})\|^2 \geq \text{CDF}_{\chi^2}^{-1}(1-p) \frac{|\mathcal{T}|}{2t_{\text{exp}}^2}$ :
         $\mathcal{F}_{\text{cand}} \leftarrow \mathcal{F}_{\text{cand}} \cup \{f\}$ 
return  $\mathcal{F}_{\text{cand}}$ 

```

In general, we find that after CFAR sinc comb and synchronization, $\mathcal{F}_{\text{laser}}$ usually contain multiple laser frequencies clustered within a sinc lobe, higher-order harmonic frequencies, and occasionally frequencies that lie in the side sinc lobes of the laser repetition frequencies. To remove spurious frequencies, we cluster all laser frequencies that lie within the same sinc lobe, keeping only the laser frequency f with the sharpest peak, *i.e.* $\max \hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$ is highest among all frequencies in the sinc lobe. Then we remove higher-order harmonics by projecting them to first-order laser fre-

Pseudocode 6.2: Frequency localization by high-res scanning

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp} , $\mathcal{F}_{\text{cand}}$, $\Delta_{f_{\text{fine}}}$
Output: \mathcal{F}_{loc} - set of localized repetition frequencies

$\mathcal{F}_{\text{loc}} \leftarrow \emptyset$
for f in $\mathcal{F}_{\text{cand}}$:
 # Frequency probing with fine step around the first harmonic of candidate frequency
 $\mathcal{P} \leftarrow \{ \|\hat{\Phi}(\mathbf{q}, f + k\Delta_{f_{\text{fine}}}, \mathcal{T})\|^2 \mid |k\Delta_{f_{\text{fine}}}| < 0.6/t_{\text{exp}}, k \in \mathbb{Z} \}$
 let f_{loc} be the frequency that satisfies $\|\hat{\Phi}(\mathbf{q}, f_{\text{loc}}, \mathcal{T})\|^2 = \max \mathcal{P}$
 $\mathcal{F}_{\text{loc}} \leftarrow \mathcal{F}_{\text{loc}} \cup \{f_{\text{loc}}\}$
return \mathcal{F}_{loc}

Pseudocode 6.3: Frequency pruning by second harmonic detection

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp} , \mathcal{F}_{loc} , $\Delta_{f_{\text{fine}}}$
Output: \mathcal{F}_{loc} - set of localized repetition frequencies

for f in \mathcal{F}_{loc} :
 # Frequency probing with fine step around the second harmonic of candidate frequency
 $\mathcal{P} \leftarrow \{ \|\hat{\Phi}(\mathbf{q}, 2f + 2k\Delta_{f_{\text{fine}}}, \mathcal{T})\|^2 \mid |k\Delta_{f_{\text{fine}}}| < 0.6/t_{\text{exp}}, k \in \mathbb{Z} \}$
 if $\max \mathcal{P} < \text{CDF}_{\chi^2}^{-1}(1-p) \frac{|\mathcal{T}|}{2t_{\text{exp}}^2}$:
 $\mathcal{F}_{\text{loc}} \leftarrow \mathcal{F}_{\text{loc}} \setminus \{f\}$
 else:
 let f_{loc} be the frequency that satisfies $\|\hat{\Phi}(\mathbf{q}, f_{\text{loc}}, \mathcal{T})\|^2 = \max \mathcal{P}$
 $\mathcal{F}_{\text{loc}} \leftarrow (\mathcal{F}_{\text{loc}} \setminus \{f\}) \cup \{ \frac{f_{\text{loc}}}{2} \}$
return \mathcal{F}_{loc}

Pseudocode 6.4: mHz frequency localization by harmonic hopping

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp} , \mathcal{F}_{loc} , $\Delta_{f_{\text{fine}}}$
Output: \mathcal{F}_{loc} - set of localized repetition frequencies

for f in \mathcal{F}_{loc} :
 for $i = 3$ to 10 :
 $n \leftarrow 2^i$
 # Frequency probing with fine step around the n-th harmonic of candidate frequency
 $\mathcal{P}_n \leftarrow \{ \|\hat{\Phi}(\mathbf{q}, nf + nk\Delta_{f_{\text{fine}}}, \mathcal{T})\|^2 \mid |k\Delta_{f_{\text{fine}}}| < 0.6/t_{\text{exp}}, k \in \mathbb{Z} \}$
 if $\max \mathcal{P}_n \geq \text{CDF}_{\chi^2}^{-1}(1-p) \frac{|\mathcal{T}|}{2t_{\text{exp}}^2}$:
 let f_{loc} be the frequency that satisfies $\|\hat{\Phi}(\mathbf{q}, f_{\text{loc}}, \mathcal{T})\|^2 = \max \mathcal{P}_n$
 $f \leftarrow \frac{f_{\text{loc}}}{n}$
 else:
 break
return \mathcal{F}_{loc}

quencies. Finally, we remove any frequency f in the side lobe by probing the frequency intervals $(f - 1.5t_{\text{exp}}, f - 0.5t_{\text{exp}})$ and $(f + 0.5t_{\text{exp}}, f + 1.5t_{\text{exp}})$ with a step size of 1 mHz and removing f if

Pseudocode 6.5: Pulse train reconstruction by harmonic probing

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp} , f
Output: $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$ - reconstructed pulse train at each timestamp t in \mathcal{T}
 $N \leftarrow \max\{n \in \mathbb{N} \mid nf < 15 \text{ GHz}\}$
for $n = -N$ to N :
 # Probe the n -th harmonic of f
 $\hat{\Phi}(\mathbf{q}, nf, \mathcal{T}) \leftarrow \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} \exp(-j2\pi nft)$
for $t \in \mathcal{T}$:
 # Reconstruct pulse train at each timestamp
 $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T}) \leftarrow \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}) \exp(j2\pi nft)$
return $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$

Pseudocode 6.6: CFAR sinc comb detection and synchronization

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , t_{exp} , \mathcal{F}_{loc} , p - probability of false alarm
Output: $\mathcal{F}_{\text{laser}}$ - set of detected pulse repetition frequencies
 $\mathcal{F}_{\text{laser}} \leftarrow \emptyset$
for f in \mathcal{F}_{loc} :
 let $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$ be the output of Pseudocode 6.5 with inputs \mathcal{T} , t_{exp} , and f
 # CFAR threshold for pulse train
 $z_p \leftarrow \text{CDF}_{\mathcal{N}}^{-1}(1 - p)$
 $\Delta \leftarrow \frac{2N+|\mathcal{T}|}{t_{\text{exp}}} + z_p \frac{\sqrt{2N|\mathcal{T}|}}{t_{\text{exp}}}$
 $\mathcal{T}_{\text{laser}} \leftarrow \{t \in \mathcal{T} \mid \hat{\phi}_f(\mathbf{q}, t, \mathcal{T}) > \Delta\}$
 # Frequency detected if pulse train exceeds CFAR threshold for at least
 # one timestamp in \mathcal{T}
 if $\mathcal{T}_{\text{laser}} \neq \emptyset$:
 $\mathcal{F}_{\text{laser}} \leftarrow \mathcal{F}_{\text{laser}} \cup \{f\}$
return $\mathcal{F}_{\text{laser}}$

we observe a higher probing value in the side lobes.

6.6 Geometric Optimization

Pulse delays from individual lasers to specific pixels impose joint geometric constraints on the 3D positions of the lasers and the scene's depth map. This is because all light paths from lasers in the scene to a given camera pixel share a common 3D segment along the ray through that pixel (Figure 6.1.2). More specifically, any pulse delay due to a direct surface reflection from a laser to a pixel must satisfy the relation

$$\underbrace{c \left(\tau_l(\mathbf{q}) + o_l \right)}_{\text{pulse delay at pixel } \mathbf{q}} = \underbrace{\|\mathbf{l}_l - d(\mathbf{q})\mathbf{v}(\mathbf{q})\|_2}_{\text{laser } l \text{ to scene point along ray}} + \underbrace{d(\mathbf{q})}_{\text{depth of } \mathbf{q}} \quad (6.6.1)$$

where c is the speed of light, \mathbf{l}_l is the laser's 3D position, and $\mathbf{v}(\mathbf{q})$ is the unit vector along the ray through pixel \mathbf{q} . It follows that Eq. (6.6.1) defines a total of LP algebraic equations for L lasers

Pseudocode 6.7: Pulse-delay map estimation

Input: \mathcal{T} - set of timestamps from pixel \mathbf{q} , $\mathcal{F}_{\text{laser}}$
Output: $o_f + \tau_f(\mathbf{q})$ - pulse delay for $f \in \mathcal{F}_{\text{laser}}$ at pixel \mathbf{q}
for f in $\mathcal{F}_{\text{laser}}$:
 $N \leftarrow \max\{n \in \mathbb{N} \mid nf < 15 \text{ GHz}\};$
 for $n = -N$ to N :
 # Probe the n -th harmonic of f
 $\hat{\Phi}(\mathbf{q}, nf, \mathcal{T}) \leftarrow \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} \exp(-j2\pi nft);$
 # Period-wrapped counterpart of pulse train $\hat{\phi}_f(\mathbf{q}, t, \mathcal{T})$
 compute $\hat{\phi}_f(\mathbf{q}, (t \bmod 1/f), \mathcal{T}) = \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}) \exp(j2\pi nf(t \bmod 1/f));$
 $o_f + \tau_f(\mathbf{q}) \leftarrow \operatorname{argmax} \hat{\phi}_f(\mathbf{q}, (t \bmod 1/f), \mathcal{T});$
return $o_f + \tau_f(\mathbf{q})$ for each f in $\mathcal{F}_{\text{laser}}$

and P pixels in terms of the P unknown depths and the $4L$ laser-specific unknowns (3D positions and clock offsets). In scenes with complex geometry, we expect these equations to result in an overdetermined (but non-convex) problem akin to bundle adjustment [242, 216]. To solve it while also accounting for shadows and outliers, we minimize the adaptive loss function of Barron [22] on the geometric error defined by Eq. (6.6.1) using stochastic gradient descent [105].

In practise, we rewrite Equation (6.6.1) to formulate the passive 3D geometry reconstruction task instead as the following optimization problem:

$$\underset{o_l, \mathbf{l}_l, d(\mathbf{q})}{\text{minimize}} \quad \rho_{\alpha, \gamma}(\|\mathbf{l}_l - d(\mathbf{q})\mathbf{v}(\mathbf{q})\| + d(\mathbf{q}) - c(\tau_l(\mathbf{q}) + o_l)) \quad (6.6.2)$$

$$\text{subject to} \quad d(\mathbf{q}) \geq 0 \text{ for all } \mathbf{q}, \alpha > 0, \text{ and } \gamma > 0, \quad (6.6.3)$$

where $\rho_{\alpha, \gamma}$ is Barron’s adaptive robust loss defined by adaptive parameters α, γ with constraint $\alpha > 0$ and $\gamma > 0$ [22]. We randomly initialize our estimation parameters $o_l, \mathbf{l}_l, d(\mathbf{q})$ in the range between $[0, 1)$, and use Adam optimizer [105] with 0.001 learning rate to run optimization until convergence. Convergence is defined as the point at which the average change of $o_l, \mathbf{l}_l, d(\mathbf{q})$ all fall below 10^{-3} . Typical number of iterations required for convergence ranges from 100K to 1M.

6.7 Experiments

In practice, 2D single-photon cameras are an emerging technology [152, 252, 177] and we do not have one available to us; nevertheless, we create an experimental setup based on a single-pixel SPAD and a single pulsed laser that accurately emulates what a single-photon camera *would capture* in a scene filled with fast ambient light signals caused by asynchronous pulsed laser sources. Using this setup we demonstrate (1) simultaneous discovery and mHz-accurate computational synchronization to multiple asynchronous laser sources, (2) post-capture separation and visualization of their asynchronously propagating wavefronts, and (3) mm-accurate 3D imaging and laser source localization over room-scale distances, including under strong ambient light with $\text{SBR} \ll 1$.

6.7.1 Hardware, 2D acquisition procedure & algorithm details

Hardware. We use the following sensor and laser sources in our experiments:

- **Free-running SPAD:** We use a single-pixel PDM Series Fast-Gated SPAD from Micro Photon Devices, operated in asynchronous mode. The SPAD has a pixel pitch of 50 micrometers, a timing jitter of 68 picoseconds, a 231-nanosecond dead time, and a quantum efficiency of 30% at 532 nm.
- **Time-to-digital converter (TDC):** The SPAD's output is converted into a stream of timestamps using a Swabian TimeTagger Ultra. The TimeTagger Ultra's timing resolution is 1 picosecond, and its RMS timing jitter is 8 picoseconds. The TDC features four independent channels, one of which was used for the SPAD.
- **Galvo mirrors:** For experiments where 2D imaging was required, we used a pair of galvo mirrors from Thorlabs (Model GVS012) to acquire measurements over a 2D field of view with our single-pixel SPAD.
- **Ultrafast high-power pulsed laser:** We used a class-4 picosecond laser with a wavelength of 532 nm (NKT Photonics Katana 05HP). We vary its repetition rate from 9.993 MHz to 10 MHz during our acquisition and set its power to 0.2W. Its FWHM was 110 ps under the experimental conditions we employed, resulting in a flux function with frequency support up to 10 GHz.
- **Mirrors:** We used Thorlabs E01-BB02 mirrors to redirect the collimated laser beam to various points in the scene.
- **Picosecond laser drivers:** We used two Alphas PLDD-100M picosecond drivers as reference clocks for measuring frequency drift.
- **Pulsed laser illumination:** We use mirrors (Thorlabs E01-BB02) to redirect the laser beam towards a piece of A4 paper taped onto the wall. The A4 paper acts as a diffuser, scattering laser light onto our scene.

2D acquisition procedure. Since we had access only to a single pixel SPAD, the only way to create concurrent imaging is to (1) scan the scene point by point, (2) repeat scanning for each laser position, and (3) compute the union of the timestamps from the pulsed lasers and the ones from ambient light while also enforcing dead time. Specifically:

1. **Temporal pixel alignment:** In order to simulate concurrent imaging, the photon timestamps of all pixels must be computationally shifted relative to each other to create the effect of a common absolute clock. We use the laser synchronization markers to provide temporal alignment of the photon timestamps within the laser repetition period.
2. **Scanning and timestamp stream collection:** We used the galvo mirrors to scan the scene one point at a time, with each point collecting photons for the same time duration (100 milliseconds unless otherwise stated).
3. **Concurrent imaging of multiple asynchronous pulsed lasers:** Since we had access to only a single pulsed laser source, we repeat the previous steps as many times as there were light sources considered in the scene (typically 3, unless stated otherwise). The final timestamp stream for each pixel is obtained by combining the individual timestamps from each pulsed laser source along with ambient timestamps, while accounting for dead time. Specifically, if two photons from different streams were detected within the dead time after alignment, one of them is re-

moved.

4. **Total acquisition time:** Total acquisition times vary by experiment and are noted in individual experiments.

Algorithm details. We clarify details about the various parts of the algorithm outlined in Sections 6.5 and 6.6:

- **Initial candidate frequencies.** We identify candidate repetition frequencies by frequency-scanning the DC-50 MHz range with step size $0.6/t_{\text{exp}}$. 50 MHz is chosen as an upper bound for repetition frequencies of typical lasers used in lidar [10]. $0.6/t_{\text{exp}}$ is chosen according to Section 4.4.1.
- **Boosting SNR by local timestamp aggregation.** We randomly choose a single 10 by 10 neighborhood of pixels to aggregate timestamps and then perform probing operation on this “superpixel” in all our experiments. This aggregation boosts SNR by up to 10x and reduces the effect of dead-time under strong ambient light conditions while preserving memory and runtime efficiency.
- **Pulse-delay map estimation.** We discretize the period-wrapped counterpart $\hat{\phi}_f(\mathbf{q}, (t \bmod 1/f), \mathcal{T})$ using 10000 evenly spaced samples over the domain $[0, 1/f)$. We estimate the pulse delay by first identifying the time at which the discretized flux reaches its maximum. We further refine this estimate by using it as the starting point for Newton’s algorithm, running it for 100 iterations to compute the pulse delay by locating the global maximum of $\hat{\phi}_f(\mathbf{q}, (\tau \bmod 1/f), \mathcal{T})$.

6.7.2 Visualizing propagating wavefront of multiple asynchronous lasers

In this scene, we detect three laser sources and separate their flux contribution to create the first visualization of light propagation from multiple separated asynchronous laser sources (Figure 6.1.1, column 3).

Scene. The scene itself is shown in Figure 6.1.1 (top left). It consisted of two mannequins separated by a whiteboard covered with black fabric. The whiteboard was intentionally placed between the mannequins to cast shadows from at least one laser. In this scene, the single-photon camera is positioned 1.12 m from right wall, and 2.99 m from the top wall. The three light sources are positioned approximately 2.16 meters apart. We scanned 512×512 points, with each scan requiring 30 hours per pulsed laser position.

Illumination. We used the procedure outlined in Section 6.7.1 to acquire timestamps corresponding to 3 pulsed lasers outside the SPAD’s field of view and approximately 2.16 m away from each other. The pulse repetition frequency for each point was set to 9.998 MHz, 9.999 MHz, and 10.000 MHz, respectively.

Scanning and timestamp stream collection. We used the galvo mirrors to scan the field of view with a spatial resolution of 0.38 cm, resulting in a 512×512 image. We used an exposure time of $t_{\text{exp}} = 0.1$ seconds for timestamp stream acquisition at each galvo position. Average photon counts at every pixel acquired is 11040 photons. Photon counts at individual pixels vary significantly due to shading and shadows: differences in photon counts between brightest and dimmest pixels are 15159, 9581, 9270 for laser 1, 2, and 3, respectively.

Frequency scanning and detection. We aggregate timestamps from a 10×10 patch. We then probe from 0 Hz to 50 MHz with a step size of 6Hz. We set the constant probability of false alarms to 1 over the number of probed frequencies for the initial candidate frequencies, *i.e.* $0.12 \cdot 10^{-7}$. From this step, 362 candidate frequencies were above the CFAR bound which were then refined as part of our frequency detection method (Algorithms 6.2—6.6). Three laser frequencies passed the CFAR sinc comb detector. The whole process from scanning to detection takes 1.9 hours.³

Laser separation. Computational laser synchronization is essential for passive 3D imaging, as it reconstructs the pulse train of each laser, enabling the generation of pulse-delay maps used as inputs for the geometric optimization. For each detected laser frequency, we reconstruct its corresponding pulse train by performing harmonic probing (Algorithm 6.5) up to the 1500th harmonic for all pixels in the image. In this scene, reconstructing the pulse train for each detected laser frequency involves computing only 1500 Fourier coefficients per detected laser frequency, resulting in a total of 4500 coefficients for the three detected lasers.

Multi-laser transients. We create transient images by integrating the incident pulse train at each pixel over a 125-picosecond exposure. This corresponds to a frame rate of 8 billion frames per second. Figure 6.7.1 shows the transient before (left) and after (right), demonstrating the flux function decomposition into pulse trains from the detected pulsed lasers. In Figure 6.7.1 (right), we represent each laser with a distinct color for visualization purposes. Notably, the separation of laser pulses reveals intriguing phenomena occurring on the picosecond timescale. For instance, in the top row of Figure 6.7.1, the top wall (annotated with a red ellipse) appears illuminated by a single wavefront in the transient image of unseparated pulsed lasers. However, our method reveals that this effect actually arises from the simultaneous illumination of the wall by wavefront from two different pulsed lasers. This phenomenon is even more prominent in Figure 6.7.1 (row 2, left). This is consistently observed across the remaining rows in Figure 6.7.1.

Computational syncing. Our approach makes it possible to automatically sync-lock to one of the pulsed lasers, effectively freezing it in time, and allowing us to observe how the other pulsed lasers drift relative to it. Figure 6.7.2 (row 1) illustrates this first-of-its-kind visualization, where the “magenta” pulsed laser is frozen, while the other two lasers move forward in space as time progresses. To freeze a pulse, we select the same relative time across different periods, aligning with the laser’s cycle. However, the other pulses, having different periods, appear to propagate at varying speeds relative to the synchronized laser. An intriguing effect occurs when we sync-lock to the pulsed laser with the shortest period. Since this pulse repeats more frequently, it appears

³For comparison, Algorithm 4.1 takes approximately 60 hours.

stationary while the slower pulses propagate in the opposite direction. This phenomenon offers a novel visualization of aliasing in the transient domain, showing an optical illusion similar to how a fan can appear to spin backward.

Comparison with passive ultra-wideband imaging. We also compare to the results of passive ultra-wideband (UWB) imaging (Algorithm 4.1) in Figure 6.7.3 (row 2) for a different scene whose details are provided in Section 6.7.4). We probe frequencies up to 15 GHz, resulting in 7191 frequencies above the CFAR bound used to reconstruct flux. At high SBR (0.50) our method produces sharper pulses (110 ps FWHM vs 120 ps FWHM), with higher intensity peaks, and fewer ringing artifacts. Our approach is almost unaffected by low SBR of 0.01 (see Section 6.7.7 for more low-SBR results), while UWB fails to recover pulse profiles. Notably, UWB cannot separate the contributions of each source and computationally synchronize to them as our approach does (colored transients in Figure 6.7.3, row 2).

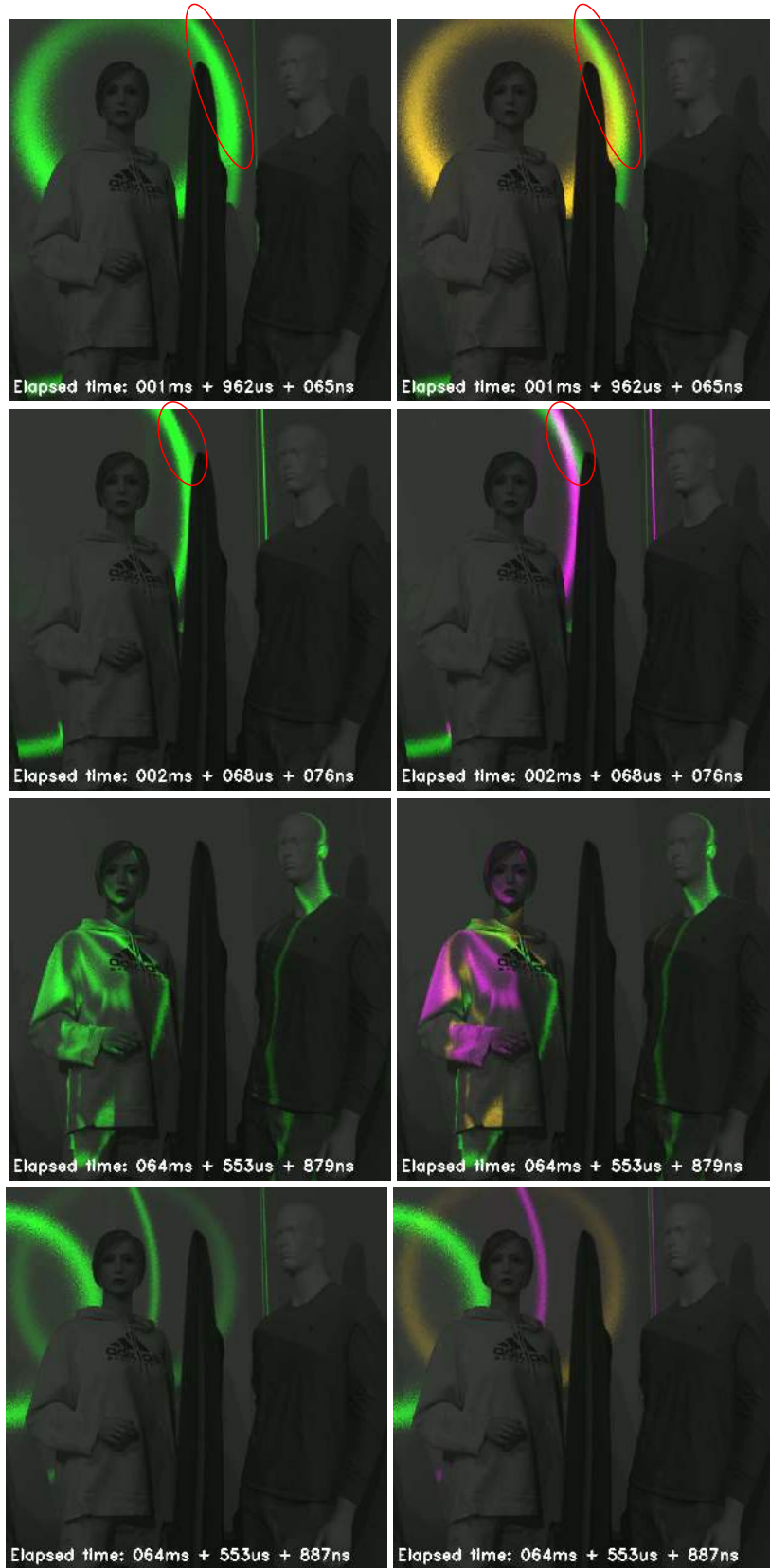


Figure 6.7.1: Post capture laser separation. **Left:** Transient images without separated pulsed laser contributions from our approach. **Right:** Transient images with separated pulsed laser contributions from our approach.

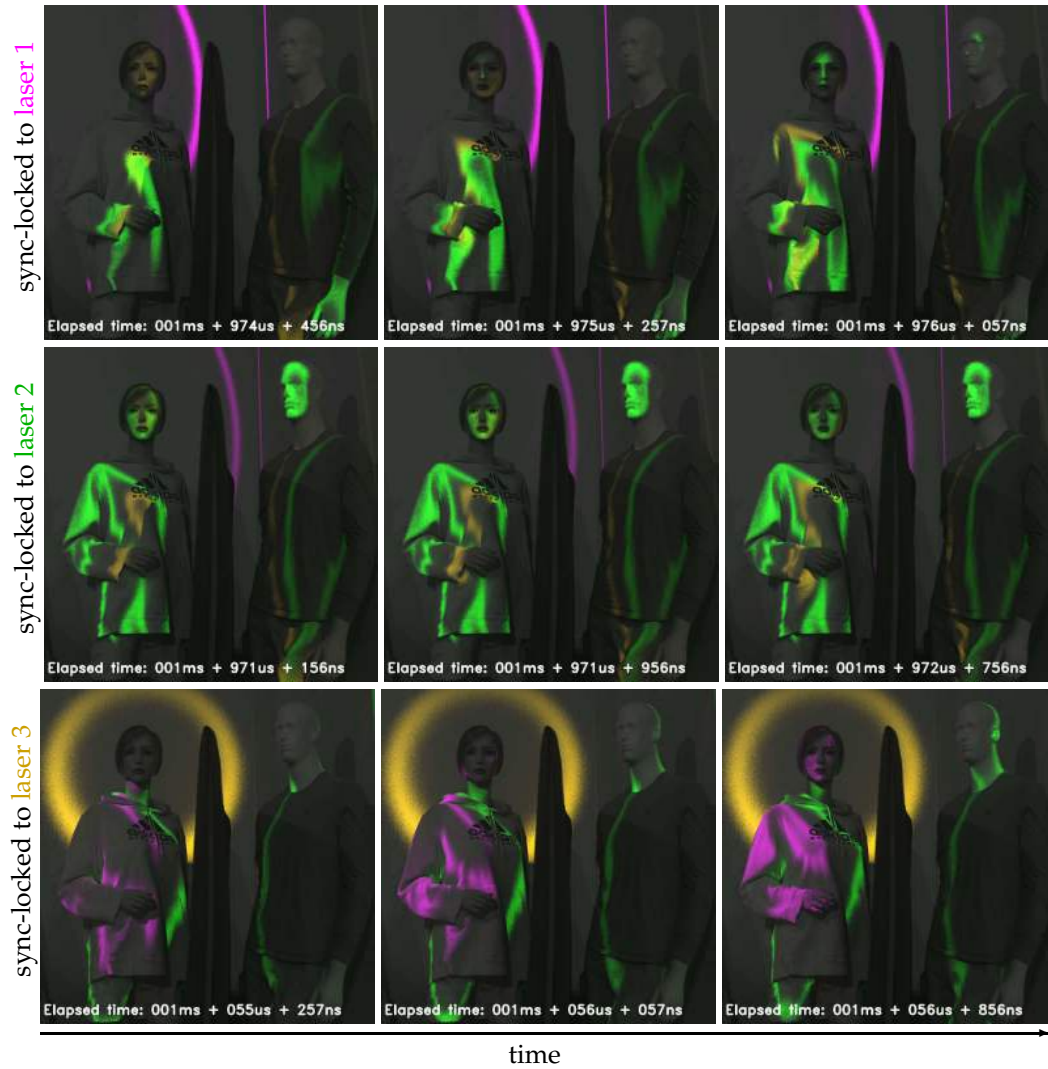


Figure 6.7.2: Computational sync locking. **Row 1:** Sync-locked to laser 1 (“magenta”), freezing its pulse in time while the other two lasers propagate forward. **Row 2:** Sync-locked to laser 2 (“green”), causing the “magenta” pulse to appear as though it is moving backward due to its longer period. **Row 3:** Sync-locked to laser 3 (“yellow”), making the other two lasers appear to move backward because of their longer periods.

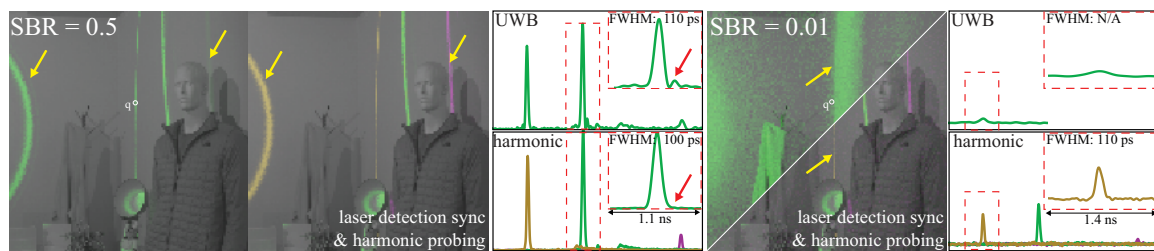


Figure 6.7.3: Our method improves over passive ultra-wideband imaging (Chapter 4) in both high SBR and photon-starved regimes, enabling detection and separation of asynchronous laser sources for the first time.

6.7.3 Opportunistic time-of-flight 3D imaging of room-scale scene

We detect three laser sources, separate them, localize their 3D position, and recover geometry and sync offsets in a room-scale scene.

Scene. We use the same scene and captured data as in Section 6.7.2. Additionally, we provide results for the case where two more pulsed lasers, with repetition rates 9.996 and 9.997 MHz, are present in the scene, resulting in a total of five pulsed laser sources.

This scene poses unique challenges for 3D reconstruction: the light sources are outside the camera’s field of view and have significant estimated sync offsets (6.57 m or 21.9 ns mean offset). Distance falloff, shading, and shadows cast by the divider cause the measurements for each laser source to have different SBRs (0.66, 0.83, 0.29) and photon counts (4,511, 4,367, and 2,161); shadowed regions result in outlier path length estimates.

Metrics for assessing 3D geometry recovery. Since we did not have access to ground truth geometry measurements, we evaluated the accuracy of the 3D reconstruction with other well-established geometric primitives. Specifically, we fit planes to the top and right walls in the scene by segmenting pixels based on their estimated depths. Pixels with depths greater than 3 meters in the left portion of the scene are extracted to represent the front wall. Similarly, pixels with depths greater than 2.1 meters in the right portion of the scene are extracted to represent the right wall. Figure 6.7.4 shows the segmented pixels for the top wall in blue (66,827 pixels) and right wall in green (22,442 pixels). We apply RANSAC [55] with an inlier threshold of 1 cm and a maximum of 1000 iterations to both sets of segmented pixels to compute the inlier ratio and plane-fitting RMSE. The angle between the two walls is calculated by taking the arccosine of the dot product of the normals of the fitted planes.

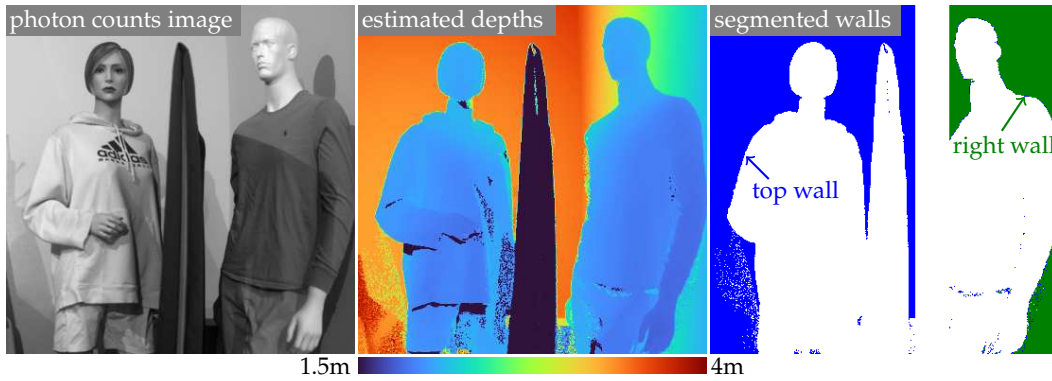


Figure 6.7.4: Left: Intensity image of the scene, gamma-corrected. Middle: Estimated depth map from our approach. Right: Wall segmentation using depth-based thresholding.

More than three light sources. We perform an experiment in which we add two more pulsed lasers to the scene with repetition frequencies of 9.996 MHz and 9.997 MHz. Using these, we perform frequency detection (Algorithms 6.1—6.6), pulse-delay estimation (Algorithm 6.7), and geometric optimization (Section 6.6) to recover depth, 3D locations of the pulsed lasers, and clock offsets.

# of pulsed lasers	Top wall plane fitting		Right wall plane fitting		Angle between fitted planes [°]
	RMSE [mm]	%Inlier	RMSE [mm]	%Inlier	
3	3.4	93.2	3.6	93.2	90.42
4	3.0	89.1	3.4	89.4	90.20
5	3.0	87.8	2.7	89.4	90.20

Table 6.1: Geometry reconstruction accuracy under different number of light sources evaluated with plane fitting.

We report quantitative metrics for plane fitting and the angle between fitted planes with varying numbers of pulsed lasers in Table 6.1. As the number of pulsed lasers increases, the RMSE values for the top wall and right wall show slight improvement, particularly for the right wall, where the RMSE decreases from 3.6 mm to 2.7 mm. The angle between the fitted planes remains consistent across all configurations, maintaining a value close to 90.2 degrees.

Figure 6.7.5 shows the quality of the reconstructed point cloud under varying numbers of pulsed lasers in the scene. Figure 6.7.6 (right) shows the reconstructed point cloud capture by Kinect Azure from a slightly different viewpoint, but at approximately the same distance from the scene (Figure 6.7.6 (left)). Table 6.2 presents the RMSE values for registering the reconstructed point clouds. Overall, our reconstructions remain consistent even with one-tenth of the exposure time.

Source point cloud	Target point cloud	%Inlier	RMSE[mm]
3 pulsed lasers, 0.01s exposure	3 pulsed lasers, 0.1s exposure	95.52	2.6
4 pulsed lasers, 0.1s exposure	3 pulsed lasers, 0.1s exposure	91.9	3.6
5 pulsed lasers, 0.1s exposure	3 pulsed lasers, 0.1s exposure	90.1	3.0
5 pulsed lasers, 0.1s exposure	4 pulsed lasers, 0.1s exposure	93.1	2.6

Table 6.2: Registration of reconstructed point clouds under varying number of light sources and exposures using ICP algorithm.

Two light sources. Optimization fails with only two light sources because occlusions result in an insufficient number of pixels observing both light sources (Figure 6.7.7).

6.7.4 3D reconstruction with dynamic camera & pulsed laser locations.

We also consider the case where the camera and the laser sources all move independently between acquisitions (Figure 6.7.8).

Scene. The scene is shown in Figure 6.7.8 and includes a mannequin wearing a jacket with millimeter-scale grooves, a figure holding a bowl, and 3D-printed staircases with step sizes 1 cm and 3 cm. We captured six different views following the acquisition procedure described in Section 6.7.1. The camera moved approximately 0.3 meters between views, and the pulsed laser positions changed across the views as well. We scanned 128×128 points, with each scan requiring 2 hours per pulsed laser position.

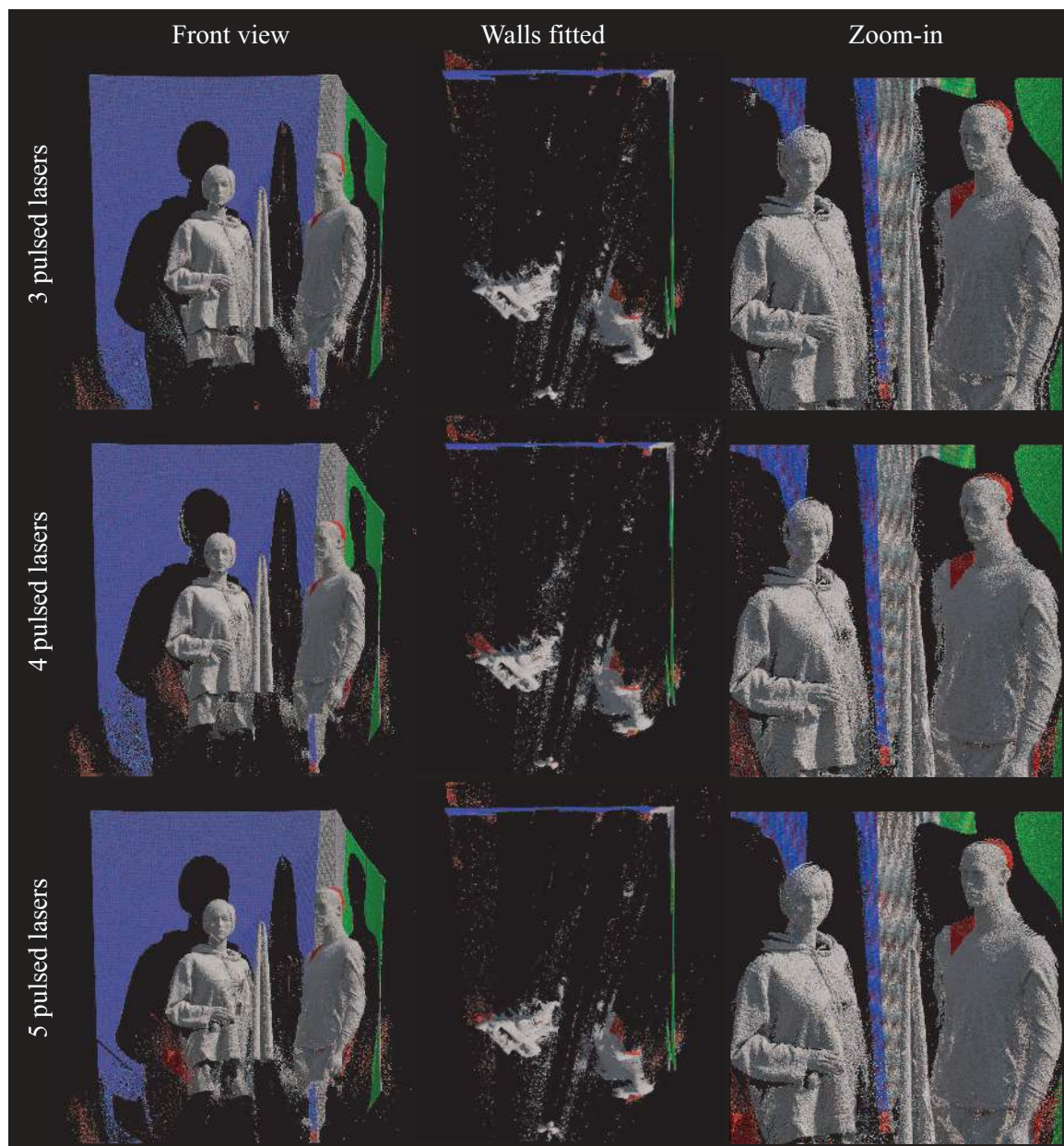


Figure 6.7.5: 3D reconstruction with a varying number of pulsed lasers. We depict inliers points used for the top wall fitting in blue and inlier points used for right wall fitting in green. Outlier points for both are shown in red, and unused points are shown in gray.

Accuracy assessment. In each acquisition, our method reliably recovers mm-accurate geometry, and successfully reconstructs concave shapes, such as a bowl shown in Figure 6.7.8 (view 1). Figure 6.7.8 (right) shows that the geometry remains consistent despite significant variations in laser source positions—registering the point clouds sequentially with ICP yields a mean RMSE of 1.1 cm, and we estimate the angle between the walls to be 90.89° by applying plane fitting to the walls of the registered point clouds.

Furthermore, using the ground truth CAD model of the staircases, we register the reconstructed point cloud of each staircase in each view with the corresponding ground truth shape using the

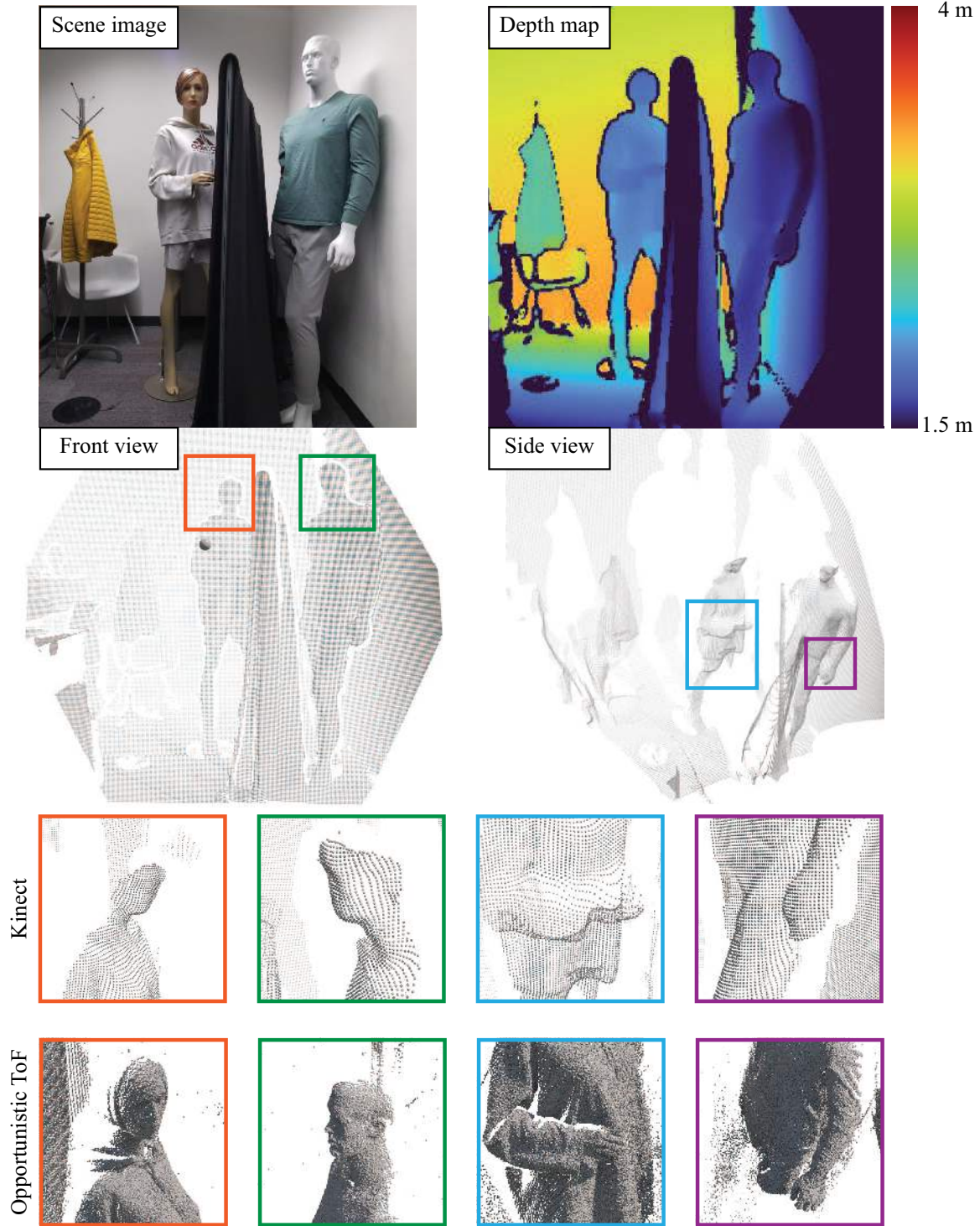


Figure 6.7.6: 3D reconstruction by Kinect Azure [18] for the scene in Figure 6.1.1. We provide zoomed in views of our reconstructed point cloud from similar vantage points.

ICP algorithm [27].⁴ The average registration RMSE error for the 1 cm and 3 cm staircase models is 2.6 mm and 2.0 mm respectively, across all views. The average distance between estimated camera

⁴We provide more details about the registration in the next section, Section 6.7.6.

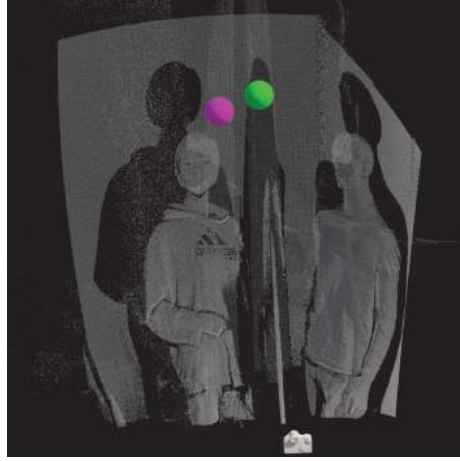


Figure 6.7.7: Optimization fails to converge with two pulsed lasers, as shown by the estimated pulsed lasers being too close in 3D space and the background surfaces appearing curved.

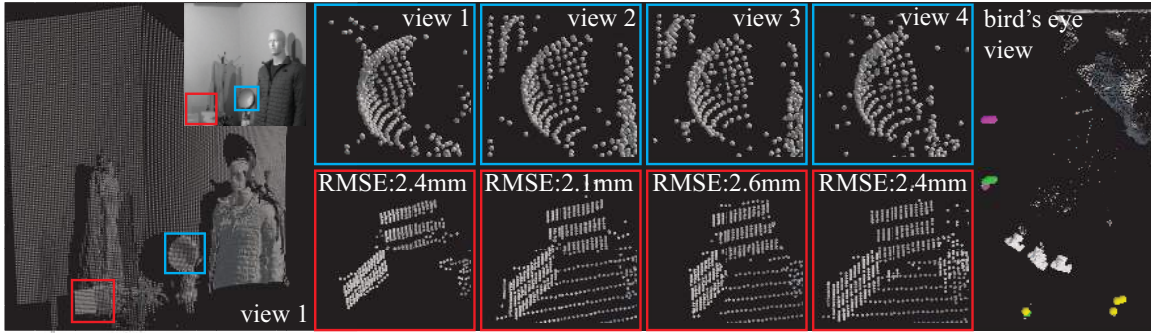


Figure 6.7.8: We turn a single-photon camera into a passive 3D imaging system capable of capturing scenes where the camera and light source move at each frame. We consistently recover mm-accurate geometry despite changes in the imaging configuration

positions from ICP is 32.7 cm.

	10cm block	5cm block	3cm block	1cm block	mixed block	overall
3 pulsed lasers	2.8mm	2.5mm	2.4mm	2.1mm	2.5mm	2.3mm
4 pulsed lasers	2.0mm	2.1mm	2.0mm	1.9mm	2.2mm	2.1mm
5 pulsed lasers	2.0mm	2.0mm	1.9mm	1.8mm	2.1mm	2.0mm
6 pulsed lasers	2.2mm	2.0mm	1.9mm	1.7mm	2.1mm	2.0mm
7 pulsed lasers	1.8mm	1.8mm	1.8mm	1.7mm	2.0mm	1.9mm
8 pulsed lasers	1.7mm	1.8mm	1.8mm	1.6mm	2.0mm	1.8mm

Table 6.3: RMSE of 5 staircase registrations from 3 to 8 light sources.

6.7.5 Quantitative experiments: pulse delay estimation against frequency error

We examine the sensitivity of our estimated frequency on the pulse delay estimation for the scene in Figure 6.1.1. Figure 6.7.9 displays frequency estimation error for the 10 MHz source in the corresponding scene; we create the error maps by applying Algorithms 6.2–6.4 to independent patches of 10×10 pixels. Under both high and low SBR conditions, our estimates are typically accurate to within 1 mHz and 10 mHz, respectively (assessed using the laser sync signal). Outlier frequency errors correlate with occlusions and shadowed regions (arrows in Figure 6.7.9). Figure 6.7.10 also

compares changes in the pulse-delay estimation resulting from frequency estimation changes ranging from 1 mHz to 10 Hz for frequencies 9.998, 9.999 and 10.0 MHz.

Figure 6.7.11 shows the mean and median pulse delay error versus the frequency errors. While we do not have access to ground truth frequencies and pulse delays, we observe that even a small change in the estimated frequency on the order of a few hertz can cause changes to nanosecond-scale shifts to the pulse delay map, which corresponds to several metres in the path length. This highlights the importance of requiring mHz-level accuracy in estimating the pulse repetition frequencies of the pulsed lasers for our approach, otherwise we will have pulse-delay maps with significant errors. Finally, this range of errors is incorporated into the simulations in Section 6.8 to evaluate the robustness of the geometric optimization.

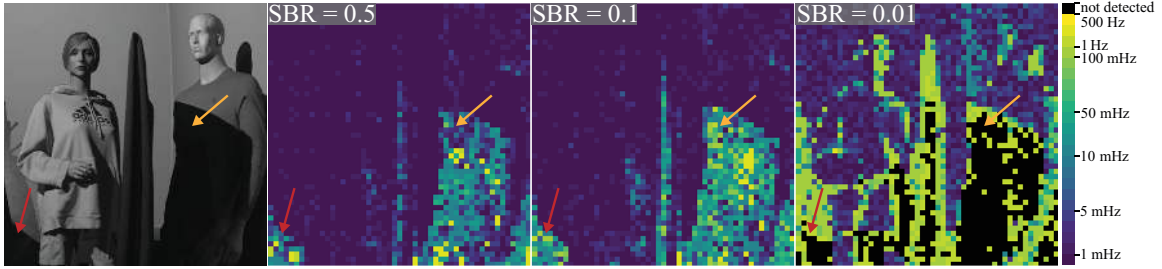


Figure 6.7.9: Per-pixel frequency error across different SBR levels; error in most areas is well below 100 mHz and detection fails only in shadow where the light source is occluded.

6.7.6 Quantitative experiments: assessing geometry with 3D-printed objects

We 3D-printed four staircases with equal step sizes of 10 cm, 5 cm, 3 cm, and 1 cm. Each camera-facing staircase face measures 20 cm x 7.5 cm. Additionally, we printed another staircase with step sizes of 3 cm, 2 cm, 1 cm, 8 mm, 6 mm, 4 mm, and 2 mm, with camera-facing faces measuring 20 cm x 3 cm. Figure 6.7.12 (row 1, right) shows a side profile of all the printed objects. The 3D printer had a precision of 0.2 mm.

Scene. The scene is shown in Figure 6.7.12 (row 1, left). It consists of five 3D printed models side-by-side at approximately 3 meters away from the SPAD. We scanned 114×407 points with $t_{\text{exp}} = 0.1$ s.

3D accuracy metrics. We segment each individual staircase based on the photon count image of the scene. We then register the reconstructed point cloud for each staircase with the ground truth staircase point clouds. The ground-truth point clouds are sampled from the CAD model of the staircases and contain the same number of points as the reconstructed staircases. Registration is performed using ICP with maximum correspondence distance of 0.01, voxel size of 0.001, initialization set to the identity matrix, tolerance of 10^{-9} and maximum number of iterations set to 5000. We summarize the overall inlier ratio and RMSE of 5 staircase registrations from 3 to 8 light sources in Table 6.3. We show pixel-wise RMSE error maps in Figure 6.7.16. Our reconstructions consistently achieve millimeter-level accuracy, regardless of the number of light sources used.

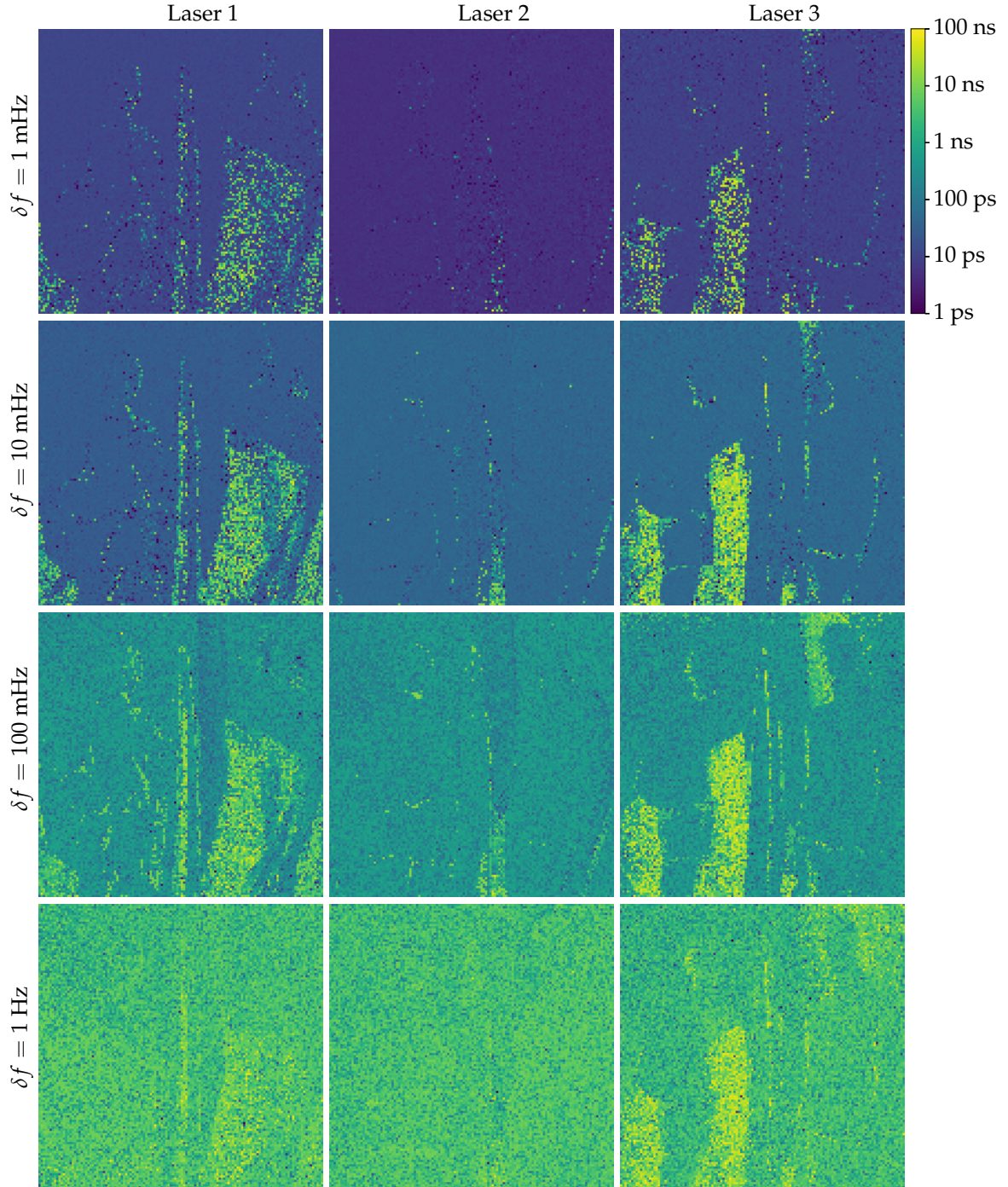


Figure 6.7.10: Pulse-delay error for varying frequency estimation errors for three different frequencies.

Comparison to Kinect. We captured the same scene using a Kinect Azure [18]. Figure 6.7.12 (rows 2–4, right) shows the reconstructed point cloud from Kinect, which is unable to accurately reconstruct the staircase with varying step sizes. Figure 6.7.12 (rows 2–4, middle) qualitatively shows that our method consistently outperforms the Kinect. Additionally, for the case of 3 light sources, our method achieved a 2.3 mm average RMSE compared to the Azure Kinect which had a 4.9 mm

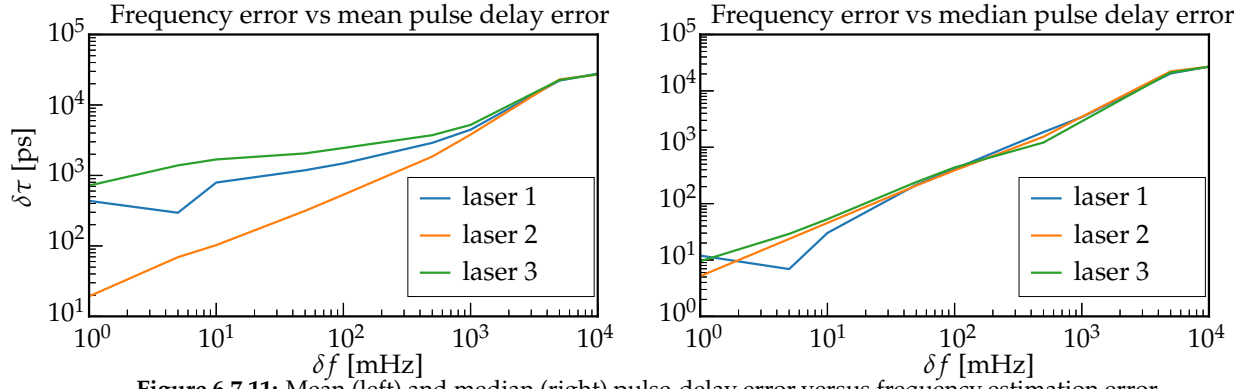


Figure 6.7.11: Mean (left) and median (right) pulse-delay error versus frequency estimation error.

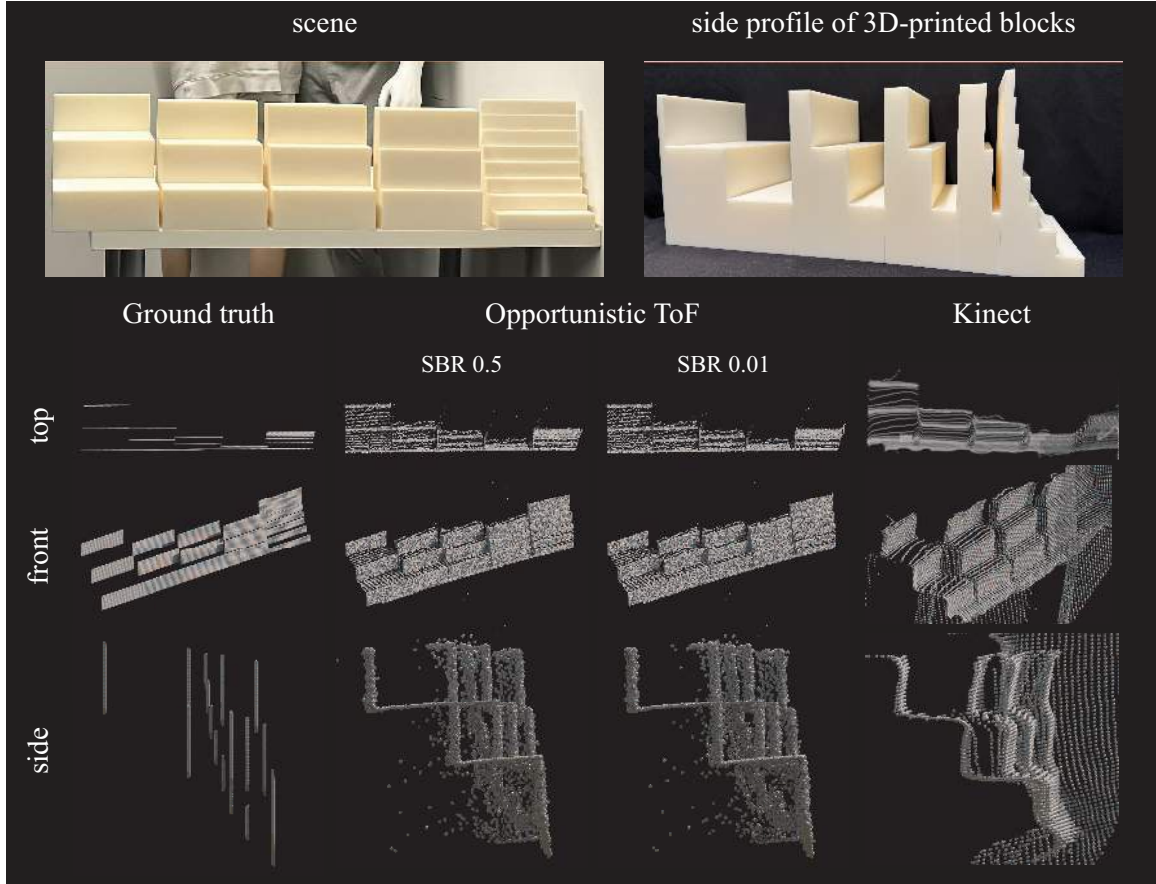


Figure 6.7.12: Reconstructing the 3D shape of calibration targets with known geometry. **Row 1 left:** View of the scene. **Row 1 right:** Side profile of the 3D-printed blocks; from left to right, the first four staircases have step sizes 10 cm, 5 cm, 3 cm and 1 cm respectively. The last one has steps of 3 cm, 2 cm, 1 cm, 8 mm, 6 mm, 4 mm, and 2 mm. **Rows 2–4 left:** Views of the ground truth CAD model. **Rows 2–4 middle:** Views of the estimated 3D point cloud from opportunistic time-of-flight under two SBR levels— high (0.5) and low (0.01). **Rows 2–4 right:** 3D reconstruction from Kinect Azure [18].

average RMSE. Reducing the SBR to 0.01 has little effect (2.5 mm avg. RMSE).

6.7.7 Opportunistic time-of-flight under strong ambient light

We also apply our method to the scene shown in Figure 6.1.1 under low SBR settings.

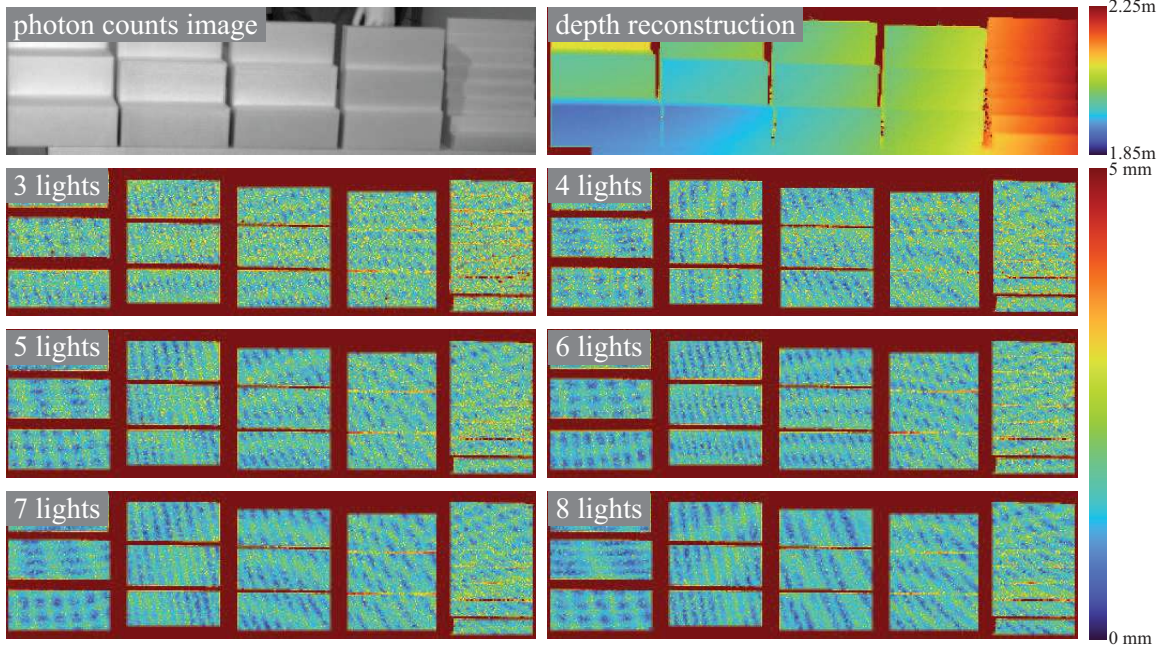


Figure 6.7.13: Pixel-wise depth error maps for 3 to 8 light sources.

SBR calculation. We define SBR β as the ratio of the number of signal photons to background photons. For the case with three lasers, the SBR is 0.5. In simulations with ambient light, we report the desired SBR level (i.e., the incident flux) rather than the resulting SBR after simulation, as dead time can introduce non-linear effects that may slightly impact the final SBR level. However, in our experiments, we observed no significant discrepancies. For example, a desired SBR level of 0.01 resulted in an average SBR of 0.0103.

Controlled low-SBR conditions. In order to evaluate performance under many levels of ambient light for low SBR and dead-time impacted conditions, we combined experimentally acquired timestamp data for a scene with simulated background light timestamps. Rather than capture ambient timestamp data for each scene, we instead simulate background light corresponding to placing a DC light source at the location of each pulsed laser in the scene with a flux equal to a multiple of the corresponding laser flux. The simulated timestamps from these DC sources have a flux which then corresponds to a multiple of the averaged flux across all pulsed laser flux levels. Our results are expressed in terms of the average SBR across all pixels.

In order to simulate a desired SBR level of β , we compute the average flux $\frac{1}{L} \sum_{l=1}^L \phi_l(\mathbf{q})$ at each pixel, i.e. the per-pixel laser photon count divided by the exposure time. We simulate ambient timestamps from a homogeneous Poisson process with flux parameter $\frac{1}{L\beta} \sum_{l=1}^L \phi_l(\mathbf{q})$. We then superimpose the laser photon stream with the ambient photon stream and apply a dead time of 231 ns. By default, the laser photons have an SBR of around 0.5 without the inclusion of ambient photons.

Scene and illumination. We use the same scene and illumination conditions outlined in Section 6.7.2.

Timestamp stream generation. We use the procedure outlined in Section 6.7.7 to generate timestamps of the Figure 6.1.1 scene with a desired SBR of 0.01. The average photon count at every pixel was 175384 photons, which is about 40.5% of the maximum photon count for a dead time of 231 ns. The average laser photon count per pixel was 5264. Compared to the high SBR case, our timestamp streams contained on average 17 times more photons and 2 times less laser photon.

Frequency scanning and detection. We aggregate timestamps from a single 10×10 patch and run our frequency detection method (Algorithms 6.2—6.6) on the aggregated timestamp stream. 80 candidate frequencies were above the CFAR bound and three laser frequencies passed the CFAR sinc comb detector. The whole process from frequency scanning to detection takes 10.3 hours.

Pulse-delay map estimation. We use the pulse-delay map estimation method (Algorithm 6.7) to recover the pulse delay per-pixel for each laser. For efficiency, we thin our timestamp data by culling 90% of the photons, resulting in average photon count of 17.5k during harmonic probing. Additionally, we recover the photon count images from the recovered pulse trains.

Figure 6.7.14 shows the photon count images, the pulse-delay maps and the pulse-delay error maps.

Discussion. We observe that the low-SBR photon count images (Figure 6.7.14 (row 1)) depict good separation of the laser photon counts, similar to the ones shown in Figure 6.1.1. However, these images are noisier as a result of the increased ambient light level, and we also observe some artifacts resulting from the ambient photons bleeding into the photon count images, highlighted in Figure 6.7.14 (top, red circle). Since the laser frequencies are well-estimated within a few mHz even at an SBR of 0.01 (Table 6.4), we observe that the pulse-delay maps (Figure 6.7.14 (row 2)) are estimated with high accuracy despite SBR of 0.01 and 10% thinning, differing from the high-SBR pulse-delay maps by 10s of picoseconds (Figure 6.7.14 (row 3)) in well-lit regions. We do observe that low-flux regions and shadows regions have significantly large pulse delay errors (> 3 ns), likely due to the reduction in signal laser photons in those areas. Figure 6.7.15 shows the estimated depth map, the reconstructed point cloud, and localized laser positions. Qualitatively, the depth maps are very similar, except the depth is noisier for more pixels in the low SBR case. This is consistent with noisy pulse delay estimations (Figure 6.7.14 (rows 2–3)). As a result, the resulting point cloud has significantly more floating points. Nevertheless, the result suggests that our method is still able to estimate depth well even under low SBR settings.

6.7.8 Quantitative experiments: frequency localization precision against SBR

We conduct an experiment in which we assess the effect of signal-background ratio (SBR) on our detection method (Algorithms 6.2—6.6).

Laser details. The laser frequencies are 9.998 MHz, 9.999 MHz, and 10.000 MHz. Our exposure time t_{exp} was 100 milliseconds.

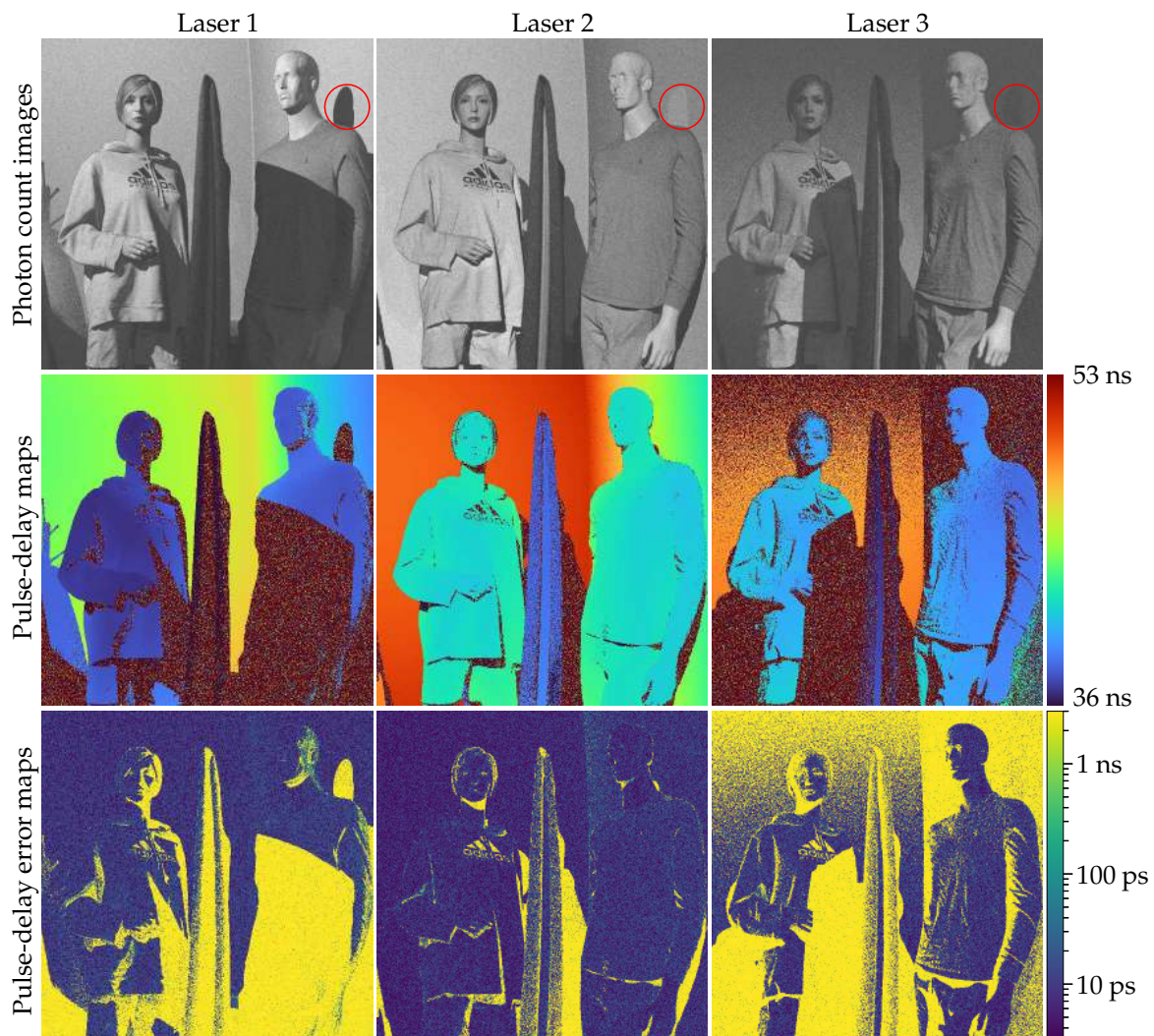


Figure 6.7.14: Figure 6.1.1 scene under SBR of 0.01. **Row 1:** Photon count images corresponding to each pulsed laser. **Row 2:** Pulse-delay maps. **Row 3:** Pulse-delay error map between high-SBR pulse delays and low-SBR pulse delays (log-scale).

Experimental procedure. We randomly select 30 pixels from the scene of Figure 6.1.1 and aggregate timestamps from a 10-by-10 neighbourhood at each pixel. For each neighbourhood, we evaluate our ability to detect the laser frequencies for SBR levels of 0.5, 0.1, 0.01, and 0.003, resulting in 120 random trials in total. For each trial, we ran Algorithms 6.2—6.6 and consider a laser frequency to be detected in this case if our method returns a frequency within a conservative bound of 500 Hz of the laser frequency.

Discussion. We report the percentage of estimated laser frequencies within 1 mHz, 10 mHz, and 100 mHz in Table 6.4 for each light source. We observe that in the high SBR (0.5 and 0.1) scenarios, we are able to estimate the laser frequencies to within sub-mHz for a majority of the trials with at least 80% of the trials having a precision within mHz. In the low SBR case, the accuracy suffered, but the majority of trials still returned mHz-precise frequency estimates. In all 120 trials, our method did not return spurious frequencies for any of the 120 trials.

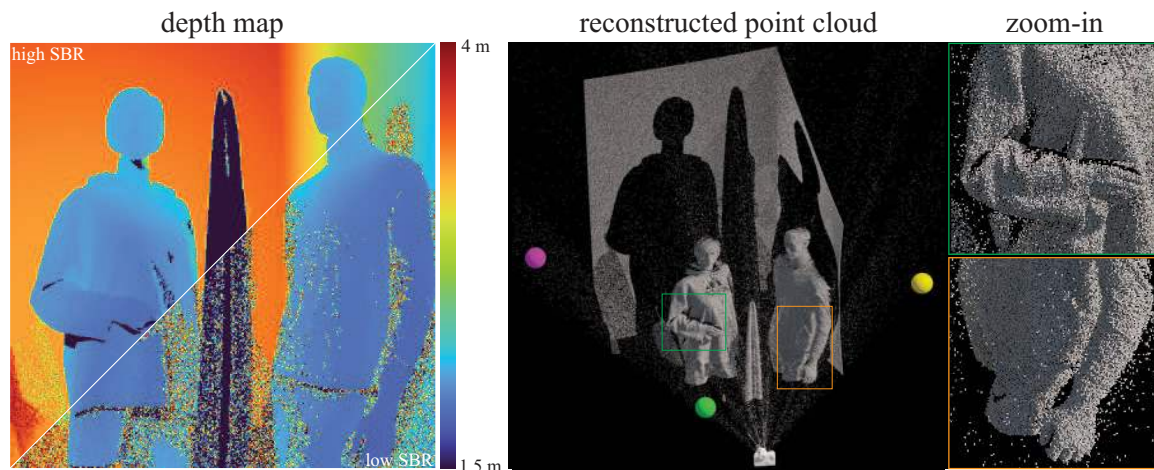


Figure 6.7.15: 3D reconstruction of Figure 6.1.1 scene under low SBR. **Left:** We show the reconstructed depth map for both high and low SBR settings. **Right:** We show the localized laser positions in 3D and the recovered point cloud.

	SBR	%<1 mHz	%<10 mHz	%<100 mHz	% detected
Laser 1	0.5	63.3	80	100	100
	0.1	56.7	93.3	96.7	100
	0.01	36.7	66.7	83.3	86.7
	0.003	33.3	60	66.7	66.7
Laser 2	0.5	96.7	100	100	100
	0.1	90.0	100	100	100
	0.01	50	100	100	100
	0.003	36.7	90	100	100
Laser 3	0.5	73.3	86.7	100	100
	0.1	46.7	86.7	100	100
	0.01	16.7	53.3	70	80
	0.003	6.7	20	40	46.7

Table 6.4: Frequency estimation accuracy across different SBR levels for the scene in Figure 6.1.1. Our method consistently detects the pulse repetition frequencies of the pulsed lasers with mHz even in 0.3% SBR.

6.7.9 Quantitative experiments: 3D geometry accuracy under strong ambient light

We evaluate the accuracy of our method’s 3D recovery for a scene with an SBR of 0.01. We emulated ambient photons for the scene in Figure 6.7.12 according to Section 6.7.7. Figure 6.7.16 shows the percentage of pixels with depth error less than 1, 2, and 3 mm respectively. Our method consistently reconstructs the same number of points with an error of less than 3 mm, even when the SBR is reduced by a factor of 50.

6.8 Simulations

In this section we use simulated scenes to evaluate individual components of opportunistic ToF against known ground truth.

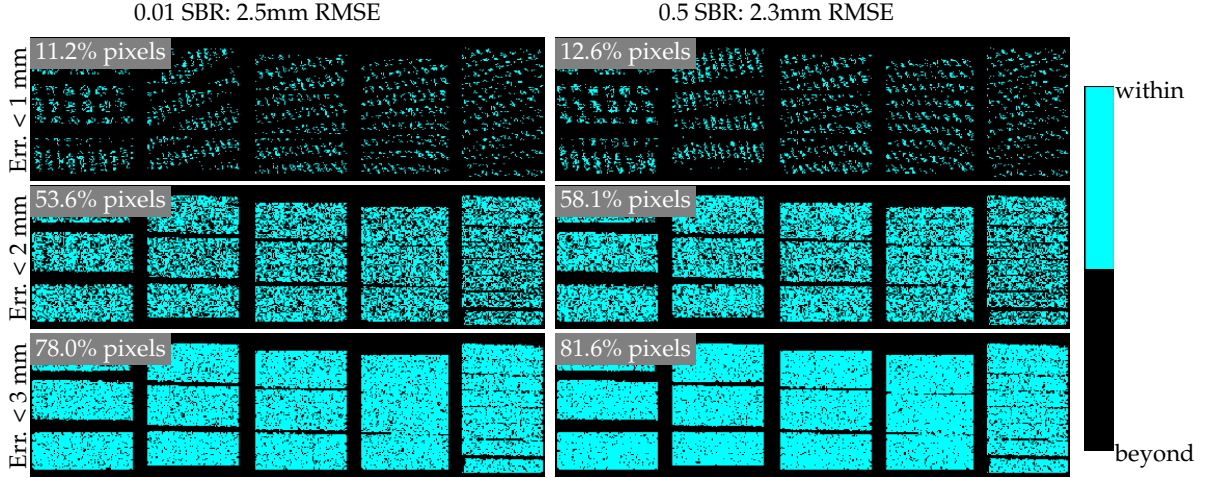


Figure 6.7.16: Pixel-wise depth error maps for 0.01 and 0.5 SBR with thresholds 1 mm, 2 mm, and 3 mm.

6.8.1 Simulation details

Simulation scenes and renderings. Our simulated scenes comprise 6 different room-scale scenes in the Mitsuba format from the collection of Bitterli [28].⁵ The depth range of the scenes varied from 4 m to 15 m. We use Mitsuba version 3.1.1 to extract the ground-truth 3D positions of each pixel, camera calibration parameters, depth maps, and pulse-delay maps for various randomly sampled 3 pulsed-laser configurations in each scene. To these pulse-delay maps, we manually add sampled clock offsets for each light source, with the offsets uniformly sampled from the range $[0, \text{maximum pulse delay at each source}]$. Unless otherwise stated, we use 128×128 pixels and $t_{\text{exp}} = 0.1$ s.

Photon timestamp generation of 2D room scenes. For realistic rendering of the 2D room scenes, we convert the Mitsuba 3 files into a Mitsuba-2-compatible format and use a transient renderer [205] for Mitsuba 2 to render a 4,000-bin histogram at every pixel for each laser. For each pixel, we then use the histogram to sample a timestamp stream for each laser using the thinning method [117]. We set the average flux to be 10 kilophotons per second which is similar to the lighting conditions of our real experiments. We combine the timestamp stream of each pulsed laser and apply a dead time of 231 ns to remove timestamps that arrive during the SPAD’s inactive period. The remaining timestamps are quantized to a resolution of four picoseconds.

Pulsed laser model from [205]. We set the laser pulse profile from [205] to a FWHM of 235 ps. The laser pulse profile is approximately Gaussian, but our use of the transient renderer enables us to capture multi-path transients. In the room scene simulation, we use 7.499, 7.5, and 7.501 mHz, as the repetition frequencies of the lasers.

Photon timestamp generation of single-pixel pulsed laser trains. In cases where we only simulate one pixel, we model a laser pulse as a Gaussian with a desired FWHM and use Cinlar’s [45] method instead to simulate the timestamps from a laser pulse train for efficiency as we have an

⁵The scenes we use are “bathroom1” (Contemporary Bathroom by Mareck), “livingroom1” (The White Room by Jay-Artist), “livingroom2” (The Modern Living Room by Wig42), “bedroom” (Bedroom by SlykDrako), “dining” (The Breakfast Room by Wig42), and “greywhiteroom” (The Grey & White Room by Wig42).

analytical expression for the laser pulse. To simulate the timestamps from L laser pulse trains, we generate L sets of timestamps for each laser pulse train separately and then we combine the L set of timestamps into a single stream. We apply a dead time of 231 ns, the nominal dead time of our system, to remove photons that arrive within the dead time period of a photon detection. We then apply a timing jitter of 8 ps and quantize the timestamps to 1 ps matching the nominal specs of our TDC.

6.8.2 3D reconstruction and localization consistency

We evaluate the consistency of passive 3D imaging by generating timestamps for 50 trials per simulated scene for a total of 300 trials. In each trial, we vary the pulsed laser positions (by uniformly sampling positions along the ceiling of each room) and clock offsets (by uniformly sampling between 0 and the laser repetition period) while keeping the pulse repetition frequency and average photon flux constant. We then use opportunistic time of flight to detect frequencies (Algorithms 6.2—6.6), estimate pulse-delay maps (Algorithm 6.7), and estimate depth, pulsed laser locations, and clock offsets (Section 6.6). Figure 6.8.1 shows some examples of rendered room scenes, the ground truth path lengths (converted from pulse delay maps via multiplication with the speed of light c) and the path length error. Table 6.5 reports the success rate of trials where our method converged, as well as the average error and standard deviation for the estimated variables in those cases. We observe that the method succeeds in roughly 94% of all trials. Failure cases occur when the pulsed lasers are positioned in locations that create significant occlusions and shadows.

% Success cases	Depth error		Pulsed laser localization error		Clock offset error	
	Average [mm]	Std. [mm]	Average [mm]	Std. [mm]	Average [mm]	std. [mm]
93.67	3.5	4.3	16.2	11.2	9.4	8.9

Table 6.5: Simulation success rate and accuracy.

6.8.3 3D reconstruction accuracy

We assess the impact of pulse-delay errors on 3D reconstruction, pulsed laser localization, and clock offset accuracy. To do this, we perturb the ground truth simulated pulse-delay maps by adding zero-mean Gaussian noise with standard deviations ranging from 0 ns to 33 ns, as determined from the analysis in Section 6.7.5, which correspond to depths errors of 0 mm to 1000 mm. Then, we run geometric optimization (Section 6.6) to recover depth, pulsed laser 3D locations and sync offsets. Table 6.6 summarizes the average depth error, pulsed-laser localization error, and clock offset error across all 6 simulated scenes. Thanks to the adaptive loss function and the overdetermined nature of our system of equations, our geometric optimization framework effectively handles significant pulse-delay errors, even those in the range of hundreds of millimeters. As noted in Section 6.7.5, such pulse-delay errors are primarily caused by frequency estimation errors of hundreds of millihertz. However, as demonstrated in our real-world experiments (Section 6.7, laser source discovery and frequency estimation), our frequency estimation errors are much smaller, ensuring robust and accurate performance in practical scenarios.

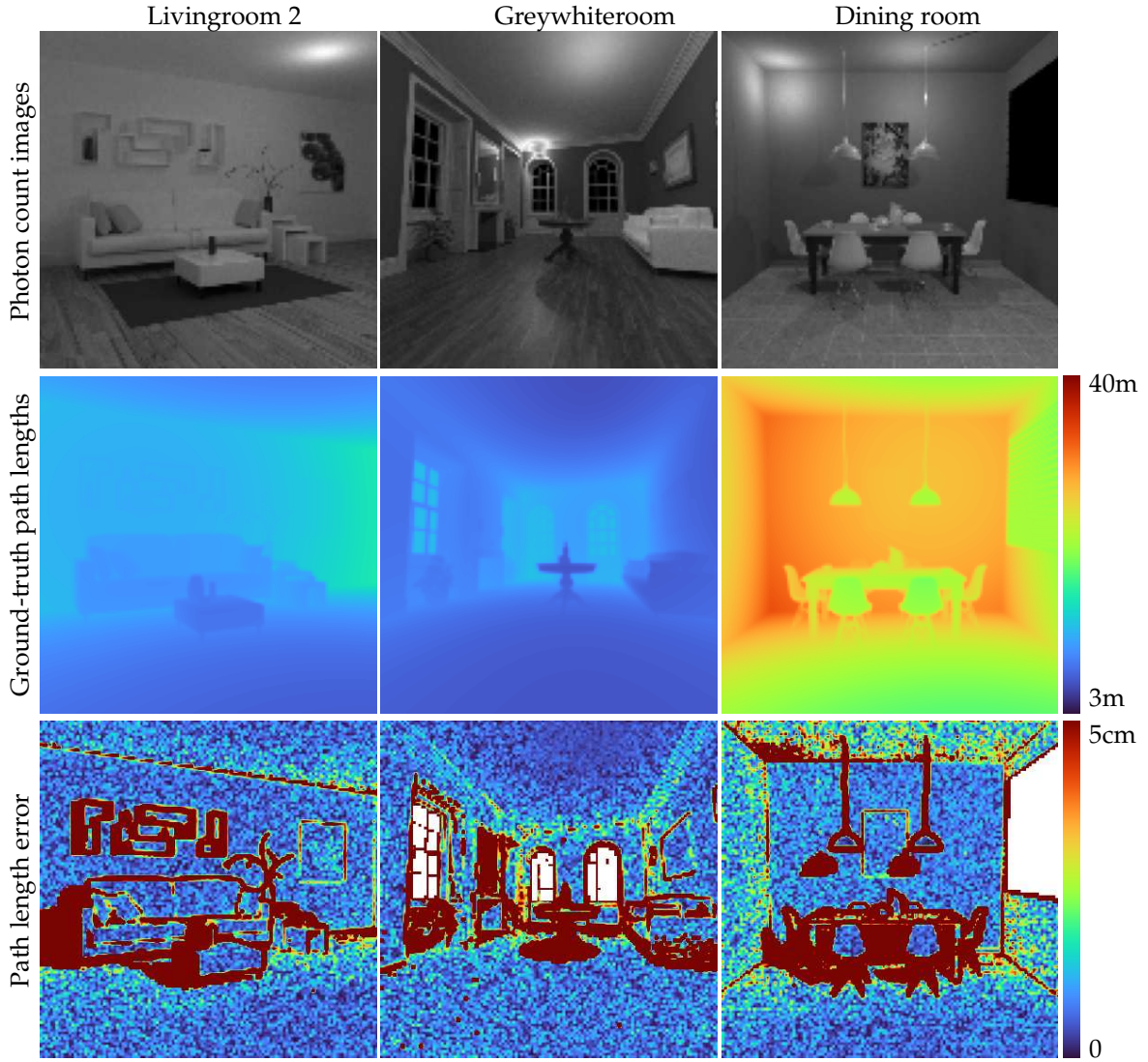


Figure 6.8.1: Path length estimation errors in livingroom2, greywhiteroom and dining room scenes. **Top:** Intensity images (gamma-corrected). **Middle:** Ground-truth path lengths from rendering. **Bottom:** Path length estimation error against ground truth.

6.8.4 Two-laser detection test

We perform simulations to determine the minimum separation between resolvable laser repetition frequencies.

Flux function. We define the flux function to be the sum of two laser pulse trains with pulse repetition frequencies of 10 MHz and $10 + \Delta f$ MHz, respectively. We model the laser pulse as a Gaussian with 110 ps FWHM and set the amplitude of each laser pulse to be 10 kilophotons per second. The pulse delay of each laser pulse train was chosen uniformly randomly from $[0, 1/f]$ where f is its pulse repetition frequency. We use the photon timestamp generation of pulsed laser trains procedure described in Section 6.8.1 to generate timestamp stream.

Experimental Procedure. We vary the frequency separation Δf from $0.2/t_{\text{exp}}$ to $4/t_{\text{exp}}$ in increments of $0.2/t_{\text{exp}}$ and use exposure times t_{exp} of 1 second, 100 milliseconds, and 10 milliseconds.

Avg. pulse-delay error $\times c$	Avg. depth error	Average source localization error	Avg. clock offset error $\times c$
0	0.38mm	0.61mm	0.32mm
5mm	1.4mm	2.4mm	1.6mm
1 cm	2.7mm	0.82mm	2.2mm
5 cm	13mm	6.0mm	2.4mm
10 cm	29.8mm	41.7mm	16.7mm
50 cm	304mm	607mm	2153mm

Table 6.6: Impact of pulse-delay estimation error in passive 3D reconstruction accuracy.

For each Δf and t_{exp} , we generated 50 timestamp streams, resulting in 1000 random trials in total. We run Algorithms 6.2—6.6 on each timestamp stream for frequency detection and consider a repetition frequency to be detected if our method estimates a frequency within $0.01/t_{\text{exp}}$ of the corresponding frequency. For each Δf and t_{exp} , we compute the two-frequency detection probability across all trials.

Figure 6.8.2 shows the simulation results.

Discussion. We observe that when the frequency separation is less than or equal to $1.4/t_{\text{exp}}$, our method consistently fails to detect both frequencies. Additionally, our simulations suggest that there is a phase transition where the detection probability goes from 0% to 100% and the duration of this phase transition decreases as the exposure time increases, likely due to the increased accuracy in the CFAR frequency detector due to an increase in the photon count. These observations suggest that our frequency detection method has a minimum two-frequency resolution between $1.4/t_{\text{exp}}$ to $1.6/t_{\text{exp}}$ as we vary Δf in steps of $0.2/t_{\text{exp}}$. We also corroborate this limit using the Rayleigh criterion limit [199]. Since the Rayleigh criterion is defined for Airy disks and not a sinc, we fit a sinc to the Airy disk by matching the FWHMs of both functions.⁶ Doing so yielded a minimum frequency separation of $1.43/t_{\text{exp}}$, which is consistent with our observed limit of $1.4/t_{\text{exp}}$ to $1.6/t_{\text{exp}}$. This suggests that we may be able to improve our frequency detection method even further to resolve even smaller frequency separation. Furthermore, since the width of the sinc lobes depends only on the exposure time, we can potentially leverage higher-order harmonics to identify laser pulse repetition frequencies with much smaller frequency separation.

We also observe that our detection method has an increased probability of detecting one of the two laser frequencies when the frequency separation is $0.4/t_{\text{exp}}$ or less. For frequency separations less than $1/t_{\text{exp}}$, the two sinc lobes of the laser frequency appear as one sinc lobe centred at $10 + \Delta f/2$, the mean of the two laser frequencies. In this case, while we are able to localize laser frequencies, they will be roughly $\Delta f/2$ away from the peak of the sinc lobe and therefore removed by our frequency detection method. However, as $\Delta f \rightarrow 0$, the laser frequencies get closer to the peak of the sinc lobe, allowing our method to detect them, which explains the increase in successful trials.

⁶We found that the Airy disk $(2kJ_1(x/k)/x)^2 \approx \text{sinc}(x/t_{\text{exp}})$ around their lobes when $k = 3.733$; J_1 is the Bessel function of the first kind of order one. Under Rayleigh criterion, $\Delta f \approx 3.8317k = 1.43/t_{\text{exp}}$ since 3.832 is the location of the first zero of J_1 .

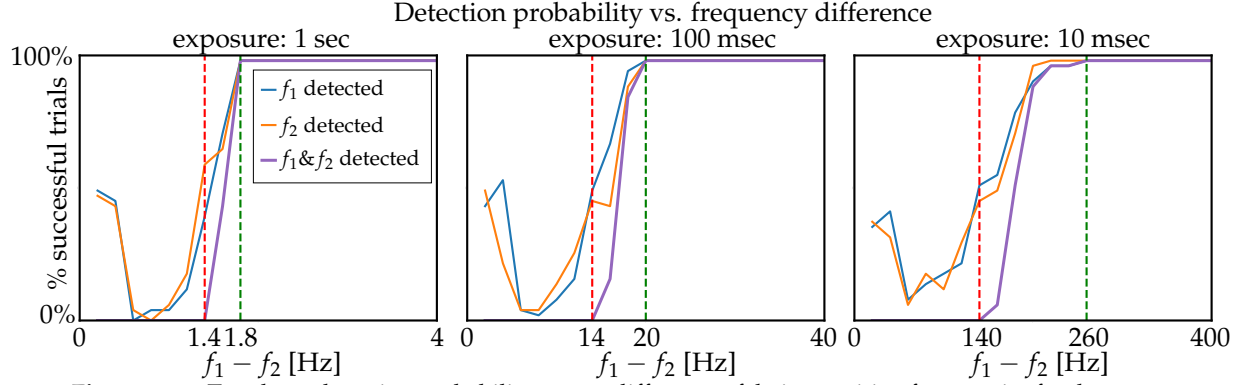


Figure 6.8.2: Two-laser detection probability versus difference of their repetition frequencies for three exposure times.

6.9 Limitations

Our path length constraint, Eq. 6.6.1, does not currently account for phase wrapping that occurs in the scene. This means that the pulse repetition frequencies of the lasers limit the maximum recoverable distance of our method. In principle, the asynchrony between laser sources should allow us to disambiguate phase wrapping, just as two co-prime frequencies can increase depth ranges [272]. Intuitively, it is easy to see that phase wrapping induces a countable set of depths consistent with the path length constraint for each laser. These laser-specific countable sets must intersect at the correct depth, and their asynchrony means that it is very unlikely for them to intersect exactly anywhere else. Incorporating phase wrapping into the loss function would be useful in increasing maximum depth ranges of our method.

Our reliance on the harmonic structure in the spectrum also assumed a very specific emission profile from the laser sources. They assume that the lasers emit short pulses. However, there are systems such as continuous-wave time-of-flight cameras [18, 17] that emit either a sinusoidal or a square wave; in these cases, localizing the frequency accurately can be tricky because either they don't have any higher order harmonics (in the case of sinusoidal waves) or many harmonics are 0 (in the case of square waves, where even-order harmonics are zero), both of which could cause our harmonic hopping algorithm to stop early.

6.10 Concluding remarks

Our work represents a small, initial step toward leveraging the harmonic structure of pulsed laser signals for 3D vision. While our preliminary results are promising, pulsed lasers are used in a far broader range of settings than the one considered here, including beam-scanning lidar, visible-light communication, and many more. Developing methods that tackle this broader range of signals is challenging because the structure of those signals is far more complex than the case considered here. Additionally, 2D SPAD cameras are still in their infancy, presenting a technological barrier to deployment of our opportunistic 3D sensing approach in the wild. Nevertheless, the specific capabilities we developed in this work—hardware-free computational synchronization, repetition-frequency estimation, low-SBR single-photon sensing, transient imaging with many asynchronous lasers, *etc.*—are directly applicable to existing single-pixel or low-pixel-counts SPADs, as well as

active sensing systems that employ pulsed lasers.

7

Conclusions

We are reaching a tipping point where new emerging sensors provide viable alternatives to conventional imaging, and new computational/algorithmic advances are breaking new grounds on imaging capabilities. In this thesis, I have looked at one facet of this, namely the problem of imaging a scene across varying timescales. Specifically, I developed several frameworks that allow emerging sensors to imaging across a discrete set of timescales as well as a continuous set of timescales . Furthermore, I introduced a method for identifying signals at relevant timescales of interest. I briefly summarize the key points of each chapter in this thesis.

Chapter 3 discussed my work on the C2B camera which allowed video acquisition across two different timescales simultaneously. Here, I formalized the image formation model and formulated the control of camera using a 3D tensor, dictating how the camera operates across the two timescales. Additionally, I derived novel tensor configurations that improved the SNR of the captured images. I then applied my framework to various imaging scenarios, wherein the finer time scale was used to multiplex images under varying illumination conditions and the coarser time scale was used to capture illumination mosaics that combine to form conventional video frames. My method demonstrated the ability to perform one-shot photometric stereo for 3D surface normal recovery as well as one-shot structured light triangulation at real time rates, ~ 20 frames a second.

In Chapter 4, I demonstrated a new regime of imaging called the passive ultra-wideband imaging regime using SPAD-based cameras, wherein the captured video can zoom into any timescale from seconds down to picoseconds, revealing different events. We empirically demonstrate this capability to record an ultra-wideband video of a spinning fan illuminated with transient laser pulses and recover a video from a non-line-of-sight configuration.

Chapter 5 introduced the flux probing theory, a novel framework for imaging across a continuous spectrum of timescales from raw photon timestamps which enabled passive ultra-wideband imaging. The flux probing theory described the relationship between the stream of absolute timestamps and projections of the flux function onto different basis functions. Using this theory, I showed that it was possible to directly probe the frequency content of the flux function, allowing a continuous reconstruction of the flux function.

Finally, in Chapter 6, I considered the problem of detecting asynchronous signals that occurred across different timescales. I introduced a frequency detection method for detecting periodic sources in a scene as well as harmonic probing that achieved much better frequency localization than our initial flux probing theory predicted. We then demonstrated how this capability can be used to capture transient videos of multiple, asynchronous laser pulses propagating through a single scene as well as enable a new form of 3D lidar, called passive ToF, using the asynchronous laser pulses in room scale scenes.

7.1 Limitations

Despite the state-of-the-art capabilities demonstrated in this thesis, there are still many limitations from a practical and theoretical perspective.

On the practical side, one main limitation lies with the hardware devices themselves. The information throughput from single-photon devices such as SPADs is extremely high [80, 221]. Even for one pixel, a SPAD sensor records tens to hundreds of thousands of photon timestamps a second compared to a conventional camera operating at 30 fps that outputs 30 intensity values a second per pixel. This is over 4 orders of magnitude in the size of the data and is further exacerbated by the fact that timestamps require more bits to represent them than intensity values. Additionally, efficient compression of the timestamps has not been as well studied as image compression. This means that there is a lot of bandwidth requirements needed to timestamp photons. Furthermore, this problem is compounded when we move to pixel arrays that may have thousands to millions of pixels, all outputting asynchronous streams of timestamps.

On the theoretical side, there are a few drawbacks to the current approach. For one, we assume some periodicity in the scene (although there is some evidence from our MGM video experiment that we may be able to recover aperiodic flux functions); our reliance on periodicity allows us to recover events that may be missed due to the dead time, but it also means that we cannot recover the flux of one-off transient events. Additionally, our criterion for selecting frequencies is currently crude and conservative. We try to minimize the amount of false positives, i.e. noise frequencies detected as signal, which potentially causes many false negatives where signal frequencies are discarded as noise. Finally, there are additional sensor non-idealities that affect the statistics of photon detections that we have not accounted for such as after pulsing, where residual electrical charges in the sensor substrate cause photon detections even in the absence of incident photons.

7.2 Future Directions

There are many new exciting opportunities for research that have emerged from the work presented in this thesis. In this section, I will outline some of these possibilities, which not only build upon the current results but also pave the way forward for novel applications and interdisciplinary explorations.

While we focused on the coded two-bucket camera and SPADs in this thesis, the landscape of emerging sensor technology opens up many opportunities for imaging across different timescales. For instance, event cameras [142, 122] and spiking cameras [283, 284] output data in the form of event streams, which can be modelled using point processes, making them a perfect fit for the flux theory proposed in Chapter 5. The application of flux probing to these event streams could enable new ways to enhance the capabilities of these devices. Additionally, event cameras are commercially viable, which means that integrating our flux probing theory with these cameras could lead to ultra-wideband imaging capabilities that is readily available.

Our current approach of using the Fourier basis function for flux reconstruction is quite crude as we risk false negatives due to missing frequencies and we need to sweep across a large spectrum of frequencies. One solution is to consider not just the null hypothesis and create an analytical model

for signal frequencies in order to use likelihood ratio tests which are known to be the uniformly most powerful (UMP) test [162]. Another potential approach is to use an optimization-based approach to select frequencies based on an objective function designed to balance the likelihood of false positives with the likelihood of false negatives. Furthermore, it may be possible to leverage machine learning methods to narrow the set of frequencies to probe to significantly reduce the search space. We have already seen the benefits of AlphaGo [229] and AlphaZero [230] in pruning the search space for games. This is an idea amenable to the case of reducing the search space for frequencies of interest. We may even be able to employ machine learning methods to directly estimate the set of relevant frequencies as well as their magnitudes or directly recover the flux from photon streams.

In this thesis, we mostly focused on recovering the flux function of the scene as well as applications in 3D estimation. However, the implications of passive ultra-wideband imaging extend beyond these areas. A promising direction lies in exploring how passive ultra-wideband imaging can improve downstream computer vision tasks [57]. Given that passive ultra-wideband imaging allows us to observe the world at different timescales and speeds, there is a wealth of information here to leverage for scene understanding, action recognition, or frame prediction. For example, different timescales can potentially reveal distinct visual cues that are complementary for understanding the dynamics of scene.

Beyond computer vision, the concept of passive ultra-wideband imaging has the potential to impact other fields. For example, in astronomy, we can apply flux probing and SPADs for analyzing light from distant stars or galaxies [51, 183], which can provide new insights and understanding to the behaviour of planetary bodies. In biology, fluorescence lifetime imaging (FLIM) [43, 285] use a laser pulse synced to single-photon detectors to visualize cells and cellular structure with higher fidelity than conventional microscopes. The ability to demultiplex asynchronous laser signals that we demonstrated in our opportunistic single-photon time-of-flight framework (Chapter 6) is useful for FLIM by multiplexing asynchronous lasers to simultaneously study the fluorescence lifetime across spectral energies, providing additional dimensions of information.

Finally, I look forward to seeing extensions of the flux probing theory to probe other properties of light such as polarization states [70], that have not been considered at all in this thesis. Probing other properties of light can have tremendous implications for optical quantum computing [168, 81] where photons and their polarization states can be treated as qubits.

A Coded Two-Bucket Camera

A.1 Proofs and Derivations

A.1.1 Gaussian approximation of bucket-ratio random variables

Proposition A.1

The random variables r and \hat{r} can be approximated by Gaussian random variable whose means and variances are given by:

$$\mu_r = \frac{j}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad , \quad (\text{A.1.1})$$

and

$$\mu_{\hat{r}} = \frac{\hat{j}}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad , \quad (\text{A.1.2})$$

where j and \hat{j} are the noiseless bucket images and σ^2 is the variance of the noise.

Proof. Consider the bucket ratios r and \hat{r} at frame F and pixel p , which are defined by combining bucket-1 and bucket-0 image intensities $\mathbf{i}^p[f]$ and $\hat{\mathbf{i}}^p[f]$ according to Eq. (3.2.5) from Section 3.2. For notational simplicity, we denote $\mathbf{i}^p[f]$ and $\hat{\mathbf{i}}^p[f]$ with i and \hat{i} respectively. We model the two-bucket intensities i and \hat{i} as having zero-mean Gaussian image noise with variance σ^2 :

$$i = j + e \quad , \quad \hat{i} = \hat{j} + \hat{e} \quad , \quad (\text{A.1.3})$$

where j, \hat{j} are the noiseless bucket-1 and bucket-0 intensities and e, \hat{e} are zero-mean Gaussian random variables with variance σ^2 . In order to apply Lemma A.1 to r and \hat{r} , we require $\frac{\sigma_{i+\hat{i}}}{j+\hat{j}} < 0.22$. This effectively amounts to requiring the SNR to be at least 5 which is an appropriate assumption for most imaging systems. Now we can apply Lemma A.1 with $k = 1$ to derive the approximate means and variances for r and \hat{r} . We first show the approximations for r :

$$\mu_r = \text{mean}\left(\frac{i}{i + \hat{i}}\right) = \frac{j}{j + \hat{j}} \quad , \quad \text{var}(r) = \text{var}\left(\frac{i}{i + \hat{i}}\right) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad . \quad (\text{A.1.4})$$

We now show the approximations for \hat{r} :

$$\mu_{\hat{r}} = \text{mean}\left(\frac{\hat{i}}{i + \hat{i}}\right) = \frac{\hat{j}}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \text{var}\left(\frac{\hat{i}}{i + \hat{i}}\right) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad . \quad (\text{A.1.5})$$

From Eq. (A.1.4) and Eq. (A.1.5), we can see that the approximate variance for both r and \hat{r} are the same. Furthermore, they are proportional to $\frac{1}{(j + \hat{j})^2}$. This means that the approximate standard deviation for both of them is proportional to $\frac{1}{j + \hat{j}}$. Since j and \hat{j} are linear function of the albedo, the approximate standard deviation for both r and \hat{r} is weakly dependence on albedo. \square

Lemma A.1

If X and Y are normally distributed with means μ_X, μ_Y respectively and variances $\sigma^2, k\sigma^2$ and $\frac{\sigma_{X+Y}}{\mu_{X+Y}} < 0.22$, then $\frac{X}{X+Y}$ can be approximated as a Gaussian random variable with the following mean and variance:

$$\text{mean}\left(\frac{X}{X+Y}\right) = \frac{\mu_X}{\mu_X + \mu_Y} \quad , \quad \text{var}\left(\frac{X}{X+Y}\right) = \frac{k\sigma^2}{(k+1)(\mu_X + \mu_Y)^2} . \quad (\text{A.1.6})$$

Proof. Let X and Y be Gaussian random variables with means μ_X, μ_Y respectively and variances $\sigma^2, k\sigma^2$ respectively where k is the ratio between the variances of X and Y . Furthermore, let us assume that $\frac{\sigma_{X+Y}}{\mu_{X+Y}} < 0.22$. First, we use the observation from [68] that $kX - Y$ and $X + Y$ have Pearson correlation coefficient ρ equal to 0:

$$\rho = \frac{\text{cov}(kX - Y, X + Y)}{\sigma_{kX-Y}\sigma_{X+Y}} = \frac{\text{cov}(kX, X) - \text{cov}(Y, Y)}{\sigma_{kX-Y}\sigma_{X+Y}} = \frac{k\text{var}(X) - \text{var}(Y)}{\sigma_{kX-Y}\sigma_{X+Y}} . \quad (\text{A.1.7})$$

Since $\text{var}(X) = \sigma^2$ and $\text{var}(Y) = k\sigma^2$, the numerator on the right side of Eq. (A.1.7), and therefore ρ , is equal to 0. Since we assume $\frac{\sigma_{X+Y}}{\mu_{X+Y}} < 0.22$, Table 1 from [175] tells us that we can reasonably approximate the ratio of two uncorrelated Gaussian random variables as a normal random variable with the following parameters:

$$\text{mean}\left(\frac{kX - Y}{X + Y}\right) = \frac{\mu_{kX-Y}}{\mu_{X+Y}} \quad , \quad \text{var}\left(\frac{kX - Y}{X + Y}\right) = \frac{\text{var}(kX - Y)}{\mu_{X+Y}^2} . \quad (\text{A.1.8})$$

Now we expand Eq. (A.1.8) to simplify it:

$$\text{mean}\left(\frac{kX - Y}{X + Y}\right) = \frac{k\mu_X - \mu_Y}{\mu_X + \mu_Y} \quad , \quad \text{var}\left(\frac{kX - Y}{X + Y}\right) = \frac{k^2\sigma^2 + k\sigma^2}{(\mu_X + \mu_Y)^2} . \quad (\text{A.1.9})$$

We can use Eq. (A.1.9) to now approximate the noise of $\frac{X}{X+Y}$. To do this, we first show that $\frac{X}{X+Y}$ is a linear function of $\frac{kX-Y}{X+Y}$:

$$\frac{X}{X+Y} = \frac{kX + X}{(k+1)(X+Y)} = \frac{kX + X - Y + Y}{(k+1)(X+Y)} = \frac{1}{k+1} + \left(\frac{1}{k+1}\right) \frac{kX - Y}{X+Y} . \quad (\text{A.1.10})$$

Combining Eq. (A.1.9) and Eq. (A.1.10), we can approximate the mean of $\frac{X}{X+Y}$:

$$\text{mean}\left(\frac{X}{X+Y}\right) = \frac{1}{k+1} + \left(\frac{1}{k+1}\right) \frac{k\mu_X - \mu_Y}{\mu_X + \mu_Y} = \frac{\mu_X + \mu_Y + k\mu_X - \mu_Y}{(k+1)(\mu_X + \mu_Y)} = \frac{\mu_X}{\mu_X + \mu_Y} , \quad (\text{A.1.11})$$

which is the noiseless ratio. We can also approximate the variance of $\frac{X}{X+Y}$ using Eq. (A.1.9) and Eq. (A.1.10):

$$\text{var}\left(\frac{X}{X+Y}\right) = \frac{1}{(k+1)^2} \text{var}\left(\frac{kX - Y}{X+Y}\right) = \frac{1}{(k+1)^2} \frac{k^2\sigma^2 + k\sigma^2}{(\mu_X + \mu_Y)^2} = \frac{k\sigma^2}{(k+1)(\mu_X + \mu_Y)^2} . \quad (\text{A.1.12})$$

□

A.1.2 Derivation of Equation 3.3.4

Consider the two image intensities, i_l and $\sum_k i_k$. We can model their values as a function of the unknown \mathbf{x} (Eq. (3.3.3) from Section 3.3):

$$i_l = a\mathbf{d}_l\mathbf{x} + e_l \quad , \quad \sum_k i_k = a \sum_k \mathbf{d}_k\mathbf{x} + e_k \quad . \quad (\text{A.1.13})$$

As $\sum_k \mathbf{d}_k$ is formed by summing the rows \mathbf{d}_k of \mathbf{D} , we can express Eq. (A.1.13) as:

$$\sum_k i_k = a\mathbf{1D}\mathbf{x} + \mathbf{1e} \quad , \quad (\text{A.1.14})$$

where $\mathbf{1}$ is a 1-by- S vector of 1's and $\mathbf{1e}$ aggregates the image noises. We can derive a constraint from the ratio of i_l and $\sum_k i_k$ from the noiseless Eq. (A.1.13) and Eq. (A.1.14):

$$\frac{i_l}{\sum_k i_k} = \frac{a\mathbf{d}_l\mathbf{x}}{a\mathbf{1D}\mathbf{x}} = \frac{\mathbf{d}_l\mathbf{x}}{\mathbf{1D}\mathbf{x}} \quad . \quad (\text{A.1.15})$$

We denote $\frac{i_l}{\sum_k i_k}$ with r_l and multiply both sides with $\mathbf{1D}\mathbf{x}$. □

A.1.3 Derivation of Equation 3.3.5

Consider two image intensities, i_l and i_k . We have the following relation (Eq. (3.3.3) from Section 3.3):

$$i_l = a\mathbf{d}_l\mathbf{x} + e_l \quad , \quad i_k = a\mathbf{d}_k\mathbf{x} + e_k \quad , \quad (\text{A.1.16})$$

where e_l and e_k are the noise for i_l and i_k respectively. Solving for \mathbf{x} normally involves using noiseless Eq. (A.1.16) directly as constraints on \mathbf{x} :

$$i_l = a\mathbf{d}_l\mathbf{x} \quad , \quad i_k = a\mathbf{d}_k\mathbf{x} \quad (\text{A.1.17})$$

However, instead of treating these as separate constraints, we can take their ratio:

$$\frac{i_l}{i_k} = \frac{a\mathbf{d}_l\mathbf{x}}{a\mathbf{d}_k\mathbf{x}} = \frac{\mathbf{d}_l\mathbf{x}}{\mathbf{d}_k\mathbf{x}} \quad , \quad (\text{A.1.18})$$

and cross multiply the terms in Eq. (A.1.18) to derive Eq. (3.3.5). □

A.1.4 Proof of Observation 3.1

Since \mathbf{W} has S columns, we trivially know that $\text{rank } \mathbf{W} \leq S$. It remains to show that $\text{rank } \mathbf{W} \leq F + 1$. Since the rank of a matrix is equal to its row rank and the row rank of a matrix is determined by the dimensions spanned by its rows, we simply need to show that the rows of \mathbf{W} which are $\mathbf{c}_1^p, \dots, \mathbf{c}_F^p, \bar{\mathbf{c}}_1^p, \dots, \bar{\mathbf{c}}_F^p$ lie in a subspace V of dimension at most $F + 1$.

First, consider \mathbf{c}_1^p and $\bar{\mathbf{c}}_1^p$ which are the first row vector of \mathbf{W} and its complement. Since they are complementary, we know their sum, trivially:

$$\mathbf{c}_1^p + \bar{\mathbf{c}}_1^p = \mathbf{1} \quad , \quad (\text{A.1.19})$$

where $\mathbf{1}$ is a 1-by- S vector of all 1's. Now let V be the subspace spanned by the vectors $\mathbf{c}_1^p, \dots, \mathbf{c}_F^p$ and $\bar{\mathbf{c}}_1^p$. Using Eq. (A.1.19), we can see that for all k , $\bar{\mathbf{c}}_k^p$ must lie in V since it can be expressed as the

linear combination of vectors in V :

$$\bar{\mathbf{c}}_k^p = \mathbf{1} - \mathbf{c}_k^p = \mathbf{c}_1^p + \bar{\mathbf{c}}_1^p - \mathbf{c}_k^p . \quad (\text{A.1.20})$$

Therefore, given V , we know that $\mathbf{c}_1^p, \dots, \mathbf{c}_F^p$, and $\bar{\mathbf{c}}_1^p$ lie in V by construction and also all $\bar{\mathbf{c}}_k^p$'s lie in V . Since V is spanned by $F + 1$ vectors, we know that $\dim(V)$ is at most $F + 1$. Therefore, the rows of \mathbf{W} span a subspace with dimension at most $F + 1$, implying that $\text{rank } \mathbf{W} \leq F + 1$. \square

A.1.5 Proof of Lemma 3.1

We derive the lower bound for the MSE in 4 steps:

Step 1: Verify that the definition of the MSE used in [198] is still equal to

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \text{ for two-bucket cameras.}$$

Step 2: Derive a lower bound on $\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right]$ as a function of $\mathbf{W}'\mathbf{W}$'s largest eigenvalue λ_1 .

Step 3: Bound the value of λ_1 .

Step 4: Combine Step 2 and Step 3 results to derive the explicit lower bound.

Step 1: verification of MSE definition. Using Eq. (3.3.2) from Section 3.3, we know that the demultiplexed intensities $i_1^p \dots i_S^p$ can be computed by applying the inverse of \mathbf{W} to the $2F$ bucket intensities \mathbf{i}^p and $\hat{\mathbf{i}}^p$. As in [198], we define the MSE of \mathbf{W} as the mean variance of the demultiplexed images:

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^S \text{var}(i_k^p) . \quad (\text{A.1.21})$$

Let us denote $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ with \mathbf{A} . Let \mathbf{A}_i denote the i^{th} row of \mathbf{A} and $\mathbf{A}_i[j]$ denote the element of \mathbf{A} at row i and column j . Then we can express $\text{var}(i_k^p)$ as:

$$\text{var}(i_k^p) = \text{var} \left(\mathbf{A}_k \begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix} \right) \quad (\text{A.1.22})$$

$$= \text{var} \left(\sum_{j=1}^F \mathbf{A}_k[j] \mathbf{i}_j^p + \mathbf{A}_k[j+F] \hat{\mathbf{i}}_j^p \right) \quad (\text{A.1.23})$$

$$= \sum_{j=1}^F \mathbf{A}_k[j]^2 \text{var}(\mathbf{i}_j^p) + \mathbf{A}_k[j+F]^2 \text{var}(\hat{\mathbf{i}}_j^p) . \quad (\text{A.1.24})$$

We will assume that \mathbf{i}^p and $\hat{\mathbf{i}}^p$ have zero-mean Gaussian noise with variance σ^2 . We can now simplify Eq. (A.1.24):

$$\text{var}(i_k^p) = \sum_{j=1}^F \mathbf{A}_k[j]^2 \sigma^2 + \mathbf{A}_k[j+F]^2 \sigma^2 = \sigma^2 \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 , \quad (\text{A.1.25})$$

and use Eq. (A.1.25) in Eq. (A.1.21):

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^S \sigma^2 \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 \quad (\text{A.1.26})$$

$$= \frac{\sigma^2}{S} \sum_{k=1}^S \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 . \quad (\text{A.1.27})$$

Since the sum of squares of elements of \mathbf{A} is equal to the trace of $\mathbf{A}'\mathbf{A}$, we can substitute this back into Eq. (A.1.27):

$$\text{MSE} = \frac{\sigma^2}{S} \text{trace}(\mathbf{A}'\mathbf{A}) \quad (\text{A.1.28})$$

$$= \frac{\sigma^2}{S} \text{trace} \left[\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \right] \quad (\text{A.1.29})$$

$$= \frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{W} \right] \quad (\text{A.1.30})$$

$$= \frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \quad (\text{A.1.31})$$

Step 2: bounding the MSE in terms of λ_1 : Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_S \geq 0$ be the eigenvalues of $\mathbf{W}'\mathbf{W}$. Then we can express $\text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right]$ as the sum of its eigenvalues:

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] = \frac{\sigma^2}{S} \sum_{i=1}^S \frac{1}{\lambda_i} \quad (\text{A.1.32})$$

$$= \frac{\sigma^2}{S\lambda_1} + \frac{\sigma^2}{S} \sum_{i=2}^S \frac{1}{\lambda_i} \quad (\text{A.1.33})$$

$$= \frac{\sigma^2}{S\lambda_1} + \left(\frac{\sigma^2(S-1)}{S} \right) \left(\frac{1}{S-1} \right) \sum_{i=2}^S \frac{1}{\lambda_i} . \quad (\text{A.1.34})$$

The right term in the addition in Eq. (A.1.34) is the arithmetic mean of $\frac{1}{\lambda_i}$'s. Since the arithmetic mean is always greater or equal than the harmonic mean, we have the following inequality:

$$\left(\frac{1}{S-1} \right) \sum_{i=2}^S \frac{1}{\lambda_i} \geq \left(\frac{\sum_{i=2}^S (\frac{1}{\lambda_i})^{-1}}{S-1} \right)^{-1} = \frac{S-1}{\sum_{i=2}^S \lambda_i} . \quad (\text{A.1.35})$$

Since $\text{trace}(\mathbf{W}'\mathbf{W})$ sums over all the eigenvalues of $\mathbf{W}'\mathbf{W}$, we know that:

$$\sum_{i=2}^S \lambda_i = \text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1 . \quad (\text{A.1.36})$$

We can now combine Eq. (A.1.35) and Eq. (A.1.36) to turn Eq. (A.1.34) into an inequality:

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \geq \frac{\sigma^2}{S\lambda_1} + \sigma^2 \left(\frac{S-1}{S} \right) \frac{S-1}{\sum_{i=2}^S \lambda_i} \quad (\text{A.1.37})$$

$$= \frac{\sigma^2}{S\lambda_1} + \sigma^2 \left(\frac{S-1}{S} \right) \frac{S-1}{\text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1} . \quad (\text{A.1.38})$$

We compute $\text{trace}(\mathbf{W}'\mathbf{W})$ before proceeding. Let \mathbf{J} be the F -by- S matrix of 1's. Then we have the

following:

$$\text{trace}(\mathbf{W}'\mathbf{W}) = \text{trace}[(\mathbf{C}^p)' \mathbf{C}^p + (\overline{\mathbf{C}}^p)' \overline{\mathbf{C}}^p] \quad (\text{A.1.39})$$

$$= \text{trace}[(\mathbf{C}^p)' \mathbf{C}^p + (\mathbf{J} - \mathbf{C}^p)' (\mathbf{J} - \mathbf{C}^p)] \quad (\text{A.1.40})$$

$$= \text{trace}[(\mathbf{C}^p)' \mathbf{C}^p + \mathbf{J}'\mathbf{J} - \mathbf{J}'\mathbf{C}^p - (\mathbf{C}^p)' \mathbf{J} + (\mathbf{C}^p)' \mathbf{C}^p] \quad (\text{A.1.41})$$

$$= 2\text{trace}[(\mathbf{C}^p)' \mathbf{C}^p] + \text{trace}(\mathbf{J}'\mathbf{J}) - \text{trace}(\mathbf{J}'\mathbf{C}^p) - \text{trace}[(\mathbf{C}^p)' \mathbf{J}] \quad (\text{A.1.42})$$

$$= 2 \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j]^2 + \sum_{k=1}^F \sum_{j=1}^S 1 - \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j] - \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j] . \quad (\text{A.1.43})$$

Since \mathbf{C}^p is a binary matrix, we know that $\mathbf{C}_k^p[j]^2 = \mathbf{C}_k^p[j]$:

$$\text{trace}(\mathbf{W}'\mathbf{W}) = 2 \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j] + FS - \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j] - \sum_{k=1}^F \sum_{j=1}^S \mathbf{C}_k^p[j] \quad (\text{A.1.44})$$

$$= SF . \quad (\text{A.1.45})$$

We can now substitute Eq. (A.1.45) into Eq. (A.1.38) yielding the following:

$$\frac{\sigma^2}{S} \text{trace}[(\mathbf{W}'\mathbf{W})^{-1}] \geq \frac{\sigma^2}{S\lambda_1} + \sigma^2 \left(\frac{S-1}{S} \right) \frac{S-1}{\text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1} \quad (\text{A.1.46})$$

$$= \frac{\sigma^2}{S\lambda_1} + \sigma^2 \left(\frac{S-1}{S} \right) \frac{S-1}{SF - \lambda_1} \quad (\text{A.1.47})$$

$$= \frac{\sigma^2}{S} \left(\frac{1}{\lambda_1} + \frac{(S-1)^2}{SF - \lambda_1} \right) \quad (\text{A.1.48})$$

$$= \frac{\sigma^2}{S} \frac{SF - \lambda_1 + \lambda_1(S-1)^2}{(SF - \lambda_1)\lambda_1} \quad (\text{A.1.49})$$

$$= \frac{\sigma^2}{S} \frac{SF + \lambda_1 S^2 - 2S}{(SF - \lambda_1)\lambda_1} \quad (\text{A.1.50})$$

$$= \sigma^2 \frac{F + \lambda_1(S-2)}{(SF - \lambda_1)\lambda_1} . \quad (\text{A.1.51})$$

Step 3: bounding the value of λ_1 : We know that the squared spectral norm of \mathbf{W} is equal to λ_1 :

$$\lambda_1 = \|\mathbf{W}\|_2^2 = \sup_{x \neq 0} \frac{\|\mathbf{W}x\|_2^2}{\|x\|_2^2} = \sup_{x \neq 0} \frac{x'(\mathbf{W}'\mathbf{W})x}{x'x} . \quad (\text{A.1.52})$$

We can use $x = \mathbf{1}$ where $\mathbf{1}$ is a S -by-1 vector of all 1's on the right side of Eq. (A.1.52) as a lower bound of λ_1 :

$$\lambda_1 \geq \frac{\mathbf{1}'(\mathbf{W}'\mathbf{W})\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \frac{\sum_{k=1}^{2F} (\sum_{j=1}^S \mathbf{W}_k[j])^2}{S} . \quad (\text{A.1.53})$$

Since \mathbf{W} is defined as the concatenation of \mathbf{C}^p and $\overline{\mathbf{C}}^p$, we know that for $k \geq F$ and j :

$$\mathbf{W}_k[j] = 1 - \mathbf{W}_{k-F}[j] . \quad (\text{A.1.54})$$

We can now substitute Eq. (A.1.54) back into Eq. (A.1.53):

$$\lambda_1 \geq \frac{1}{S} \sum_{k=1}^F \left(\sum_{j=1}^S \mathbf{w}_k[j] \right)^2 + \frac{1}{S} \sum_{k=F+1}^{2F} \left(\sum_{j=1}^S 1 - \mathbf{w}_{k-F}[j] \right)^2 \quad (\text{A.1.55})$$

$$= \frac{1}{S} \sum_{k=1}^F \left(\sum_{j=1}^S \mathbf{w}_k[j] \right)^2 + \left(\sum_{j=1}^S 1 - \mathbf{w}_k[j] \right)^2 \quad (\text{A.1.56})$$

$$= \frac{1}{S} \sum_{k=1}^F \left(\sum_{j=1}^S \mathbf{w}_k[j] \right)^2 + \left(S - \sum_{j=1}^S \mathbf{w}_k[j] \right)^2 \quad (\text{A.1.57})$$

$$= \frac{1}{2S} \sum_{k=1}^F 2 \left(\sum_{j=1}^S \mathbf{w}_k[j] \right)^2 + 2 \left(S - \sum_{j=1}^S \mathbf{w}_k[j] \right)^2 . \quad (\text{A.1.58})$$

The inequality of arithmetic mean and geometric mean which states $a^2 + b^2 \geq 2ab$. This implies that $2a^2 + 2b^2 \geq a^2 + b^2 + 2ab \geq (a + b)^2$. We can use this on the summand of Eq. (A.1.58):

$$\lambda_1 \geq \frac{1}{2S} \sum_{k=1}^F \left(\sum_{j=1}^S \mathbf{w}_k[j] + S - \sum_{j=1}^S \mathbf{w}_k[j] \right)^2 \quad (\text{A.1.59})$$

$$= \frac{1}{2S} \sum_{k=1}^F S^2 \quad (\text{A.1.60})$$

$$= \frac{SF}{2} . \quad (\text{A.1.61})$$

Additionally, we will prove an upper bound on λ_1 . Let v be an eigenvector with eigenvalue λ_1 and at least one positive element and let $v[k]$ be the maximum element in v . Since $\mathbf{W}'\mathbf{W}$ only has non-negative values, we can bound λ_1 in the following way:

$$\lambda_1 v[k] = \sum_{j=1}^S (\mathbf{W}'\mathbf{W})_k[j] v[j] \leq \sum_{j=1}^S (\mathbf{W}'\mathbf{W})_k[j] v[k] . \quad (\text{A.1.62})$$

As we chose v to have at least one positive element, we know that $v[k] > 0$ so we can divide both sides of Eq. (A.1.62) with $v[k]$ and preserve the inequality:

$$\lambda_1 \leq \sum_{j=1}^S (\mathbf{W}'\mathbf{W})_k[j] . \quad (\text{A.1.63})$$

Now, we simply bound the k^{th} row sum for $\mathbf{W}'\mathbf{W}$ to bound λ_1 :

$$\sum_{j=1}^S (\mathbf{W}'\mathbf{W})_k[j] = \sum_{j=1}^S \sum_{i=1}^{2F} \mathbf{w}_i[k] \mathbf{w}_i[j] \leq \sum_{j=1}^S \sum_{i=1}^{2F} \mathbf{w}_i[j] . \quad (\text{A.1.64})$$

Since \mathbf{W} is defined as the concatenation of \mathbf{C}^p and $\overline{\mathbf{C}}^p$, it has exactly SF non-zero elements all equal to 1. The right side of Eq. (A.1.64) sums over all the elements of \mathbf{W} so we can replace it with SF :

$$\sum_{j=1}^S (\mathbf{W}'\mathbf{W})_k[j] \leq \sum_{j=1}^S \sum_{i=1}^{2F} \mathbf{w}_i[j] = SF . \quad (\text{A.1.65})$$

So the k^{th} row sum of $\mathbf{W}'\mathbf{W}$ is bounded above by SF . Combining Eq. (A.1.61) and Eq. (A.1.65) gives us the interval in which λ_1 must lie:

$$\lambda_1 \in \left[\frac{SF}{2}, SF \right] . \quad (\text{A.1.66})$$

Step 4: compute the explicit bound. Eq. (A.1.66) tells us that λ_1 has to be between $\frac{SF}{2}$ and SF . Eq. (A.1.51) expresses the lower bound of the MSE as the ratio between two functions of λ_1 . The denominator, $(SF - \lambda_1)\lambda_1$, is a simple quadratic function of λ_1 which is maximized at $\lambda_1 = \frac{SF}{2}$ in the range $[\frac{SF}{2}, SF]$. The numerator, $F + \lambda_1(S - 2)$, is a simple linear function of λ_1 that is minimized also at $\lambda_1 = \frac{SF}{2}$ in the range $[\frac{SF}{2}, SF]$. Combining this information in Eq. (A.1.51) yields:

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \geq \sigma^2 \frac{F + \lambda_1(S - 2)}{(SF - \lambda_1)\lambda_1} \quad (\text{A.1.67})$$

$$\geq \sigma^2 \frac{F + \frac{SF}{2}(S - 2)}{(SF - \frac{SF}{2})\frac{SF}{2}} \quad (\text{A.1.68})$$

$$= 2\sigma^2 \frac{2F + SF(S - 2)}{(SF)^2} \quad (\text{A.1.69})$$

$$= 2\sigma^2 \frac{1 + (S - 1)^2}{S^2 F} . \quad (\text{A.1.70})$$

In the minimal configuration, $F = S - 1$. We can substitute this into Eq. (A.1.70):

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] \geq 2\sigma^2 \frac{(S - 1)^2 + 1}{(S - 1)S^2} . \quad (\text{A.1.71})$$

□

A.1.6 Proof of Proposition 3.1

The proof proceeds in 3 steps:

Step 1: We derive a simple expression of $\mathbf{W}'\mathbf{W}$.

Step 2: We compute an explicit formula for $(\mathbf{W}'\mathbf{W})^{-1}$.

Step 3: We compute the MSE of $(\mathbf{W}'\mathbf{W})^{-1}$ using the explicit formula and show that it is equal to the lower bound in Eq. (A.1.70).

First, we clarify some notation. Let $\mathbf{1}$ be a S -by-1 vector of 1's, \mathbf{J} be the $S - 1$ -by- S matrix of 1's, \mathbf{H} be the S -by- S Hadamard matrix, and \mathbf{I}_S be the S -by- S identity matrix. By construction, $\begin{bmatrix} \mathbf{1} & \tilde{\mathbf{H}}' \end{bmatrix} = \mathbf{H}'$.

Step 1: simple expression for $\mathbf{W}'\mathbf{W}$. We begin by expanding $\mathbf{W}'\mathbf{W}$:

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} (\mathbf{C}^p)' & (\bar{\mathbf{C}}^p)' \end{bmatrix} \begin{bmatrix} \mathbf{C}^p \\ \bar{\mathbf{C}}^p \end{bmatrix} \quad (\text{A.1.72})$$

$$= (\mathbf{C}^p)' \mathbf{C}^p + (\bar{\mathbf{C}}^p)' \bar{\mathbf{C}}^p \quad (\text{A.1.73})$$

$$= (\mathbf{C}^p)' \mathbf{C}^p + (\mathbf{J} - \mathbf{C}^p)' (\mathbf{J} - \mathbf{C}^p) \quad (\text{A.1.74})$$

$$= \frac{1}{4} (\tilde{\mathbf{H}} + \mathbf{1})' (\tilde{\mathbf{H}} + \mathbf{1}) + (\mathbf{J} - \frac{1}{2} (\tilde{\mathbf{H}} + \mathbf{1}))' (\mathbf{J} - \frac{1}{2} (\tilde{\mathbf{H}} + \mathbf{1})) \quad (\text{A.1.75})$$

$$= \frac{1}{4} (\tilde{\mathbf{H}} + \mathbf{J})' (\tilde{\mathbf{H}} + \mathbf{J}) + \frac{1}{4} (\mathbf{J} - \tilde{\mathbf{H}})' (\mathbf{J} - \tilde{\mathbf{H}}) \quad (\text{A.1.76})$$

$$= \frac{1}{4} (\tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \mathbf{J}' \mathbf{J} + \tilde{\mathbf{H}}' \mathbf{J} + \mathbf{J}' \tilde{\mathbf{H}}) + \frac{1}{4} (\tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \mathbf{J}' \mathbf{J} - \tilde{\mathbf{H}}' \mathbf{J} - \mathbf{J}' \tilde{\mathbf{H}}) \quad (\text{A.1.77})$$

$$= \frac{1}{2} (\tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \mathbf{J}' \mathbf{J}) . \quad (\text{A.1.78})$$

Since \mathbf{J} is a $S - 1$ by S matrix of 1's, $\mathbf{J}'\mathbf{J}$ is a S by S matrix of $S - 1$'s which can be expressed as $(S - 1)\mathbf{1}\mathbf{1}'$:

$$\mathbf{W}'\mathbf{W} = \frac{1}{2}(\tilde{\mathbf{H}}'\tilde{\mathbf{H}} + (S - 1)\mathbf{1}\mathbf{1}') \quad (\text{A.1.79})$$

$$= \frac{1}{2}(\tilde{\mathbf{H}}'\tilde{\mathbf{H}} + \mathbf{1}\mathbf{1}' + (S - 2)\mathbf{1}\mathbf{1}') \quad (\text{A.1.80})$$

$$= \frac{1}{2}\left(\begin{bmatrix} \mathbf{1} & \tilde{\mathbf{H}}' \end{bmatrix} \begin{bmatrix} \mathbf{1}' \\ \tilde{\mathbf{H}} \end{bmatrix} + (S - 2)\mathbf{1}\mathbf{1}'\right) \quad (\text{A.1.81})$$

$$= \frac{1}{2}(\mathbf{H}'\mathbf{H} + (S - 2)\mathbf{1}\mathbf{1}') . \quad (\text{A.1.82})$$

$\mathbf{H}'\mathbf{H}$ is easy to compute as it is orthogonal. In particular, \mathbf{H} satisfies the condition:

$$\mathbf{H}'\mathbf{H} = S\mathbf{I}_S . \quad (\text{A.1.83})$$

We substitute Eq. (A.1.83) into Eq. (A.1.82):

$$\mathbf{W}'\mathbf{W} = \frac{S}{2}\mathbf{I}_S + \frac{S - 2}{2}\mathbf{1}\mathbf{1}' . \quad (\text{A.1.84})$$

Step 2: explicit formula for $(\mathbf{W}'\mathbf{W})^{-1}$. The form $\mathbf{W}'\mathbf{W}$ takes in Eq. (A.1.82) is very convenient as we can compute its inverse using the Sherman-Morrison formula which states:

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + v'A^{-1}u} , \quad (\text{A.1.85})$$

where A is a matrix and u, v are vectors. Since we will be using a special case of the formula where A is the identity matrix and u, v are both equal to $\sqrt{\frac{S-2}{S}}\mathbf{1}$, the formula becomes even simpler:

$$(A + uv')^{-1} = (\mathbf{I}_S + \frac{S - 2}{S}\mathbf{1}\mathbf{1}')^{-1} \quad (\text{A.1.86})$$

$$= \mathbf{I}_S - \frac{\frac{S-2}{S}\mathbf{1}\mathbf{1}'}{1 + \frac{S-2}{S}\mathbf{1}\mathbf{1}} \quad (\text{A.1.87})$$

$$= \mathbf{I}_S - \frac{\frac{S-2}{S}\mathbf{1}\mathbf{1}'}{1 + S - 2} \quad (\text{A.1.88})$$

$$= \mathbf{I}_S - \frac{S - 2}{(S - 1)S}\mathbf{1}\mathbf{1}' . \quad (\text{A.1.89})$$

We can use Eq. (A.1.89) to compute $(\mathbf{W}'\mathbf{W})^{-1}$:

$$(\mathbf{W}'\mathbf{W})^{-1} = \left(\frac{S}{2}\mathbf{I}_S + \frac{S - 2}{2}\mathbf{1}\mathbf{1}'\right)^{-1} \quad (\text{A.1.90})$$

$$= \frac{2}{S}(\mathbf{I}_S + \frac{S - 2}{S}\mathbf{1}\mathbf{1}')^{-1} \quad (\text{A.1.91})$$

$$= \frac{2}{S}\left(\mathbf{I}_S - \frac{S - 2}{(S - 1)S}\mathbf{1}\mathbf{1}'\right) . \quad (\text{A.1.92})$$

Step 3: compute MSE for \mathbf{W} . Using Eq. (A.1.92), we have:

$$\frac{\sigma^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \right] = \frac{\sigma^2}{S} \text{trace} \left[\frac{2}{S} \left(\mathbf{I}_S - \frac{S-2}{(S-1)S} \mathbf{1}\mathbf{1}' \right) \right] \quad (\text{A.1.93})$$

$$= \frac{2\sigma^2}{S^2} \left[\text{trace}(\mathbf{I}_S) - \text{trace} \left(\frac{S-2}{(S-1)S} \mathbf{1}\mathbf{1}' \right) \right] \quad (\text{A.1.94})$$

$$= \frac{2\sigma^2}{S^2} \left[S - \frac{S-2}{(S-1)S} S \right] \quad (\text{A.1.95})$$

$$= \frac{2\sigma^2}{S^2} \frac{S^2 - 2S + 2}{S-1} \quad (\text{A.1.96})$$

$$= 2\sigma^2 \frac{(S-1)^2 + 1}{(S-1)S^2}. \quad (\text{A.1.97})$$

□

A.1.7 Generalization of Equation 3.3.2 and Lemma 3.1 to bucket ratios

The results of Eq. (3.3.2) and Lemma 3.1 apply to raw intensities captured by C2B frames. Since the BRD method relies on demosaicing and then demultiplexing ratio mosaics, it is not obvious that we can use the same \mathbf{W} to demultiplexing the ratio mosaics. Furthermore, \mathbf{W} might not even be optimal for demultiplexing ratio mosaics. This is because the MSE is intensity-based and defined using the noise model of the bucket images which differs from the noise model of the bucket ratios.

Proposition A.2 below generalizes Eq. (3.3.2) from Section 3.2 and shows that we can use the same \mathbf{W} to demultiplex the bucket ratios to recover the illumination ratios, defined in Eq. (3.3.4) from Section 3.4. Additionally, Proposition A.3 generalizes Lemma 3.1 and shows that we can define an approximate MSE using the normal approximation of the demultiplexed illumination ratios which can be used to find optimal multiplexing matrices \mathbf{W} . We exhaustively search for matrices that minimized the approximate MSE and verified that \mathbf{W} 's, constructed from optimal \mathbf{C}^p 's in Table 3.2 from Section 3.4, that minimize the intensity-based MSE also minimize the approximate MSE.

Proposition A.2

If i_1, \dots, i_S are demultiplexed image intensities obtained from bucket images \mathbf{i} and $\hat{\mathbf{i}}$ and multiplexing matrix \mathbf{W} , then r_1, \dots, r_S are demultiplexed illumination ratios obtained from bucket ratios \mathbf{r} and $\hat{\mathbf{r}}$ and multiplexing matrix \mathbf{W} :

$$\begin{bmatrix} r_1 \\ \vdots \\ r_S \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}' \begin{bmatrix} \mathbf{r} \\ \hat{\mathbf{r}} \end{bmatrix}. \quad (\text{A.1.98})$$

Proof. Given F frames, we have $2F$ images, \mathbf{i} , the bucket-1 images, and $\hat{\mathbf{i}}$, the bucket-0 images. From \mathbf{i} and $\hat{\mathbf{i}}$, we can compute ratios \mathbf{r} and $\hat{\mathbf{r}}$ using Eq. (3.2.5) from Section 3.2:

$$\mathbf{r} = \begin{bmatrix} \frac{\mathbf{i}[1]}{\mathbf{i}[1] + \hat{\mathbf{i}}[1]} \\ \vdots \\ \frac{\mathbf{i}[F]}{\mathbf{i}[F] + \hat{\mathbf{i}}[F]} \end{bmatrix}, \quad \hat{\mathbf{r}} = \begin{bmatrix} \frac{\hat{\mathbf{i}}[1]}{\mathbf{i}[1] + \hat{\mathbf{i}}[1]} \\ \vdots \\ \frac{\hat{\mathbf{i}}[F]}{\mathbf{i}[F] + \hat{\mathbf{i}}[F]} \end{bmatrix}. \quad (\text{A.1.99})$$

For a given frame f , $\mathbf{i}[f] + \hat{\mathbf{i}}[f]$ is equal to $\sum_k i_k$ (Eq. 3.3.1 from Section 3.3):

$$\mathbf{i}[f] + \hat{\mathbf{i}}[f] = \mathbf{c}_f \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} + \bar{\mathbf{c}}_f \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \mathbf{1} \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \sum_k i_k \mathbf{1}, \quad (\text{A.1.100})$$

where $\mathbf{1}$ is a 1-by- S vector of 1's. Equation (A.1.100) tells us \mathbf{r} and $\hat{\mathbf{r}}$ are equal to \mathbf{i} and $\hat{\mathbf{i}}$ scaled by $1/\sum_k i_k$. Therefore, demultiplexing the bucket ratios will allow us to recover S demultiplexed intensities i_1, \dots, i_S also scaled by $1/\sum_k i_k$:

$$\begin{bmatrix} \frac{i_1}{\sum_k i_k} \\ \vdots \\ \frac{i_S}{\sum_k i_k} \end{bmatrix} = \frac{1}{\sum_k i_k} \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \frac{1}{\sum_k i_k} (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}' \begin{bmatrix} \mathbf{i} \\ \hat{\mathbf{i}} \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}' \begin{bmatrix} \mathbf{r} \\ \hat{\mathbf{r}} \end{bmatrix}. \quad (\text{A.1.101})$$

Since we define the illumination ratio r_l as $i_l/\sum_k i_k$, the left side of Eq. (A.1.101) is exactly the vector of the S illumination ratios. \square

Proposition A.3

The approximate MSE of \mathbf{W} for demultiplexing bucket ratios can be defined as:

$$\text{MSE} = \frac{\lambda^2}{S} \text{trace} \left[(\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}' \begin{bmatrix} \mathbf{I}_F & -\mathbf{I}_F \\ -\mathbf{I}_F & \mathbf{I}_F \end{bmatrix} \mathbf{W} (\mathbf{W}'\mathbf{W})^{-1} \right], \quad (\text{A.1.102})$$

where λ^2 is the approximate variance of the bucket ratios.

Proof. Similarly to what we did in the proof of Lemma 3.1, we define the MSE as the mean variance of the demultiplexed illumination ratios:

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^S \text{var}(r_k). \quad (\text{A.1.103})$$

Let us denote $(\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'$ with \mathbf{A} . Let \mathbf{A}_i denote the i^{th} row of \mathbf{A} and $\mathbf{A}_i[j]$ denote the element of \mathbf{A} at row i and column j . Let $\mathbf{r}[j]$ and $\hat{\mathbf{r}}[j]$ denote the j^{th} element in \mathbf{r} and $\hat{\mathbf{r}}$, respectively. Using Proposition A.2, we can express $\text{var}(r_k)$ as:

$$\text{var}(r_k) = \text{var} \left(\mathbf{A}_k \begin{bmatrix} \mathbf{r} \\ \hat{\mathbf{r}} \end{bmatrix} \right) = \text{var} \left(\sum_{j=1}^F \mathbf{A}_k[j] \mathbf{r}[j] + \mathbf{A}_k[j+F] \hat{\mathbf{r}}[j] \right). \quad (\text{A.1.104})$$

Notice that $\mathbf{r}[j]$ and $\hat{\mathbf{r}}[j]$ sum up to 1:

$$\hat{\mathbf{r}}[j] = \frac{\hat{\mathbf{i}}[j]}{\mathbf{i}[j] + \hat{\mathbf{i}}[j]} = \frac{\hat{\mathbf{i}}[j] + \mathbf{i}[j] - \mathbf{i}[j]}{\mathbf{i}[j] + \hat{\mathbf{i}}[j]} = 1 - \frac{\mathbf{i}[j]}{\mathbf{i}[j] + \hat{\mathbf{i}}[j]} = 1 - \mathbf{r}[j]. \quad (\text{A.1.105})$$

Now we can substitute Eq. (A.1.105) into Eq. (A.1.104):

$$\text{var}(r_k) = \text{var}\left(\sum_{j=1}^F \mathbf{A}_k[j] \mathbf{r}[j] + \mathbf{A}_k[j+F](1 - \mathbf{r}[j])\right) \quad (\text{A.1.106})$$

$$= \text{var}\left(\sum_{j=1}^F (\mathbf{A}_k[j] - \mathbf{A}_k[j+F]) \mathbf{r}[j]\right) \quad (\text{A.1.107})$$

$$= \sum_{j=1}^F (\mathbf{A}_k[j] - \mathbf{A}_k[j+F])^2 \text{var}(\mathbf{r}[j]) . \quad (\text{A.1.108})$$

From Eq. (A.1.108), we see that the definition of the MSE depends on $\text{var}(\mathbf{r}[j])$. Using Lemma A.2, we can denote the approximate variance of $\mathbf{r}[j]$ as λ^2 :

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^S \sum_{j=1}^F (\mathbf{A}_k[j] - \mathbf{A}_k[j+F])^2 \lambda^2 \quad (\text{A.1.109})$$

$$= \frac{\lambda^2}{S} \sum_{k=1}^S \sum_{j=1}^F (\mathbf{A}_k[j] - \mathbf{A}_k[j+F])^2 . \quad (\text{A.1.110})$$

To simplify Eq. (A.1.110), we first define a matrix whose elements are equal to $\mathbf{A}_k[j] - \mathbf{A}_k[j+F]$. Since we are always subtracting elements exactly F columns apart and in the same row. We know that this matrix is difference between the submatrix formed by taking the first F columns of \mathbf{A} and the submatrix formed by taking the last F columns of \mathbf{A} . We will denote these submatrices with \mathbf{A}^L and \mathbf{A}^R respectively. We can compute them as follows:

$$\mathbf{A}^L = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ 0 \end{bmatrix} \quad \mathbf{A}^R = \mathbf{A} \begin{bmatrix} 0 \\ -\mathbf{I}_F \end{bmatrix} , \quad (\text{A.1.111})$$

where \mathbf{I}_F is the F -by- F identity matrix. Taking their difference gives us the following expression:

$$\mathbf{A}^L - \mathbf{A}^R = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ 0 \end{bmatrix} - \mathbf{A} \begin{bmatrix} 0 \\ -\mathbf{I}_F \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ -\mathbf{I}_F \end{bmatrix} , \quad (\text{A.1.112})$$

where \mathbf{I}_F is the F -by- F identity matrix. Now summing all the squared elements of $\mathbf{A}^L - \mathbf{A}^R$ equates to computing the trace of $(\mathbf{A}^L - \mathbf{A}^R)'(\mathbf{A}^L - \mathbf{A}^R)$. Using this fact and substituting Eq. (A.1.112) into Eq. (A.1.110) allows us to simplify the MSE:

$$\text{MSE} = \frac{\lambda^2}{S} \sum_{k=1}^S \sum_{j=1}^F (\mathbf{A}_k[j] - \mathbf{A}_k[j+F])^2 \quad (\text{A.1.113})$$

$$= \frac{\lambda^2}{S} \text{trace}[(\mathbf{A}^L - \mathbf{A}^R)(\mathbf{A}^L - \mathbf{A}^R)'] \quad (\text{A.1.114})$$

$$= \frac{\lambda^2}{S} \text{trace}\left(\mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ -\mathbf{I}_F \end{bmatrix} \begin{bmatrix} \mathbf{I}_F & -\mathbf{I}_F \end{bmatrix} \mathbf{A}'\right) \quad (\text{A.1.115})$$

$$= \frac{\lambda^2}{S} \text{trace}\left(\mathbf{A} \begin{bmatrix} \mathbf{I}_F & -\mathbf{I}_F \\ -\mathbf{I}_F & \mathbf{I}_F \end{bmatrix} \mathbf{A}'\right) . \quad (\text{A.1.116})$$

We can now substitute $\mathbf{A} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ into Eq. (A.1.116) to derive Eq. (A.1.102). \square

Lemma A.2

If the $2F$ bucket images $\mathbf{i}[1], \dots, \mathbf{i}[F]$ and $\hat{\mathbf{i}}[1], \dots, \hat{\mathbf{i}}[F]$ all have zero-mean Gaussian noise with variance σ^2 , then for any frame f , the bucket ratios r and \hat{r} for $\mathbf{i}[f]$ and $\hat{\mathbf{i}}[f]$ both have approximate variance λ^2 which is independent of f .

Proof. From Proposition A.1, we know the approximate variances for r and \hat{r} are equal:

$$\text{var}(r) = \text{var}(\hat{r}) = \frac{\sigma^2}{2(j + \hat{j})^2}, \quad (\text{A.1.117})$$

where j and \hat{j} are the means of $\mathbf{i}[f]$ and $\hat{\mathbf{i}}[f]$, respectively. From Eq. (A.1.100), we already know that $\mathbf{i}[f] + \hat{\mathbf{i}}[f]$ is equal to the sum of the S noiseless demultiplexed image intensities, $\sum_k i_k$. Therefore, there are no terms on the right side of Eq. (A.1.117) that depends on f and we set $\lambda^2 = \frac{\sigma^2}{2(\sum_k i_k)^2}$. \square

B Flux Probing Theory

B.1 Preliminaries

We include some clarifications and definitions from [46] that will be helpful for some of our core proofs.

For convenience, the proofs of results from Chapters 4 and 5 are integrated into this appendix.

The exposure time. In Chapters 4 and 5, we use t to denote the current instant in time and use $[0, t]$ to denote the time interval since the beginning of acquisition. As a result, t can be thought of as representing two quantities that are usually treated as distinct, *i.e.*, the current instant in time and the exposure time. This distinction is important in conventional settings (*e.g.*, exposure time for a depth measurement in active imaging or a video frame in conventional imaging) because increasing the exposure time eventually leads to temporal blur and information loss. In contrast, expanding the interval $[0, t]$ in our framework leads to a *net gain* of information. This is because flux probing does not introduce temporal blur as t increases and, in fact, frequency scanning and detection can be performed at higher resolution as the acquisition interval expands (see Section 4.4.1). Thus, aside from computational efficiency reasons, it is advantageous to probe flux by taking the entire timestamp stream into account: as time t increases and more photons are detected, the reconstruction of the *entire* flux function from $[0, t]$ can be revised by incorporating timestamps as they arrive in order to achieve higher accuracy. This motivates using t to denote both the current time instant and the total exposure time. For the sake of clarity, however, these quantities are treated as distinct in this chapter, with the latter denoted by t_{exp} .

A key property of martingales is that their square can be expressed as the sum of another martingale $V(t)$ and a non-decreasing function that can be thought of as a stochastic generalization of the concept of variance:

Definition B.1 (Quadratic Variation (Remark 11.1.2 in [46]))

The **quadratic variation** of a martingale $M(t)$, denoted by $\langle M \rangle$, is the unique increasing process:

$$\langle M \rangle = M^2(t) - V(t), \tag{B.1.1}$$

where $V(t)$ is a martingale.

We leverage Definition B.1 for many proofs and derivations throughout this chapter.

B.2 Proofs and Derivations for Negligible Dead Time

B.2.1 Proof of Proposition 5.1

We start by restating Proposition 5.1.

Proposition 5.1

Let $N(t)$ be the counting process of an inhomogeneous Poisson process with flux function $\varphi(t)$ and let \mathbf{F} be a filtration such that $N(t)$ is \mathbf{F} -adapted. The process

$$M(t) = N(t) - \int_0^t \varphi(u) du, \quad t \geq 0, \quad (5.4.3)$$

is an \mathbf{F} -martingale.

Proof Sketch of Proposition 5.1.

The proof proceeds in two steps:

Step 1: We show $M(t)$ satisfies $E[|M(t)|] < \infty$ which follows from triangle inequality and each term in the triangle inequality being finite.

Step 2: We show $E[M(t) | \mathbf{F}_s] = M(s)$ which follows from the memoryless property of Poisson processes. \square

Proof of Proposition 5.1.

Step 1: Let $t \geq 0$ and $M(t) = N(t) - \int_0^t \varphi(u) du$. Then, we have:

$$E[|M(t)|] = E\left[\left|N(t) - \int_0^t \varphi(u) du\right|\right] \quad (B.2.1a)$$

$$\leq \underbrace{E[|N(t)|]}_{|N(t)|=N(t)} + E\left[\underbrace{\left|\int_0^t \varphi(u) du\right|}_{\varphi(u) \geq 0}\right] \quad (B.2.1b)$$

$$\leq \underbrace{E[N(t)]}_{=\int_0^t \varphi(u) du} + \int_0^t \varphi(u) du \quad (B.2.1c)$$

$$= 2 \int_0^t \varphi(u) du < \infty. \quad (B.2.1d)$$

Step 2: Let $s < t$ and $M(t) = N(t) - \int_0^t \varphi(u) du$. We have to show that

$$E[M(t) | \mathbf{F}_s] = M(s). \quad (B.2.2)$$

By substituting Eq. (5.4.3) in Eq. (B.2.2), we have:

$$E[M(t) | \mathbf{F}_s] = E[N(t) - \int_0^t \varphi(u) du | \mathbf{F}_s] \quad (\text{B.2.3a})$$

$$= E[N(t) | \mathbf{F}_s] - \int_0^t \varphi(u) du \quad (\text{B.2.3b})$$

$$= E[N(t) - N(s) + N(s) | \mathbf{F}_s] - \int_0^t \varphi(u) du \quad (\text{B.2.3c})$$

$$= E[N(t) - N(s) | \mathbf{F}_s] + E[N(s) | \mathbf{F}_s] - \int_0^t \varphi(u) du. \quad (\text{B.2.3d})$$

Due to the independent increment property of Poisson process, the photon count, $N(t) - N(s)$, in the interval $[s, t]$ do not depend on the arrival of photons before s , *i.e.*, its history \mathbf{F}_s . As such, $N(t) - N(s)$ is conditionally independent of \mathbf{F}_s so $E[N(t) - N(s) | \mathbf{F}_s] = E[N(t) - N(s)] = \int_s^t \varphi(u) du$. In a similar vein, the count at s is determined entirely by its history \mathbf{F}_s and so $E[N(s) | \mathbf{F}_s] = N(s)$. Substituting these back into Eq. (B.2.3) gives us:

$$E[M(t) | \mathbf{F}_s] = \int_s^t \varphi(u) du + E[N(s) | \mathbf{F}_s] - \int_0^t \varphi(u) du \quad (\text{B.2.4a})$$

$$= \int_s^t \varphi(u) du + N(s) - \int_0^t \varphi(u) du \quad (\text{B.2.4b})$$

$$= N(s) - \int_0^s \varphi(u) du \quad (\text{B.2.4c})$$

$$= M(s). \quad (\text{B.2.4d})$$

Therefore $M(t)$ is an \mathbf{F} -martingale. \square

Remark. While the decomposition in Eq. (5.4.3) is reminiscent of the Doob-Meyer decomposition in stochastic processes [46], the proofs below do not rely on that decomposition. Instead, we derive it directly from first principles because our proofs require an explicit construction.

B.2.2 Flux probing operation (proof of Proposition 4.1)

We restate Proposition 4.1:

Proposition 4.1 (Flux Probing Equation)

The inner product of $p(t)$ and the unknown flux function $\varphi(t)$ over the time interval $[0, t_{\text{exp}}]$ can be expressed as

$$\langle p, \varphi \rangle = p(\mathcal{T}) + M_p(t_{\text{exp}}),$$

where $p(\mathcal{T})$ are “probing measurements” which sum the values of the probing function at the photon timestamps

$$p(\mathcal{T}) \stackrel{\text{def}}{=} \sum_{t \in \mathcal{T}} p(t) ,$$

$M_p(t)$ is a martingale, and the inner product is $\int_0^{t_{\text{exp}}} p(u) \varphi(u) du$.

Proof Sketch of Proposition 4.1.

The proof proceeds in three steps:

- Step 1:** We express $\langle p, \varphi \rangle$ in terms of its Lebesgue-Stieltjes integral $\int_0^{t_{\text{exp}}} p(u) d\Phi(u)$ where $\Phi(t) = \int_0^t \varphi(u) du$. Furthermore, we substitute in the decomposition in Eq. (1) into this Lebesgue-Stieltjes integral.
- Step 2:** We then simplify the first term of the Lebesgue-Stieltjes integral which is the integral of $p(t)$ with respect to the counting process.
- Step 3:** We characterize the Lebesgue-Stieltjes integral with respect to a **F**-martingale $M(t)$ using Definition 5.8, which we restate for convenience:

Definition 5.8 (Definition 8.3.2.1(i) in [98])

Let $M(t)$ be an **F**-martingale. For any bounded deterministic function p , the process $M_p(t)$ given by

$$M_p(t) = - \int_0^t p(u) dM(u),$$

is a **F**-martingale.

By definition then, the integral of a function p with respect to an **F**-martingale is effectively still an **F**-martingale. \square

Proof of Proposition 4.1.

Step 1: We start by expressing $\langle p, \varphi \rangle$ in terms of the Lebesgue-Stieltjes integral:

$$\langle p, \varphi \rangle = \int_0^{t_{\text{exp}}} p(u) \varphi(u) du \quad (\text{B.2.5a})$$

$$= \int_0^{t_{\text{exp}}} p(u) d\Phi(u), \quad (\text{B.2.5b})$$

where $\Phi(t) = \int_0^t \varphi(u) du$. Substituting Eq. (1) into the integrator gives us

$$\langle p, \varphi \rangle = \int_0^{t_{\text{exp}}} p(u) dN(u) - \int_0^{t_{\text{exp}}} p(u) dM(u). \quad (\text{B.2.6})$$

Step 2: We evaluate the integral of $p(t)$ with respect to the counting process $N(t)$. Since $N(t)$ is a piecewise constant function, the Lebesgue-Stieltjes integral is simply the sum of $p(t)$ evaluated at the location of the jumps in $N(t)$ weighted by the size of the jump. In our case, these jumps occur exactly at our timestamps \mathcal{T} and their magnitude is 1 and therefore we have:

$$\int_0^{t_{\text{exp}}} p(u) dN(u) = \underbrace{\sum_{t \in \mathcal{T}} p(t)}_{p(\mathcal{T})}. \quad (\text{B.2.7})$$

Step 3: Now we consider the integral of $p(t)$ with respect to the **F**-martingale $M(t)$. Let us define $M_p(t_{\text{exp}})$ according to Eq. (5.4.4). Using Definition 5.8, we know that $M_p(t_{\text{exp}})$ is a **F**-martingale. Combining Definition 5.8 and Eq. (B.2.7) gives us the flux probing equation:

$$\langle p, \varphi \rangle = \sum_{t \in \mathcal{T}} p(t) + M_p(t_{\text{exp}}). \quad (\text{B.2.8})$$

\square

B.2.3 Maximum frequency (proof of Proposition 4.2)

In this subsection, we prove Proposition 4.2:

Proposition 4.2

Given timing resolution Q , the maximum recoverable frequency is $\frac{1}{2Q}$.

An analogous observation can be found in [53] for irregular samples corresponding to conventional (*i.e.*, non-photon based) astronomical measurements of a variable star's flux.

Proof Sketch of Proposition 4.2.

The main idea here is that every quantized timestamp can be written as an integer multiple of the quantization resolution $\frac{1}{Q}$. Therefore when we probe with frequencies above $\frac{1}{2Q}$, the probing function gets wrapped into a lower frequency. Specifically, we show that for any frequency $f > \frac{1}{2Q}$, there is a frequency $f_0 < f$ such that probing with f_0 is equivalent to probing with f . In other words, there is aliasing for any frequency above $\frac{1}{2Q}$.

The proof proceeds in two steps:

Step 1: We show that $\forall f > \frac{1}{Q}, f_0 = f \bmod Q$ is an alias.

Step 2: We show that $\forall f \in [\frac{1}{2Q}, \frac{1}{Q}], f_0 = \frac{1}{Q} - f$ is an alias. □

Proof of Proposition 4.2.

Let us start by defining the quantization of timestamps with a quantization function $q(t)$ as follows:

$$q(t) = Q \left\lfloor \frac{t}{Q} + \frac{1}{2} \right\rfloor. \quad (\text{B.2.9})$$

where Q is the SPAD's timing resolution. Then given timestamps \mathcal{T} detected during $[0, t_{\text{exp}}]$, we have the set of quantized timestamps $q(\mathcal{T}) = \{q(t)\}_{t \in \mathcal{T}}$.

For simplicity, we consider a flux function $\varphi(t) = N + A \cos(2\pi ft + \phi)$ and probing functions $p_f(t) = \frac{1}{t_{\text{exp}}} e^{-2\pi j f t}$.

Step 1 ($f > \frac{1}{Q}$): Since $f > \frac{1}{Q}$, there exist $f_0 \in [0, \frac{1}{Q}]$ and $k \in \mathbb{N}$ such that $f = \frac{k}{Q} + f_0$. It is easy to see that $f_0 < f$ by construction.

We show that probing with f_0 is equivalent to probing with f :

$$p_{f_0}(q(\mathcal{T})) = \sum_{t \in q(\mathcal{T})} p_{f_0}(t) \quad (\text{B.2.10a})$$

$$= \sum_{t \in \mathcal{T}} p_{f_0}(q(t)) \quad (\text{B.2.10b})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} e^{-2\pi j f_0 q(t)} \quad (\text{B.2.10c})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} e^{-2\pi j f_0 Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor} \quad (\text{B.2.10d})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} e^{-2\pi j f_0 Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor} \underbrace{e^{-2\pi j \frac{k}{Q} Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor}}_{=1} \quad (\text{B.2.10e})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} e^{-2\pi j (\frac{k}{Q} + f_0) Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor} \quad (\text{B.2.10f})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}} e^{-2\pi j f q(t)} \quad (\text{B.2.10g})$$

$$= p_f(q(\mathcal{T})). \quad (\text{B.2.10h})$$

Step 2 $\left(f \in (\frac{1}{2Q}, \frac{1}{Q}]\right)$: Consider $f_0 = \frac{1}{Q} - f$. It is easy to confirm that $f_0 < f$. We make the following observation:

$$p_{f_0}(q(t)) = \frac{1}{t_{\text{exp}}} e^{2\pi j f_0 q(t)} \quad (\text{B.2.11a})$$

$$= \frac{1}{t_{\text{exp}}} e^{-2\pi j f_0 Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor} \quad (\text{B.2.11b})$$

$$= \frac{1}{t_{\text{exp}}} e^{-2\pi j (\frac{1}{Q} - f) Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor} \quad (\text{B.2.11c})$$

$$= \frac{1}{t_{\text{exp}}} \underbrace{e^{-2\pi j \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor}}_{=1} \underbrace{e^{2\pi j f Q \lfloor \frac{t}{Q} + \frac{1}{2} \rfloor}}_{=p_{-f}(q(t))} \quad (\text{B.2.11d})$$

$$= p_{-f}(q(t)). \quad (\text{B.2.11e})$$

From Eq. (B.2.11e), we have $|p_{f_0}(\mathcal{T})| = |p_f(\mathcal{T})|$ and $\angle(p_{f_0}(\mathcal{T})) = -\angle(p_f(\mathcal{T}))$. Therefore, the timestamps generated from $\varphi(t)$ are indistinguishable from those generated from $\varphi_0(t) = N + A \cos(2\pi f_0 t - \phi)$. Combining the results of Step 1 and Step 2, we have that for any frequency $f > \frac{1}{2Q}$ in our flux function φ , there exists a frequency $f_0 < f$ such that f_0 is an alias. Therefore, the maximum frequency we can detect with probing is $\frac{1}{2Q}$. \square

B.2.4 Distribution of probing measurements (proof of Proposition 4.3)

Proposition 4.3 (Distribution of Probing Measurements)

The probing measurements $p(\mathcal{T})$ are approximately normally distributed with mean $\langle p, \varphi \rangle$ and variance $\langle p^2, \varphi \rangle$.

Proof Sketch of Proposition 4.3.

The proof proceeds in three steps:

Step 1: We show that the mean of the probing measurements is the inner product between the probing function and the flux function. We use the following lemma and corollary here:

Lemma B.1 (Proposition 8.3.2.1 in [98])

Let $N(t)$ be a counting process with flux function $\varphi(t)$, \mathbf{F} be a filtration such that $N(t)$ is \mathbf{F} -adapted. Given a deterministic function p such that $\forall t \geq 0$, $\int_0^t |p(u)|\varphi(u)du < \infty$. Then

$$\int_0^t p(u)\varphi(u)du = E\left[\int_0^t p(u)dN\right]. \quad (\text{B.2.12})$$

Corollary B.1

The \mathbf{F} -martingale $M_p(t) = \int_0^t p(u)dM$ has $E[M_p(t)] = 0$.

Step 2: We show that the variance of the probing measurements is the inner product between the square of the probing function and the flux function. We make use of the following lemmas and corollary:

Lemma B.2 (Itô Isometry (Lemma 12.1.4 in [46]))

Let $M(t)$ be a martingale and $p(t)$ be a deterministic function such that $E[\int_0^t p^2(u)d\langle M \rangle] < \infty$. Then,

$$E\left[\left(\int_0^t p(u)dM\right)^2\right] = E\left[\int_0^t p(u)^2 d\langle M \rangle\right]. \quad (\text{B.2.13})$$

Lemma B.3 (Proposition 8.3.2.1(ii) in [98])

Let $M(t)$ be an \mathbf{F} -martingale. For any bounded deterministic function p , the process $V_p(t)$ given by:

$$V_p(t) = \left(\int_0^t p(u)dM\right)^2 - \int_0^t p^2(u)\varphi(u)du, \quad (\text{B.2.14})$$

is a \mathbf{F} -martingale.

Corollary B.2

The quadratic variation of the martingale $M(t)$ defined in Eq. (5.4.3) is $\int_0^t \varphi(u)du$.

Step 3: We show that the probing measurements are approximately normally distributed. \square

Proof of Proposition 4.3.

Step 1: Taking the expectation of the flux probing equation (Eq. (4.3.2)) and using the facts that (a) the inner product is deterministic and (b) $E[M(t)] = 0$ (Corollary B.1), it follows that $E[p(\mathcal{T})] = \langle p, \varphi \rangle$.

Step 2: We now turn our attention to the variance of the probing measurements. For brevity, we denote $M(t)$ as M . Starting from Eq. (4.3.2) we have:

$$\text{var}\left[\sum_{t \in \mathcal{T}} p(t)\right] = \underbrace{\text{var}\left[\int_0^{t_{\text{exp}}} p(u)\varphi(u)du\right]}_{=0 \text{ (deterministic)}} + \text{var}[M_p(t_{\text{exp}})] \quad (\text{B.2.15a})$$

$$= E[M_p(t_{\text{exp}})^2] - \underbrace{E[M_p(t_{\text{exp}})]^2}_{=0 \text{ (Cor. B.1)}} \quad (\text{B.2.15b})$$

$$= E\left[\left(\int_0^{t_{\text{exp}}} p(u)dM\right)^2\right] \quad (\text{B.2.15c})$$

$$= E\left[\underbrace{\int_0^{t_{\text{exp}}} p(u)^2 d\langle M \rangle}_{\text{It\^o isometry}}\right] \quad (\text{B.2.15d})$$

$$= E\left[\int_0^{t_{\text{exp}}} p(u)^2 d \int_0^u \varphi(s)ds\right] \quad (\text{B.2.15e})$$

$$= E\left[\underbrace{\int_0^{t_{\text{exp}}} p(u)^2 \varphi(u)du}_{\text{deterministic}}\right] \quad (\text{B.2.15f})$$

$$= \langle p^2, \varphi \rangle. \quad (\text{B.2.15g})$$

The second term in Eq. (B.2.15b) is equal to 0 from Corollary B.1. Eq. (B.2.15d) uses It\^o isometry (Lemma B.2) and Eq. (B.2.15e) uses the quadratic variation $\langle M \rangle$ from Corollary B.2.

Step 3: Finally, we focus on the distribution of the probing measurements. To this end, we observe that the probing measurement is the sum of independent identically-distributed random variables because each timestamp is a random variable. As the number of terms in the sum increases, the density approaches a normal density under the central limit theorem; in practice, a sum of approximately ten such random variables results in a density that closely approximates a normal distribution [3]. \square

We also verify this property empirically, via simulation, in Section 4.6.1.

Proof of Corollary B.1. We know from Eq. (1) that $M_p(t) = \int_0^t p(u)\varphi(u)du - \int_0^t p(u)dN$. Taking the expectation of both sides and using Lemma B.1, we have that $E[M_p(t)] = 0$. \square

Proof of Corollary B.2. By setting $p(t) = 1$ in Eq. (B.2.14) we have that $M^2(t) - \int_0^t \varphi(u)du$ is a martingale. Thus from the definition of the quadratic variation in Definition B.1 we obtain that $\langle M \rangle = \int_0^t \varphi(u)du$. It is straightforward to show that $\langle M \rangle(0) = 0$. \square

B.2.5 Covariance of probing measurements

In this section, we prove the following proposition which we then use to prove Corollary 4.1:

Proposition B.1 (Covariance of Probing Measurements)

For arbitrary probing functions $p_1(t)$, $p_2(t)$ we have that $\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = \int_0^{t_{\text{exp}}} p_1(u)p_2(u)\varphi(u)du$.

Proof Sketch of Proposition B.1. The proof proceeds in two steps:

Step 1: We use the definition $\mathbf{Cov}(X, Y) = E[XY] - E[X]E[Y]$ to expand $\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T}))$ and then simplify $E[p_1(\mathcal{T})p_2(\mathcal{T})]$.

Step 2: We simplify the expansion to derive $\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = \int_0^{t_{\text{exp}}} p_1(u)p_2(u)\varphi(u)du$. \square

Proof of Proposition B.1.

Step 1: From the definition of covariance, we have:

$$\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = E[p_1(\mathcal{T})p_2(\mathcal{T})] - E[p_1(\mathcal{T})]E[p_2(\mathcal{T})]. \quad (\text{B.2.16})$$

First, we focus on the first term of the right-hand side of Eq. (B.2.16):

$$E[p_1(\mathcal{T})p_2(\mathcal{T})] = E\left[\left(\int_0^{t_{\text{exp}}} p_1 dM + \int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du\right)\left(\int_0^{t_{\text{exp}}} p_2 dM + \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du\right)\right] \quad (\text{B.2.17a})$$

$$= E\left[\int_0^{t_{\text{exp}}} p_1 dM \int_0^{t_{\text{exp}}} p_2 dM + \int_0^{t_{\text{exp}}} p_1 dM \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du \right. \quad (\text{B.2.17b})$$

$$\left. + \int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du \int_0^{t_{\text{exp}}} p_2 dM + \int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du\right] \quad (\text{B.2.17c})$$

$$= \underbrace{E\left[\int_0^{t_{\text{exp}}} p_1 dM \int_0^{t_{\text{exp}}} p_2 dM\right]}_{\int_0^{t_{\text{exp}}} p_1(u)p_2(u)d\langle M \rangle \text{ (It\^o Isometry B.2)}} + \underbrace{E\left[\int_0^{t_{\text{exp}}} p_1 dM\right] \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du}_{=0 \text{ (Corollary B.1)}} \quad (\text{B.2.17d})$$

$$+ \int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du \underbrace{E\left[\int_0^{t_{\text{exp}}} p_2 dM\right]}_{=0 \text{ (Corollary B.1)}} + \underbrace{E\left[\int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du\right]}_{\text{deterministic}} \quad (\text{B.2.17e})$$

$$= E\left[\int_0^{t_{\text{exp}}} p_1(u)p_2(u)d\langle M \rangle\right] + \underbrace{\int_0^{t_{\text{exp}}} p_1(u)\varphi(u)du \int_0^{t_{\text{exp}}} p_2(u)\varphi(u)du}_{=E[p_1(\mathcal{T})]E[p_2(\mathcal{T})] \text{ (Proposition 4.1)}} \quad (\text{B.2.17f})$$

$$= \int_0^{t_{\text{exp}}} p_1(u)p_2(u)\varphi(u)du + E[p_1(\mathcal{T})]E[p_2(\mathcal{T})] \quad (\text{B.2.17g})$$

where Eq. (B.2.17g) was obtained from Eq. (B.2.17f) by substituting the quadratic variation $\langle M \rangle$ from Corollary B.2.

Step 2: Inserting Eq. (B.2.17g) in Eq. (B.2.16) we obtain:

$$\text{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = E[p_1(\mathcal{T})p_2(\mathcal{T})] - E[p_1(\mathcal{T})]E[p_2(\mathcal{T})] \quad (\text{B.2.18a})$$

$$= \int_0^{t_{\text{exp}}} p_1(u)p_2(u)\varphi(u)du + E[p_1(\mathcal{T})]E[p_2(\mathcal{T})] - E[p_1(\mathcal{T})]E[p_2(\mathcal{T})] \quad (\text{B.2.18b})$$

$$= \int_0^{t_{\text{exp}}} p_1(u)p_2(u)\varphi(u)du. \quad (\text{B.2.18c})$$

□

B.2.6 Distribution of Fourier probing (proof of Corollary 4.1)

In this section, we prove the following corollary:

Corollary 4.1 (Distribution of Fourier Probing)

The Fourier probing measurements $p_f(\mathcal{T})$ approximately follow a complex normal distribution with mean and covariance matrix

$$\mu = \begin{bmatrix} \langle \cos(2\pi ft), \varphi(t) \rangle & \langle -\sin(2\pi ft), \varphi(t) \rangle \\ \langle \cos^2(2\pi ft), \varphi(t) \rangle & 0 \\ 0 & \langle \sin^2(2\pi ft), \varphi(t) \rangle \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \langle \cos^2(2\pi ft), \varphi(t) \rangle & 0 \\ 0 & \langle \sin^2(2\pi ft), \varphi(t) \rangle \end{bmatrix}.$$

Proof Sketch of Corollary 4.1.

The proof proceeds in three steps:

- Step 1:** We derive the mean of the real and imaginary parts of the Fourier probing measurements. This follows as a consequence of Euler's formula.
- Step 2:** We derive the covariance between the real and imaginary parts of the Fourier probing measurements and show that it is always 0.
- Step 3:** We state the conditions under which the Fourier probing measurements approximately follow a complex normal distribution. □

Proof of Corollary 4.1.

Step 1: Applying Euler's formula,

$$\text{Re}[p_f(\mathcal{T})] = \sum_{t \in \mathcal{T}} \cos(2\pi ft) \iff \quad (\text{B.2.19a})$$

$$E[\text{Re}[p_f(\mathcal{T})]] = E\left[\sum_{t \in \mathcal{T}} \cos(2\pi ft)\right] \iff \quad (\text{B.2.19b})$$

$$E[\text{Re}[p_f(\mathcal{T})]] \underbrace{=} \langle \cos(2\pi ft), \varphi(t) \rangle. \quad (\text{B.2.19c})$$

Proposition 4.3

Similarly, we have $E[\text{Im}[p_f(\mathcal{T})]] = \langle -\sin(2\pi ft), \varphi(t) \rangle$.

Step 2: Using Proposition B.1 and setting $p_1(\mathcal{T}) = \text{Re}[p_f(\mathcal{T})]$ and $p_2(\mathcal{T}) = \text{Im}[p_f(\mathcal{T})]$ in Eq. (B.2.18),

we have

$$\mathbf{Cov}(\text{Re}[p_f(\mathcal{T})], \text{Im}[p_f(\mathcal{T})]) = \int_0^{t_{\text{exp}}} \cos(2\pi fu) \sin(2\pi fu) \varphi(u) du \quad (\text{B.2.20a})$$

$$= \frac{1}{2} \langle \sin(4\pi ft), \varphi(t) \rangle. \quad (\text{B.2.20b})$$

We integrate over full periods. Letting A_{2f} be the amplitude of frequency $2f$ in φ , Eq. (B.2.20b) tells us that the covariance is bounded by $\frac{A_{2f}}{2}$. We note the variances:

$$\langle \cos^2(2\pi ft), \varphi(t) \rangle = \frac{1}{2} \underbrace{\langle 1, \varphi(t) \rangle}_{=N(t)} + \frac{1}{2} \langle \cos(4\pi ft), \varphi(t) \rangle, \quad (\text{B.2.21a})$$

$$\langle \sin^2(2\pi ft), \varphi(t) \rangle = \frac{1}{2} \underbrace{\langle 1, \varphi(t) \rangle}_{=N(t)} - \frac{1}{2} \langle \cos(4\pi ft), \varphi(t) \rangle. \quad (\text{B.2.21b})$$

In general, we expect $N(t)$ to be greater than A_{2f} , so the covariance matrix Σ is approximately diagonal. In the case where $A_{2f} = 0$, this is exact.

Step 3: We know from Proposition 4.3 that the real and imaginary parts of the Fourier probing measurements are normally distributed. As a consequence, their linear combinations are also normally distributed and thus their joint distribution is a bivariate normal. We also know from Step 2 that their covariance is 0. Finally, we know that random variables whose joint probability density function is a normal distribution with covariance 0 are independent.

To conclude, the real and imaginary parts are independent normal distributions and by definition the complex random variable $p_f(\mathcal{T}) = \text{Re}[p_f(\mathcal{T})] + j\text{Im}[p_f(\mathcal{T})]$ is a complex normal random variable with mean

$$\mu = \begin{bmatrix} \langle \cos(2\pi ft), \varphi(t) \rangle & \langle -\sin(2\pi ft), \varphi(t) \rangle \end{bmatrix}, \quad (\text{B.2.22})$$

and covariance matrix

$$\Sigma = \begin{bmatrix} \text{var}[\text{Re}[p_f(\mathcal{T})]] & 0 \\ 0 & \text{var}[\text{Im}[p_f(\mathcal{T})]] \end{bmatrix} \quad (\text{B.2.23a})$$

$$\stackrel{\text{Proposition 4.3}}{=} \begin{bmatrix} \langle \cos^2(2\pi ft), \varphi(t) \rangle & 0 \\ 0 & \langle \sin^2(2\pi ft), \varphi(t) \rangle \end{bmatrix}. \quad (\text{B.2.23b}) \quad \square$$

B.2.7 Distribution of Fourier probing energy (proof of Corollary 4.2)

In this section, we prove the following:

Corollary 4.2 (Distribution of Fourier Probing Energy)

The normalized energy of the Fourier basis probing measurements

$$p_f^{\mathcal{E}}(\mathcal{T}) \stackrel{\text{def}}{=} \text{Re} \left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{1,1}}} \right]^2 + \text{Im} \left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{2,2}}} \right]^2$$

follows a non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter:

$$\frac{\mu_1^2}{\Sigma_{1,1}} + \frac{\mu_2^2}{\Sigma_{2,2}}.$$

Proof Sketch of Corollary 4.2. The proof proceeds in three steps:

Step 1: We show that $\text{Re}\left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{1,1}}}\right] \sim \mathcal{N}(\mu_1, 1)$.

Step 2: We show that $\text{Im}\left[\frac{p_f(\mathcal{T})}{\sqrt{\Sigma_{2,2}}}\right] \sim \mathcal{N}(\mu_2, 1)$.

Step 3: We derive the statistics of the non-central χ^2 distribution. □

Proof of Corollary 4.2.

Step 1: Let X be the random variable $\text{Re}[p_f(\mathcal{T})]$. From Corollary 4.1, we have $X \sim \mathcal{N}(\mu_1, \Sigma_{1,1})$. Then we know that

$$\frac{1}{\sqrt{\Sigma_{1,1}}}X \sim \mathcal{N}\left(\mu_1, \frac{\Sigma_{1,1}}{\sqrt{\Sigma_{1,1}}^2}\right) = \mathcal{N}(\mu_1, 1). \quad (\text{B.2.24})$$

Step 2: Let Y be the random variable $\text{Im}[p_f(\mathcal{T})]$. From Corollary 4.1, we have $Y \sim \mathcal{N}(\mu_2, \Sigma_{2,2})$. Then we know that

$$\frac{1}{\sqrt{\Sigma_{2,2}}}Y \sim \mathcal{N}\left(\mu_2, \frac{\Sigma_{2,2}}{\sqrt{\Sigma_{2,2}}^2}\right) = \mathcal{N}(\mu_2, 1). \quad (\text{B.2.25})$$

Step 3: From Corollary 4.1, we know that X and Y are independent. Therefore, $X/\sqrt{\Sigma_{1,1}}$ and $Y/\sqrt{\Sigma_{2,2}}$ are also independent. We define Z as follows:

$$Z = \frac{X^2}{\Sigma_{1,1}} + \frac{Y^2}{\Sigma_{2,2}}. \quad (\text{B.2.26})$$

Then Z is a non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter

$$E\left[\frac{X}{\sqrt{\Sigma_{1,1}}}\right]^2 + E\left[\frac{Y}{\sqrt{\Sigma_{2,2}}}\right]^2 = \frac{\mu_1^2}{\Sigma_{1,1}} + \frac{\mu_2^2}{\Sigma_{2,2}}. \quad (\text{B.2.27})$$

□

B.2.8 Unbiased estimators from probing

In this section, we show that our estimators from probing are unbiased. Specifically, we show that

$$\forall \epsilon > 0, \lim_{t_{\text{exp}} \rightarrow \infty} \text{Pr}(|p_f(\mathcal{T}) - \langle p_f, \varphi \rangle| > \epsilon) = 0 \quad (\text{B.2.28})$$

and

$$\forall \epsilon > 0, \lim_{t_{\text{exp}} \rightarrow \infty} \text{Pr}(|p_f^2(\mathcal{T}) - \langle p_f^2, \varphi \rangle| > \epsilon) = 0. \quad (\text{B.2.29})$$

Therefore, these estimators converge to the true mean and variance in probability. We make a mild assumption here that the flux is always finite, *i.e.*, $\|\varphi\|_\infty < \infty$.

Proof Sketch.

The proof proceeds in two steps:

Step 1: We show that the estimated mean converges to the true mean in probability. We leverage Chebyshev's inequality for this.

Step 2: Similarly, we show that the estimated variance converges to the true variance in probability. \square

Proof.

Step 1: Consider a set of N timestamps \mathcal{T} . To recover the Fourier coefficient of frequency f for φ in $[0, t_{\text{exp}}]$, we probe with $p_f(t) = \frac{1}{t_{\text{exp}}} e^{-2\pi i f t}$. From Proposition 4.3, we have:

$$E[p_f(\mathcal{T})] = \langle p_f, \varphi \rangle. \quad (\text{B.2.30})$$

Using Chebyshev's inequality we have:

$$\Pr(|p_f(\mathcal{T}) - \langle p_f, \varphi \rangle| > \epsilon) \leq \frac{\text{var}(p_f(\mathcal{T}))}{\epsilon^2}. \quad (\text{B.2.31})$$

We now show that the real component of the estimated mean converges to the real component of the true mean in probability:

$$\Pr(|\text{Re}[p_f(\mathcal{T})] - \text{Re}[\langle p_f, \varphi \rangle]| > \epsilon) \leq \frac{\text{var}(\text{Re}[p_f(\mathcal{T})])}{\epsilon^2} \quad (\text{B.2.32a})$$

$$\leq \frac{\langle \text{Re}[p_f]^2, \varphi \rangle}{\epsilon^2} \quad (\text{B.2.32b})$$

$$\leq \frac{1}{\epsilon^2 t_{\text{exp}}^2} \langle \cos^2(2\pi f t), \varphi \rangle \quad (\text{B.2.32c})$$

$$\leq \frac{1}{\epsilon^2 t_{\text{exp}}^2} \underbrace{\|\cos^2(2\pi f t)\|_1 \|\varphi\|_\infty}_{\text{Hölder's inequality}} \quad (\text{B.2.32d})$$

$$\leq \frac{\|\varphi\|_\infty}{\epsilon^2 t_{\text{exp}}^2} \underbrace{\frac{4\pi f t_{\text{exp}} + \sin(4\pi f t_{\text{exp}})}{8\pi f}}_{\in \mathcal{O}(t_{\text{exp}})} \quad (\text{B.2.32e})$$

$$\leq \frac{\|\varphi\|_\infty \mathcal{O}(t_{\text{exp}})}{\epsilon^2 t_{\text{exp}}^2} \rightarrow 0. \quad (\text{B.2.32f})$$

We use Hölder's inequality with $p = 1$ and $q = \infty$. Since $\|\sin^2(2\pi f t)\|_1 \in \mathcal{O}(t_{\text{exp}})$, the same inequalities in Eq. (B.2.32a) hold for the imaginary component of the estimated mean. Therefore, the means of both the real and imaginary component converge in probability to their true mean.

Step 2: Let $p_f^2(\mathcal{T})$ denote the estimated variance of $p_f(\mathcal{T})$. We show that $p_f^2(\mathcal{T})$ converges to $\langle p_f^2, \varphi \rangle$ in probability. From Proposition 4.3,

$$\text{var}(p_f(\mathcal{T})) = \langle p_f^2, \varphi \rangle. \quad (\text{B.2.33})$$

Using Proposition 4.1, we have:

$$\langle p_f^2, \varphi \rangle = p_f^2(\mathcal{T}) + m. \quad (\text{B.2.34})$$

It follows from Proposition 4.3 that $\text{var}(m) = \langle p_f^4, \varphi \rangle$.

From Eq. (B.2.34) and Chebyshev's inequality,

$$\Pr(|p_f^2(\mathcal{T}) - \langle p_f^2, \varphi \rangle| > \epsilon) = \Pr(|m| > \epsilon) \quad (\text{B.2.35})$$

$$\leq \frac{\text{var}(m)}{\epsilon^2}. \quad (\text{B.2.36})$$

To recover the Fourier coefficient for φ in $[0, t_{\text{exp}}]$, we probe with $p_f(t) = \frac{1}{t_{\text{exp}}} e^{-2\pi j f t}$. We now show that the variance of the real and imaginary components of m converges in probability to the true variance.

We start with the real component:

$$\Pr(|\text{Re}[m]| > \epsilon) \leq \frac{\text{var}(\text{Re}[m])}{\epsilon^2} \quad (\text{B.2.37a})$$

$$\leq \frac{1}{\epsilon^2 t_{\text{exp}}^4} \langle \cos^4(2\pi f t), \varphi \rangle \quad (\text{B.2.37b})$$

$$\leq \frac{1}{\epsilon^2 t_{\text{exp}}^4} \underbrace{\|\cos^4(2\pi f t)\|_1 \|\varphi\|_\infty}_{\text{Hölder's inequality}} \quad (\text{B.2.37c})$$

$$\leq \frac{\|\varphi\|_\infty}{\epsilon^2 t_{\text{exp}}^4} \underbrace{24\pi f t_{\text{exp}} + 8 \sin(4\pi f t_{\text{exp}}) + \sin(8\pi f t_{\text{exp}})}_{\in \mathcal{O}(t_{\text{exp}})} \quad (\text{B.2.37d})$$

$$\leq \frac{\|\varphi\|_\infty \mathcal{O}(t_{\text{exp}})}{\epsilon^2 t_{\text{exp}}^4} \rightarrow 0. \quad (\text{B.2.37e})$$

We use Hölder's inequality with $p = 1$ and $q = \infty$.

Since $\|\sin^4(2\pi f t)\|_1 \in \mathcal{O}(t_{\text{exp}})$, the same result holds for the imaginary component. Therefore, the variance of both the real and imaginary component converges in probability to the true variance. \square

B.2.9 Frequency detection

In this section, we prove Equation 4.3.7, justifying our threshold criterion:

$$|p_f(\mathcal{T})|^2 \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \frac{N(t_{\text{exp}})}{2t_{\text{exp}}^2}.$$

Proof Sketch.

The proof proceeds in two steps:

Step 1: We show that the real and imaginary parts of the Fourier Probing Energy $p_f(\mathcal{T})$ have roughly equal variance.

Step 2: We derive the threshold for the amplitude of the Fourier probing measurements. \square

Proof.

Step 1: We probe with $p_f(t) = \frac{1}{t_{\text{exp}}} e^{-2\pi j f t}$. We know from Corollary 4.1 that the variance of the random variables $\text{Re}[p_f(\mathcal{T})]$ and $\text{Im}[p_f(\mathcal{T})]$ is, respectively,

$$\text{var}[\text{Re}[p_f(\mathcal{T})]] = \left\langle \frac{\cos^2(2\pi f t)}{t_{\text{exp}}^2}, \varphi(t) \right\rangle = \frac{N(t)}{2t_{\text{exp}}^2} + \frac{\overbrace{\langle \cos(4\pi f t), \varphi(t) \rangle}^{A_{2f}}}{2t_{\text{exp}}^2}, \quad (\text{B.2.38a})$$

$$\text{var}[\text{Im}[p_f(\mathcal{T})]] = \left\langle \frac{\sin^2(2\pi f t)}{t_{\text{exp}}^2}, \varphi(t) \right\rangle = \frac{N(t)}{2t_{\text{exp}}^2} - \frac{\langle \cos(4\pi f t), \varphi(t) \rangle}{2t_{\text{exp}}^2}, \quad (\text{B.2.38b})$$

where A_{2f} is the amplitude of frequency $2f$. In general, we expect $N(t)$ to be greater than A_{2f} . Therefore,

$$\underbrace{\text{var}[\text{Re}[p_f(\mathcal{T})]]}_{\Sigma_{1,1}} = \underbrace{\text{var}[\text{Im}[p_f(\mathcal{T})]]}_{\Sigma_{2,2}} \approx \frac{N(t_{\text{exp}})}{2t_{\text{exp}}^2}. \quad (\text{B.2.39})$$

Step 2: Combining Step 1 with the definition of the Fourier Probing Energy from Corollary 4.2, we obtain

$$p_f^{\mathcal{E}}(\mathcal{T}) = \frac{2t_{\text{exp}}^2}{N(t_{\text{exp}})} |p_f(\mathcal{T})|^2. \quad (\text{B.2.40a})$$

From the CFAR detection test, we detect frequency f if

$$p_f^{\mathcal{E}}(\mathcal{T}) \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \quad (\text{B.2.41a})$$

$$|p_f(\mathcal{T})|^2 \frac{2t_{\text{exp}}^2}{N(t_{\text{exp}})} \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \quad (\text{B.2.41b})$$

$$|p_f(\mathcal{T})|^2 \geq \text{CDF}_{\chi^2}^{-1}(1 - \alpha) \frac{N(t_{\text{exp}})}{2t_{\text{exp}}^2}, \quad (\text{B.2.41c})$$

which is Equation 4.3.7. □

B.3 Generalization Proofs for Non-Negligible Dead Time

In this section we extend our theory to the case where dead time is not negligible.

B.3.1 Generalized flux probing equation (proof of Proposition 5.2)

We start by restarting Proposition 5.2:

Proposition 5.2 (Generalized Flux Probing Equation)

The inner product of $p(t)$ and the unknown flux function $\varphi(t)$ over the time interval $[0, t_{\text{exp}}]$ can be expressed as:

$$\langle p, \varphi \rangle_{\text{active}} = p(\mathcal{T}) + M(t_{\text{exp}}),$$

where $\langle p, \varphi \rangle_{\text{active}} = \int_{\Omega_{\text{active}}} p(u) \varphi(u) du$ and Ω_{active} is the subset of $[0, t_{\text{exp}}]$ where the SPAD is active.

Proof Sketch of Proposition 5.2.

The proof proceeds in three steps:

Step 1: We recap **marked point processes** as a model for timestamp detections and introduce relevant conditional probabilities.

Step 2: We define the **flux process** $\Phi(t)$ in terms of these conditional probabilities and show that it is equivalent to $\varphi(t)$ while the SPAD is active and 0 otherwise¹. This gives us a decomposition similar to Eq. (1) but takes dead time into account.

Step 3: We derive the dead-time flux probing equation. \square

Proof of Proposition 5.2.

Note that for clarity purposes, statements of Lemmas B.4-B.8 used in the proof below can be found after the proof of Proposition 5.2.

Step 1: Recall that the sequence (t_n, m_n) is called a *marked point process* [32] if in addition to the set of photon arrival timestamps \mathcal{T} themselves, each timestamp t_n is assigned a binary mark m_n , representing the two possible outcomes, *i.e.*, detected (1) and not-detected (0). Then, from Lemma B.8, it follows that the conditional probability P_j of the j^{th} timestamp arriving at time t and being assigned mark $m_j = 1$ is

$$P_j(t_j = t, m_j = 1 | \mathbf{F}_{t_{j-1}}) = \mathbf{1}_{t > t_i + D} \varphi(t) e^{-\int_{t_{j-1}}^t \varphi(u) du}, \quad (\text{B.3.1})$$

where t_i is the last timestamp before t_j for which $m_i = 1$, and $\mathbf{1}_{t > t_i + D}$ is the indicator function for t being outside the dead time (which occurs when $t > t_i + D$).

Step 2: Let us define the following stochastic process $\Phi(t)$:

$$\Phi(t) = \sum_{n \geq 0} \frac{P_{n+1}(t_{n+1} = t, m_{n+1} = 1 | \mathbf{F}_{t_n})}{1 - \int_0^{t-t_n} P_{n+1}(t_{n+1} = u + t_n | \mathbf{F}_{t_n}) du} \mathbf{1}_{t_n \leq t < t_{n+1}}. \quad (\text{B.3.2})$$

We call $\Phi(t)$ the *flux process*. The reason for this will become evident shortly. Now consider n such that $t_n \leq t < t_{n+1}$. Then we have:

$$\Phi(t) = \frac{P_{n+1}(t_{n+1} = t, m_{n+1} = 1 | \mathbf{F}_{t_n})}{1 - \int_0^{t-t_n} P_{n+1}(t_{n+1} = u + t_n | \mathbf{F}_{t_n}) du} \quad (\text{B.3.3a})$$

$$= \frac{\mathbf{1}_{t > t_i + D} \varphi(t) e^{-\int_{t_n}^t \varphi(u) du}}{1 - \int_0^{t-t_n} \varphi(u + t_n) e^{-\int_{t_n}^{u+t_n} \varphi(v) dv} du} \quad (\text{B.3.3b})$$

$$= \frac{\mathbf{1}_{t > t_i + D} \varphi(t) e^{-\int_{t_n}^t \varphi(u) du}}{1 - \int_{t_n}^t \varphi(u) e^{-\int_{t_n}^u \varphi(v) dv} du}. \quad (\text{B.3.3c})$$

Using Lemma B.4 with interval $[t_n, t]$ and Eq. (B.3.3c) gives us

$$\Phi(t) = \mathbf{1}_{t > t_i + D} \varphi(t). \quad (\text{B.3.4})$$

¹The approach taken here can be thought of as a generalization of [194] to the case of flux probing.

Let Ω_{active} be the set of active SPAD times in the interval $[0, t_{\text{exp}}]$. Then

$$\int_0^{t_{\text{exp}}} \Phi(u) du = \int_0^{t_{\text{exp}}} \mathbf{1}_{t > t_i + D} \varphi(t) du = \int_{\Omega_{\text{active}}} \varphi(t) du. \quad (\text{B.3.5})$$

Now we use Lemma B.6 to show that the stochastic process

$$M(t_{\text{exp}}) = N^{(1)}(t_{\text{exp}}) - \int_0^{t_{\text{exp}}} \Phi(u) du = N^{(1)}(t_{\text{exp}}) - \int_{\Omega_{\text{active}}} \varphi(t) du, \quad (\text{B.3.6})$$

is an \mathbf{F} -martingale on $[0, t_{\text{exp}}]$.

Step 3: Finally, from Lemma B.7,

$$E\left[\int_0^{t_{\text{exp}}} p(u) dN^{(1)}(u)\right] = E\left[\int_0^{t_{\text{exp}}} p(u) \Phi(u) du\right]. \quad (\text{B.3.7})$$

Therefore, we have

$$M(t_{\text{exp}}) = \int_0^{t_{\text{exp}}} p(u) dN^{(1)}(u) - \int_0^{t_{\text{exp}}} p(u) \Phi(u) du \quad (\text{B.3.8a})$$

$$= \sum_{t \in \mathcal{T}^{(1)}} p(t) - \int_{\Omega_{\text{active}}} p(u) \varphi(t) du, \quad (\text{B.3.8b})$$

where $\mathcal{T}^{(1)}$ is the set of detected timestamps. □

Lemma B.4

Given an integrable function $\varphi(t)$ and an interval $[a, b]$, the following holds:

$$\frac{e^{-\int_a^b \varphi(u) du}}{1 - \int_a^b \varphi(u) e^{-\int_a^u \varphi(v) dv} du} = 1. \quad (\text{B.3.9})$$

Proof. It suffices to show that

$$1 - \int_a^b \varphi(u) e^{-\int_a^u \varphi(v) dv} du = e^{-\int_a^b \varphi(u) du}. \quad (\text{B.3.10})$$

Let us define the definite integral $\Phi(t) = \int_a^t \varphi(u) du$. It is easy to verify that

$$\frac{d\Phi}{dt}(t) = \varphi(t), \quad \Phi(a) = 0. \quad (\text{B.3.11})$$

By substituting Eq. (B.3.11) into the integral of the left-hand side of Eq. (B.3.10) and using a change

of variables, we have

$$-\int_a^b \varphi(u) e^{-\int_a^u \varphi(v) dv} du = \int_a^b -\frac{d\Phi}{dt}(u) e^{-\Phi(u)} du \quad (\text{B.3.12a})$$

$$= \int_{-\Phi(a)}^{-\Phi(b)} e^u du \quad (\text{B.3.12b})$$

$$= e^{-\Phi(b)} - e^{-\Phi(a)}. \quad (\text{B.3.12c})$$

Substituting Eq. (B.3.12c) back into the left-hand side of Eq. (B.3.10), we obtain

$$1 - \int_a^b \varphi(u) e^{-\int_a^u \varphi(v) dv} du = 1 + e^{-\Phi(b)} - e^{-\Phi(a)} \quad (\text{B.3.13a})$$

$$= 1 + e^{-\Phi(b)} - e^0 \quad (\text{B.3.13b})$$

$$= 1 + e^{-\Phi(b)} - 1 \quad (\text{B.3.13c})$$

$$= e^{-\Phi(b)} \quad (\text{B.3.13d})$$

$$= e^{-\int_a^b \varphi(u) du}. \quad (\text{B.3.13e})$$

□

Lemma B.5 (Eq 2.22 [232])

Let \mathcal{T} be the set of timestamps from a Poisson process with flux function $\varphi(t)$. The conditional probability density function P_n of t_n given t_{n-1}, \dots, t_1 is

$$P_n(t_n | t_{n-1}, \dots, t_1) = \varphi(t_n) e^{-\int_{t_{n-1}}^{t_n} \varphi(u) du}. \quad (\text{B.3.14})$$

Lemma B.6 (III, T7α [32])

Consider a m -variate point process $\{t_n, m_n\}_{n \in \mathbb{N}}$ such that $\{t_n\}_{n \in \mathbb{N}}$ is a point process and $m_n \in \{0, \dots, m-1\}$. For $0 \leq i < m$, the point process $N^{(i)}(t)$ is defined by

$$N^{(i)}(t) = \sum_{n \geq 1} \mathbf{1}_{t_n \leq t} \mathbf{1}_{m_n = i}. \quad (\text{B.3.15})$$

If the processes $\Phi(t)$ and $A^{(i)}(t)$ are given by

$$\Phi(t) = \sum_{n \geq 0} \frac{P_{n+1}(t_{n+1} = t, m_{n+1} = i | \mathbf{F}_{t_n})}{1 - \int_0^{t-t_n} P_{n+1}(t_{n+1} = u + t_n | \mathbf{F}_{t_n}) du} \mathbf{1}_{t_n \leq t < t_{n+1}} \quad (\text{B.3.16})$$

and

$$A^{(i)}(t) = \int_0^t \Phi(u) du, \quad (\text{B.3.17})$$

then the process $M^{(i)}(t) = N^{(i)}(t) - A^{(i)}(t)$ is an \mathbf{F} -martingale.

Lemma B.7 (III, T7 β [32])

If $M^{(i)}(t) = N^{(i)}(t) - A^{(i)}(t)$ is a \mathbf{F} -martingale, then

$$E\left[\int_0^t p(u) dN^{(1)}(u)\right] = E\left[\int_0^t p(u) \Phi(u) du\right] \quad (\text{B.3.18})$$

for all non-negative continuous functions $p(t)$.

Corollary B.3

If $M^{(i)}(t) = N^{(i)}(t) - A^{(i)}(t)$ is a \mathbf{F} -martingale, then

$$E\left[\int_0^t p(u) dN^{(1)}(u)\right] = E\left[\int_0^t p(u) \Phi(u) du\right] \quad (\text{B.3.19})$$

for all continuous functions $p(t)$.

Proof. Let us express $p(t)$ as $p^+(t) - p^-(t)$ where $p^+(t)$ and $p^-(t)$ are non-negative continuous functions. The result follows from the linearity of expectation and Lemma B.7. \square

Lemma B.8

If the history $\mathbf{F}_{t_{j-1}}$ is the sequence $(t_n, m_n)_{n < j}$, the conditional probability P_j of the j^{th} timestamp arriving at time t and being assigned $m_j = 1$ is

$$P_j(t_j = t, m_j = 1 | \mathbf{F}_{t_{j-1}}) = \mathbf{1}_{t > t_i + D} \varphi(t) e^{-\int_{t_{j-1}}^t \varphi(u) du}. \quad (\text{B.3.20})$$

Proof. From the definition of conditional probability, we have:

$$P_j(t_j = t, m_j = 1 | \mathbf{F}_{t_{j-1}}) = P_j(t_j = t | \mathbf{F}_{t_{j-1}}) P_j(m_j = 1 | t_j = t, \mathbf{F}_{t_{j-1}}) \quad (\text{B.3.21})$$

We also have

$$P_j(t_j = t | \mathbf{F}_{t_{j-1}}) = \varphi(t) e^{-\int_{t_{j-1}}^t \varphi(u) du} \quad (\text{B.3.22})$$

and

$$P_j(m_j = 1 | t_j = t, \mathbf{F}_{t_{j-1}}) = \mathbf{1}_{t > t_i + D}. \quad (\text{B.3.23})$$

Equation (B.3.23) says that $m_j = 1$ with probability 1 if t_j arrives after the dead time and $m_j = 0$ otherwise. Substituting Eqs. (B.3.22) and (B.3.23) into Eq. (B.3.21) gives us Eq. (B.3.20). \square

B.4 Derivations using Monte Carlo estimators

In this section, we provide an alternative perspective on the flux probing equation when dead time is short enough relative to the photon inter-arrivals that it can be ignored. Under such conditions, photon detections can be modelled as an inhomogeneous Poisson process. Leveraging the independence of photon detections in such settings, the probing measurement can also be viewed as a

Monte Carlo estimator. The following proofs that I have included in my thesis are all **courtesy of anonymous Reviewer 1** for the paper submission version [262] of Chapters 4 and 5.

B.4.1 Alternate proof of Proposition 4.3

We show that treating probing as Monte Carlo estimation provides an alternate proof for the mean and variance of the probing measurement $p(\mathcal{T})$. We restate Proposition 4.3 (without the normality claims) here:

Proposition 4.3 (Mean and Variance of Probing Measurements)

The probing measurements $p(\mathcal{T})$ have mean $\langle p, \varphi \rangle$ and variance $\langle p^2, \varphi \rangle$.

Proof Sketch of Proposition 4.3.

The proof proceeds in two steps:

Step 1: We invoke the law of total expectation,

$$E[X] = E[E[X|Y]], \quad (\text{B.4.1})$$

where X and Y are random variables, to show that the mean is $\langle p, \varphi \rangle$.

Step 2: We invoke the law of total variance:

$$\text{var}[X] = \text{var}[E[X|Y]] + E[\text{var}[X|Y]], \quad (\text{B.4.2})$$

where X and Y are random variables, to show that the variance is $\langle p^2, \varphi \rangle$. \square

Proof of Proposition 4.3.

Step 1: When dead time is negligible, the sequence of timestamps \mathcal{T} in the interval $[0, t_{\text{exp}}]$ follows an inhomogeneous Poisson process with flux equal to $\varphi(t)$. The probability of an event occurring at time $t \in [0, t_{\text{exp}}]$ is equal to $\varphi(t)/E[N]$ [232] where $E[N]$ is the expected number of photon detections in the interval $[0, t_{\text{exp}}]$.

Applying Eq. (B.4.1),

$$E[p(\mathcal{T})] = E[E[p(\mathcal{T})|N]], \quad (\text{B.4.3})$$

where N is the number of photon detections in the interval $[0, t_{\text{exp}}]$. Since the timestamps in \mathcal{T} are independent of each other, for the inner expectation we have:

$$E[p(\mathcal{T})|N] = E\left[\sum_{t \in \mathcal{T}} p(t)|N\right] = \sum_{i=1}^N E[p(t_i)] = N \frac{\langle p, \varphi \rangle}{E[N]}. \quad (\text{B.4.4})$$

Therefore, the total expectation becomes:

$$E[p(\mathcal{T})] = E\left[N \frac{\langle p, \varphi \rangle}{E[N]}\right] = E[N] \frac{\langle p, \varphi \rangle}{E[N]} = \langle p, \varphi \rangle. \quad (\text{B.4.5})$$

Step 2: Similarly, applying Eq. (B.4.2), we obtain

$$\mathbf{var}[p(\mathcal{T})] = \mathbf{var}[E[p(\mathcal{T})|N]] + E[\mathbf{var}[p(\mathcal{T})|N]]. \quad (\text{B.4.6})$$

The first term of the RHS of Eq. (B.4.6) can be simplified using Eq. (B.4.4) and the fact that $\mathbf{var}[N] = E[N]$:

$$\mathbf{var}[E[p(\mathcal{T})|N]] = \mathbf{var}\left[N \frac{\langle p, \varphi \rangle}{E[N]}\right] = \mathbf{var}[N] \frac{\langle p, \varphi \rangle^2}{E[N]^2} = \frac{\langle p, \varphi \rangle^2}{E[N]}. \quad (\text{B.4.7})$$

For the second term of the RHS of Eq. (B.4.6), we have

$$\mathbf{var}[p(\mathcal{T})|N] = E[p^2(\mathcal{T})|N] - E[p(\mathcal{T})|N]^2 \quad (\text{B.4.8a})$$

$$= E[p^2(\mathcal{T})|N] - N^2 \frac{\langle p, \varphi \rangle^2}{E[N]^2}. \quad (\text{B.4.8b})$$

Expanding the first term of Eq. (B.4.8b),

$$E[p^2(\mathcal{T})|N] = E\left[\sum_{i=1}^N p(t_i) \cdot \sum_{i=1}^N p(t_i)\right] \quad (\text{B.4.9a})$$

$$= E\left[\sum_{i=1}^N p^2(t_i) + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} p(t_i)p(t_j)\right] \quad (\text{B.4.9b})$$

$$= E\left[\sum_{i=1}^N p^2(t_i)\right] + 2E\left[\sum_{i=1}^N \sum_{j=1}^{i-1} p(t_i)p(t_j)\right] \quad (\text{B.4.9c})$$

$$= \sum_{i=1}^N E[p^2(t_i)] + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} E[p(t_i)p(t_j)] \quad (\text{B.4.9d})$$

$$= N \frac{\langle p^2, \varphi \rangle}{E[N]} + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} E[p(t_i)] \cdot E[p(t_j)] \quad (\text{B.4.9e})$$

$$= N \frac{\langle p^2, \varphi \rangle}{E[N]} + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} \frac{\langle p, \varphi \rangle^2}{E[N]^2} \quad (\text{B.4.9f})$$

$$= N \frac{\langle p^2, \varphi \rangle}{E[N]} + \frac{2(N-1)N}{2} \frac{\langle p, \varphi \rangle^2}{E[N]^2} \quad (\text{B.4.9g})$$

$$= N \frac{\langle p^2, \varphi \rangle}{E[N]} + (N-1)N \frac{\langle p, \varphi \rangle^2}{E[N]^2}. \quad (\text{B.4.9h})$$

Combining Eqs. (B.4.9h) and (B.4.8b),

$$\mathbf{var}[p(\mathcal{T})|N] = N \frac{\langle p^2, \varphi \rangle}{E[N]} + (N-1)N \frac{\langle p, \varphi \rangle^2}{E[N]^2} - N^2 \frac{\langle p, \varphi \rangle^2}{E[N]^2} \quad (\text{B.4.10a})$$

$$= N \frac{\langle p^2, \varphi \rangle}{E[N]} - N \frac{\langle p, \varphi \rangle^2}{E[N]^2}. \quad (\text{B.4.10b})$$

The expectation of $\text{var}[p(\mathcal{T})|N]$ is

$$E[\text{var}[p(\mathcal{T})|N]] = E\left[N \frac{\langle p^2, \varphi \rangle}{E[N]} - N \frac{\langle p, \varphi \rangle^2}{E[N]^2}\right] \quad (\text{B.4.11a})$$

$$= \langle p^2, \varphi \rangle - \frac{\langle p, \varphi \rangle^2}{E[N]}. \quad (\text{B.4.11b})$$

Finally, by substituting Eqs. (B.4.11b) and (B.4.7) into Eq. (B.4.6), we obtain

$$\text{var}[p(\mathcal{T})] = \text{var}[E[p(\mathcal{T})|N]] + E[\text{var}[p(\mathcal{T})|N]] \quad (\text{B.4.12a})$$

$$= \frac{\langle p, \varphi \rangle^2}{E[N]} + \langle p^2, \varphi \rangle - \frac{\langle p, \varphi \rangle^2}{E[N]} \quad (\text{B.4.12b})$$

$$= \langle p^2, \varphi \rangle. \quad (\text{B.4.12c})$$

□

B.4.2 Alternate proof of Proposition B.1

We show that the Monte Carlo interpretation of flux probing also provides an alternate proof of Proposition B.1 for the covariance, which we restate here:

Proposition B.1 (Covariance of Probing Measurements)

For arbitrary probing functions $p_1(t)$, $p_2(t)$ we have that $\text{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = \int_0^{t_{\text{exp}}} p_1(u) p_2(u) \varphi(u) du$.

Proof Sketch of Proposition B.1.

Let $h(t) = p_1(t)p_2(t)$. We invoke the law of total covariance

$$\text{Cov}(X, Y) = \text{Cov}(E[X|Z], E[Y|Z]) + E[\text{Cov}(X, Y|Z)], \quad (\text{B.4.13})$$

and conditional covariance

$$\text{Cov}(X, Y|Z) = E[XY|Z] - E[X|Z]E[Y|Z], \quad (\text{B.4.14})$$

where X , Y , and Z are random variables, to prove the proposition. □

Proof of Proposition B.1.

Invoking Eq. (B.4.13), we have:

$$\text{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = \text{Cov}(E[p_1(\mathcal{T})|N], E[p_2(\mathcal{T})|N]) + E[\text{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N)]. \quad (\text{B.4.15})$$

The first term of the RHS of Eq. (B.4.15) can be simplified using Eq. (B.4.4):

$$\mathbf{Cov}(E[p_1(\mathcal{T})|N], E[p_2(\mathcal{T})|N]) = \mathbf{Cov}\left(N \frac{\langle p_1, \varphi \rangle}{E[N]}, N \frac{\langle p_2, \varphi \rangle}{E[N]}\right) \quad (\text{B.4.16a})$$

$$= \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2} \mathbf{Cov}(N, N) \quad (\text{B.4.16b})$$

$$= \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2} \underbrace{\mathbf{var}[N]}_{= E[N]} \quad (\text{B.4.16c})$$

$$= \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]}. \quad (\text{B.4.16d})$$

Equation (B.4.14) allows us to expand $\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N)$:

$$\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N) = E[p_1(\mathcal{T})p_2(\mathcal{T})|N] - E[p_1(\mathcal{T})|N]E[p_2(\mathcal{T})|N]. \quad (\text{B.4.17})$$

We now proceed by simplifying the first term of the RHS of Eq. (B.4.17):

$$E[p_1(\mathcal{T})p_2(\mathcal{T})|N] = E\left[\sum_{i=1}^N p_1(t_i) \cdot \sum_{i=1}^N p_2(t_i)\right] \quad (\text{B.4.18a})$$

$$= E\left[\sum_{i=1}^N p_1(t_i)p_2(t_i) + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} p_1(t_i)p_2(t_j)\right] \quad (\text{B.4.18b})$$

$$= E\left[\sum_{i=1}^N h(t_i)\right] + 2E\left[\sum_{i=1}^N \sum_{j=1}^{i-1} p_1(t_i)p_2(t_j)\right] \quad (\text{B.4.18c})$$

$$= \sum_{i=1}^N E[h(t_i)] + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} E[p_1(t_i)p_2(t_j)] \quad (\text{B.4.18d})$$

$$= N \frac{\langle h, \varphi \rangle}{E[N]} + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} E[p_1(t_i)] \cdot E[p_2(t_j)] \quad (\text{B.4.18e})$$

$$= N \frac{\langle h, \varphi \rangle}{E[N]} + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2} \quad (\text{B.4.18f})$$

$$= N \frac{\langle h, \varphi \rangle}{E[N]} + \frac{2(N-1)N}{2} \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2} \quad (\text{B.4.18g})$$

$$= N \frac{\langle h, \varphi \rangle}{E[N]} + (N-1)N \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2}, \quad (\text{B.4.18h})$$

where $h(t) = p_1(t)p_2(t)$.

From Eq. (B.4.4), it follows that the second term of Eq. (B.4.17) is equal to

$$E[p_1(\mathcal{T})|N]E[p_2(\mathcal{T})|N] = N^2 \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2}. \quad (\text{B.4.19})$$

We can now express $E[\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N)]$ as

$$E[\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N)] = E\left[N\frac{\langle h, \varphi \rangle}{E[N]} + (N-1)N\frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2} - N^2\frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2}\right] \quad (\text{B.4.20a})$$

$$= E\left[N\frac{\langle h, \varphi \rangle}{E[N]} - N\frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]^2}\right] \quad (\text{B.4.20b})$$

$$= \langle h, \varphi \rangle - \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]}. \quad (\text{B.4.20c})$$

Combining Eqs. (B.4.15), (B.4.16) and (B.4.20c), we finally obtain

$$\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})) = \mathbf{Cov}(E[p_1(\mathcal{T})|N], E[p_2(\mathcal{T})|N]) + E[\mathbf{Cov}(p_1(\mathcal{T}), p_2(\mathcal{T})|N)] \quad (\text{B.4.21a})$$

$$= \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]} + \langle h, \varphi \rangle - \frac{\langle p_1, \varphi \rangle \langle p_2, \varphi \rangle}{E[N]} \quad (\text{B.4.21b})$$

$$= \langle h, \varphi \rangle \quad (\text{B.4.21c})$$

$$= \int_0^{t_{\text{exp}}} h(u) \varphi(u) du \quad (\text{B.4.21d})$$

$$= \int_0^{t_{\text{exp}}} p_1(u) p_2(u) \varphi(u) du. \quad (\text{B.4.21e})$$

□



Opportunistic Single-Photon Time of Flight

C.1 Proofs and Derivations

C.1.1 Distribution of periodic flux estimator (proof of Proposition 6.1)

We start by restating our proposition:

Proposition 6.1 (Distribution of Periodic Flux Estimator)

If $\Phi(\mathbf{q}, f) = 0$ and $t \in \mathcal{T}$, then $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+T}{t_{\text{exp}}} + \langle p, \phi \rangle, \langle p^2, \phi \rangle\right)$. where $p(t) = \frac{2}{t_{\text{exp}}} \sum_{n=1}^N \cos(2\pi n f(t - t_k))$.

Proof Sketch of Proposition 6.1.

The proof proceeds in two steps:

Step 1: We show that $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})$ corresponds to flux probing with the following probing function $p(t)$:

$$p(t) = \frac{2}{t_{\text{exp}}} \sum_{n=1}^N \cos(2\pi n f(t - t_k)).$$

Step 2: We characterize the distribution $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})$ using Proposition 4.3 that we restate for convenience:

Proposition 4.3 (Distribution of Probing Measurements)

The probing measurements $p(\mathcal{T})$ are approximately normally distributed with mean $\langle p, \phi \rangle$ and variance $\langle p^2, \phi \rangle$.

Proof of Proposition 6.1.

Step 1: Let us consider a single timestamp in \mathcal{T} and denote it as t_k , we index with subscript k to differentiate this timestamp from the rest in \mathcal{T} . For simplicity, let $T = |\mathcal{T}|$ be the total number of

timestamps and let $\mathcal{T}_- = \mathcal{T} \setminus \{t_k\}$ be the set of timestamps excluding t_k . Then we have

$$\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) = \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}) \exp(j2\pi n f t_k) \quad (\text{C.1.1a})$$

$$= \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}_-) \exp(j2\pi n f t_k) + \sum_{n=-N}^N \underbrace{\hat{\Phi}(\mathbf{q}, nf, \{t_k\})}_{= \exp(-j2\pi n f t_k) / t_{\text{exp}}} \exp(j2\pi n f t_k) \quad (\text{C.1.1b})$$

$$= \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}_-) \exp(j2\pi n f t_k) + \sum_{n=-N}^N \frac{1}{t_{\text{exp}}} \quad (\text{C.1.1c})$$

$$= \sum_{n=-N}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}_-) \exp(j2\pi n f t_k) + \frac{2N+1}{t_{\text{exp}}} \quad (\text{C.1.1d})$$

$$= \sum_{n=1}^N \hat{\Phi}(\mathbf{q}, nf, \mathcal{T}_-) \exp(j2\pi n f t_k) + \sum_{n=1}^N \hat{\Phi}(\mathbf{q}, -nf, \mathcal{T}_-) \exp(-j2\pi n f t_k) + \underbrace{\frac{T-1}{t_{\text{exp}}}}_{\text{from } n=0} + \frac{2N+1}{t_{\text{exp}}}. \quad (\text{C.1.1e})$$

$$(\text{C.1.1f})$$

Let us consider one term in the summation corresponding to the n^{th} order harmonic, nf .

$$\hat{\Phi}(\mathbf{q}, nf, \mathcal{T}_-) \exp(j2\pi n f t_k) + \hat{\Phi}(\mathbf{q}, -nf, \mathcal{T}_-) \exp(-j2\pi n f t_k) = \frac{1}{t_{\text{exp}}} \sum_{t \in \mathcal{T}_-} \exp(j2\pi n f (t_k - t)) + \exp(-j2\pi n f (t_k - t)) \quad (\text{C.1.2a})$$

$$= \frac{2}{t_{\text{exp}}} \sum_{t \in \mathcal{T}_-} \cos(2\pi n f (t - t_k)). \quad (\text{C.1.2b})$$

$$(\text{C.1.2c})$$

Let us define the following function $p(t)$:

$$p(t) = \frac{2}{t_{\text{exp}}} \sum_{n=1}^N \cos(2\pi n f (t - t_k)). \quad (\text{C.1.3})$$

Therefore, we can use $p(t)$ into simplify Eq. (C.1.1):

$$\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) = \frac{2N+T}{t_{\text{exp}}} + \sum_{t \in \mathcal{T}_-} p(t). \quad (\text{C.1.4})$$

Step 2: From Proposition 4.3, we know that:

$$\sum_{t \in \mathcal{T}_-} p(t) \sim \mathcal{N}(\langle p, \phi \rangle, \langle p^2, \phi \rangle).^1 \quad (\text{C.1.5})$$

Substituting Eq. (C.1.5) into Eq. (C.1.4) gives us:

$$\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+T}{t_{\text{exp}}} + \langle p, \phi \rangle, \langle p^2, \phi \rangle\right). \quad (\text{C.1.6})$$

Equation C.1.6 tells us therefore that we can model $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})$ as a normal distribution with the desired mean and variance. \square

C.1.2 Proof of Corollary 6.1

We start by restating our corollary:

Corollary 6.1 (Periodic Flux Estimator under Negligible Dead Time)

If $\Phi(\mathbf{q}, f) = 0$, $t \in t_k$, and dead time is negligible, then $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+|T|}{t_{\text{exp}}}, \frac{2N|T|}{t_{\text{exp}}^2}\right)$.

¹We assume that removing t_k from \mathcal{T} affects its underlying flux function ϕ negligibly.

Proof. We already know that $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+T}{t_{\text{exp}}} + \langle p, \phi \rangle, \langle p^2, \phi \rangle\right)$ from Proposition 6.1.

In general, when f is a noise frequency, we expect that $\Phi(\mathbf{q}, nf) = 0$ for all $n \in \mathbb{N}$. Therefore, without loss of generality, we assume $\Phi(\mathbf{q}, nf) = 0$ for all $n \leq 2N$.

Under negligible dead time, we can derive closed-form expressions for the mean and variance of $\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})$.² We start by simplifying $\langle p, \phi \rangle$:

$$\langle p, \phi \rangle = \int_0^{t_{\text{exp}}} p(t)\phi(t)dt \quad (\text{C.1.7a})$$

$$= \frac{2}{t_{\text{exp}}} \sum_{n=1}^N \int_0^{t_{\text{exp}}} \cos(2\pi nf(t - t_k))\phi(t)dt \quad (\text{C.1.7b})$$

$$= \frac{1}{t_{\text{exp}}} \sum_{n=1}^N \Phi(\mathbf{q}, nf) + \Phi(\mathbf{q}, -nf) \quad (\text{C.1.7c})$$

$$= 0. \quad (\text{C.1.7d})$$

Now, we simplify $\langle p^2, \phi \rangle$:

$$\langle p^2, \phi \rangle = \int_0^{t_{\text{exp}}} p^2(t)\phi(t)dt \quad (\text{C.1.8a})$$

$$= \frac{4}{t_{\text{exp}}^2} \sum_{n_1=1}^N \sum_{n_2=1}^N \int_0^{t_{\text{exp}}} \underbrace{\cos(2\pi n_1 f(t - t_k)) \cos(2\pi n_2 f(t - t_k))}_{\text{use trig identity for } \cos A \cos B} \phi(t)dt \quad (\text{C.1.8b})$$

$$= \frac{2}{t_{\text{exp}}^2} \sum_{n_1=1}^N \sum_{n_2=1}^N \int_0^{t_{\text{exp}}} \underbrace{\cos(2\pi(n_1 - n_2)f(t - t_k))\phi(t) + \cos(2\pi(n_1 + n_2)f(t - t_k))\phi(t)}_{= 0 \text{ when } n_1 \neq n_2} dt \quad (\text{C.1.8c})$$

$$= \frac{2}{t_{\text{exp}}^2} \sum_{n=1}^N \underbrace{\int_0^{t_{\text{exp}}} \phi(t)dt}_{\approx T} \quad (\text{C.1.8d})$$

$$= \frac{2NT}{t_{\text{exp}}^2}. \quad (\text{C.1.8e})$$

Therefore, under negligible dead time, we have:

$$\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \sim \mathcal{N}\left(\frac{2N+T}{t_{\text{exp}}}, \frac{2NT}{t_{\text{exp}}^2}\right). \quad (\text{C.1.9})$$

□

C.1.3 Derivation of CFAR sinc detector threshold

We start by restating the definition of frequency detection:

Definition 6.1 (CFAR sinc comb detector)

To achieve a constant probability of false alarm p , a candidate frequency f is detected if:

$$\mathcal{T}_{\text{dom}} = \{t \in \mathcal{T} \mid \hat{\phi}_f(t) > \Delta\} \neq \emptyset,$$

²Under non-negligible dead time, our noise model in Eq. (C.1.6) still holds, however we do not have a closed-form expression for the mean and variance. It may be possible to leverage the sample mean and variance for the CFAR detection. Nevertheless, we find in our experiments and simulations that even with an SBR of 0.01, that our analytical noise model is valid.

where $\Delta = \frac{2N+|\mathcal{T}|}{t_{\text{exp}}} + \text{CDF}_{\mathcal{N}}^{-1}(1-p) \frac{\sqrt{2N|\mathcal{T}|}}{t_{\text{exp}}}$, the CFAR sinc detector threshold, N is the number of harmonics, and t_{exp} is the exposure time.

The main intuition of Definition 6.1 is that a frequency f is only detected when it is statistically significant, *i.e.*, f is a noise frequency under our null hypothesis. Given that we know the distribution of our flux estimates and its parameters under the null hypothesis from Eq. (C.1.9), we set our threshold Δ according to the z-score needed for a one-sided Z-test [234] to reject our null hypothesis with a significance level of $1 - p$. In our detection scheme, we set $p = 1/|\mathcal{F}_{\text{cand}}|T$, *i.e.*, the likelihood of observing at least one outlier sample in the total number of flux estimates we evaluate. For reconstructed flux values corresponding to noise frequencies, this can be expressed as follows:

$$P(\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) > \Delta) = p \quad (\text{C.1.10a})$$

$$P(\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) \leq \Delta) = 1 - p \quad (\text{C.1.10b})$$

$$P\left(\frac{\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T}) - \mu}{\sigma} \leq \frac{\Delta - \mu}{\sigma}\right) = 1 - p, \quad (\text{C.1.10c})$$

where $\mu = E[\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})]$ and $\sigma^2 = \text{var}[\hat{\phi}_f(\mathbf{q}, t_k, \mathcal{T})]$. Substituting Eq. (C.1.9) into Eq. (C.1.10) and inverting both sides gives:

$$\Delta = \frac{2N+|\mathcal{T}|}{t_{\text{exp}}} + \text{CDF}_{\mathcal{N}}^{-1}(1-p) \frac{\sqrt{2N|\mathcal{T}|}}{t_{\text{exp}}}. \quad (\text{C.1.11})$$

where $\text{CDF}_{\mathcal{N}}^{-1}$ denotes the inverse cumulative distribution function of the standard normal distribution.

Bibliography

- [1] IEC 60050-845:2020. *International Electrotechnical Vocabulary (IEV). Part 845: Lighting*. Tech. rep. Geneva, CH: International Electrotechnical Commission, 2020. URL: <https://webstore.iec.ch/en/publica>
- [2] ISO 80000-7:2019(E). *Quantities and units. Part 7: Light*. Tech. rep. Geneva, CH: International Organization for Standardization, 2019. URL: <https://www.iso.org/standard/64977.html>.
- [3] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.
- [4] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. “Capturing the human figure through a wall”. In: *ACM Trans. Graph. (TOG)* 34.6 (2015), pp. 1–13.
- [5] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. “3D tracking via body radio reflections”. In: *Proc. USENIX Symp. Networked Syst. Des. Implement. (NSDI)*. 2014, pp. 317–329.
- [6] Fadel Adib and Dina Katabi. “See through walls with WiFi!” In: *Proc. ACM SIGCOMM Conf.* 2013, pp. 75–86.
- [7] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. “Smart homes that monitor breathing and heart rate”. In: *Proc. ACM Conf. on Human Factors in Computing Systems*. 2015, pp. 837–846.
- [8] Faisal Ahmed, Miguel Heredia Conde, Paula López Martínez, Thomas Kerstein, and Bernd Buxbaum. “Pseudo-passive time-of-flight imaging: Simultaneous illumination, communication, and 3D sensing”. In: *IEEE Sens. J.* 22.21 (2022), pp. 21218–21231.
- [9] Hüseyin Akcan, Vassil Kriakov, Hervé Brönnimann, and Alex Delis. “GPS-Free node localization in mobile wireless sensor networks”. In: *Proc. ACM Int. Workshop on Data Engineering for Wireless and Mobile Access*. 2006, pp. 35–42.
- [10] ams OSRAM Group. *TMF8820/21/28 Time-of-Flight Sensor Datasheet*. Accessed: 2024-11-14. 2024. URL: <https://look.ams-osram.com/m/52236c476132a095/original/TMF8820-21-28-Multizone-Time->
- [11] Apple. *iPhone 16*. <https://www.apple.com/ca/iphone-16/>. 2024.
- [12] Edward Victor Appleton and Miles AF Barnett. “On some direct evidence for downward atmospheric reflection of electric rays”. In: *Proc. R. Soc. London, Ser. A* 109.752 (1925), pp. 621–641.
- [13] Eitam Arnon, Shlomo Cain, Assaf Uzan, Ran Nathan, Orr Spiegel, and Sivan Toledo. “Robust Time-of-Arrival Location Estimation Algorithms for Wildlife Tracking”. In: *Sensors* 23.23 (2023), p. 9460.
- [14] Seung-Hwan Baek, Noah Walsh, Ilya Chugunov, Zheng Shi, and Felix Heide. “Centimeter-Wave Free-Space Neural Time-of-Flight Imaging”. In: *ACM Trans. Graph. (SIGGRAPH)* 42.1 (2022), pp. 1–18.
- [15] Rashmi Bajaj, Samantha Lalinda Ranaweera, and Dharma P Agrawal. “GPS: location-tracking technology”. In: *Computer* 35.4 (2002), pp. 92–94.

- [16] Henrik Baktoft, Karl Øystein Gjelland, Finn Økland, and Uffe Høgsbro Thygesen. "Positioning of aquatic animals based on time-of-arrival and random walk models using YAPS (Yet Another Positioning Solver)". In: *Sci. Rep.* 7.1 (2017), p. 14294.
- [17] Cyrus Bamji, John Godbaz, Minseok Oh, Swati Mehta, Andrew Payne, Sergio Ortiz, Satyadev Nagaraja, Travis Perry, and Barry Thompson. "A review of indirect time-of-flight technologies". In: *IEEE Trans. on Electron. Devices* 69.6 (2022), pp. 2779–2793.
- [18] Cyrus S Bamji, Swati Mehta, Barry Thompson, Tamer Elkhatib, Stefan Wurster, Onur Akkaya, Andrew Payne, John Godbaz, Mike Fenton, Vijay Rajasekaran, et al. "1Mpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3 μ m global shutter pixels and analog binning". In: *IEEE Int. Solid-State Circuits Conf. (ISSCC)*. IEEE. 2018, pp. 94–96.
- [19] Cyrus S Bamji, Patrick O'Connor, Tamer Elkhatib, Swati Mehta, Barry Thompson, Lawrence A Prather, Dane Snow, Onur Can Akkaya, Andy Daniel, Andrew D Payne, Travis Perry, Mike Fenton, and Vei-Han Chan. "A 0.13 μ m CMOS System-on-Chip for a 512x424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC". In: *IEEE J. Solid-State Circuits* 50.1 (2015), pp. 303–319.
- [20] Richard G Baraniuk, Thomas Goldstein, Aswin C Sankaranarayanan, Christoph Studer, Ashok Veeraraghavan, and Michael B Wakin. "Compressive Video Sensing: Algorithms, architectures, and applications". In: *IEEE Signal Proc. Mag.* 34.1 (2017), pp. 52–66.
- [21] Alexander H. Barnett, Jeremy Magland, and Ludvig Af Klinteberg. "A Parallel Nonuniform Fast Fourier Transform Library Based on an "Exponential of Semicircle" Kernel". In: *SIAM J. Sci. Comput.* 41.5 (2019), pp. C479–C504.
- [22] Jonathan T Barron. "A general and adaptive robust loss function". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 4331–4339.
- [23] M. S. Bartlett. "The Spectral Analysis of Point Processes". In: *J. R. Stat. Soc., B: Stat. Methodol.* 25.2 (1963), pp. 264–296.
- [24] Maurice S Bartlett. "Periodogram analysis and continuous spectra". In: *Biometrika* 37.1/2 (1950), pp. 1–16.
- [25] Bryce E Bayer. "Color imaging array". 1976.
- [26] Wolfgang Becker. "Fluorescence lifetime imaging—techniques and applications". In: *J. Microsc.* 247.2 (2012), pp. 119–136.
- [27] Paul J Besl and Neil D McKay. "Method for registration of 3-D shapes". In: *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie. 1992, pp. 586–606.
- [28] Benedikt Bitterli. *Rendering resources*. <https://benedikt-bitterli.me/resources/>. 2016.
- [29] Robert W Boyd. *Radiometry and the Detection of Optical Radiation*. Wiley, 1983.
- [30] Willard S Boyle and George E Smith. "Charge coupled semiconductor devices". In: *Bell System Technical Journal* 49.4 (1970), pp. 587–593.
- [31] D. J. Bradley, B. Liddy, and W. E. Sleat. "Direct linear measurement of ultrashort light pulses with a picosecond streak camera". In: *Opt. Commun.* 2.8 (1971), pp. 391–395.
- [32] Pierre Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer, 1981.

- [33] Christopher M Brown. "Multiplex Imaging and Random Arrays". PhD thesis. University of Chicago, 1972.
- [34] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. "Non-line-of-sight imaging using a time-gated single photon avalanche diode". In: *Opt. Express* 23.16 (2015), pp. 20997–21011.
- [35] Clara Callenberg, Felix Heide, Gordon Wetzstein, and Matthias B Hullin. "Snapshot difference imaging using correlation time-of-flight sensors". In: *Proc. SIGGRAPH Asia*. Bangkok: ACM, 2017.
- [36] Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B. Hullin. "Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging". In: *ACM Trans. Graph. (SIGGRAPH)* 40.4 (July 2021).
- [37] W Carey and N Yen. "The formation of a synthetic aperture with towed hydrophones". In: *J. Acoust. Soc. Am.* 75.S1 (1984), S62–S63.
- [38] G. Clifford Carter. "Time delay estimation for passive sonar signal processing". In: *IEEE Trans. Acoust. Speech Signal Process.* 29.3 (1981), pp. 463–470.
- [39] T. C. Carusone, D. Johns, and K. Martin. *Analog Integrated Circuit Design*. Wiley, 2011.
- [40] Ayan Chakrabarti and Kalyan Sunkavalli. "Single-Image RGB Photometric Stereo with Spatially-Varying Albedo". In: *International Conference on 3D Vision (3DV)*. 2016.
- [41] Ningyuan Chen, Donald K. K. Lee, and Sahand N. Negahban. "Super-resolution estimation of cyclic arrival rates". In: *Ann. Stat.* 47.3 (2019), pp. 1754–1775.
- [42] Wei-Yu Chen, Aswin C Sankaranarayanan, Anat Levin, and Matthew O'Toole. "Coherence As Texture-Passive Textureless 3D Reconstruction by Self-interference". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2024, pp. 25058–25066.
- [43] Yan Chen, Joachim D. Müller, Peter T. C. So, and Enrico Gratton. "The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy". In: *Biophys. J.* 77.1 (1999), pp. 553–567.
- [44] Yiheng Chi, Abhiram Gnanasambandam, Vladlen Koltun, and Stanley H Chan. "Dynamic low-light imaging with quanta image sensors". In: *Proc. Eur. Conf. on Computer Vision (ECCV)*. Springer. 2020, pp. 122–138.
- [45] Erhan Çinlar. *Introduction to Stochastic Processes*. Prentice Hall, 1975.
- [46] Samuel N Cohen and Robert James Elliott. *Stochastic Calculus and Applications*. Vol. 2. Birkhäuser New York, 2015.
- [47] O. Cossairt, M. Gupta, and S. K. Nayar. "When Does Computational Imaging Improve Performance?" In: *IEEE Transactions on Image Processing* (2013).
- [48] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. "Visual vibrometry: Estimating material properties from small motion in video". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 5335–5343.

- [49] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham Mysore, Fredo Durand, and William T. Freeman. "The Visual Microphone: Passive Recovery of Sound from Video". In: *ACM Trans. Graph. (SIGGRAPH)* 33.4 (2014), 79:1–79:10.
- [50] Joseph L Doob. *Stochastic Processes*. Wiley New York, 1953.
- [51] PM Echternach, BJ Pepper, T Reck, and CM Bradford. "Single photon detection of 1.5 THz radiation with the quantum capacitance detector". In: *Nature Astronomy* 2.1 (2018), pp. 90–97.
- [52] A Einstein. "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt". In: *Annalen der Physik* 322.6 (1905), pp. 132–148.
- [53] Eyer, L. and Bartholdi, P. "Variable stars: Which Nyquist frequency?" In: *Astron. Astrophys. Suppl. Ser.* 135.1 (1999), pp. 1–3. URL: <https://doi.org/10.1051/aas:1999102>.
- [54] Zahid Farid, Rosdiadee Nordin, and Mahamod Ismail. "Recent advances in wireless indoor localization techniques and system". In: *J. Comput. Networks Commun.* 2013.1 (2013), p. 185138.
- [55] Martin A Fischler and Robert C Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24.6 (1981), pp. 381–395.
- [56] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data". In: *IEEE transactions on image processing* 17.10 (2008), pp. 1737–1754.
- [57] David A Forsyth and Jean Ponce. *Computer vision: a modern approach*. Prentice Hall, 2002.
- [58] Tara Fortier and Esther Baumann. "20 years of developments in optical frequency comb technology and applications". In: *Commun. Phys.* 2.1 (2019), p. 153. eprint: 1909.05384.
- [59] Eric R Fossum. "CMOS image sensors: Electronic camera-on-a-chip". In: *IEEE transactions on electron devices* 44.10 (1997), pp. 1689–1698.
- [60] Eric R Fossum and Donald B Hondongwa. "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors". In: *IEEE J. Electron Devices* 2.3 (2014), pp. 33–43.
- [61] Eric R Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. "The quanta image sensor: Every photon counts". In: *Sensors* 16.8 (2016), p. 1260.
- [62] Eric R Fossum and Robert Nixon. *CMOS active pixel sensor type imaging system on a chip*. US Patent 5,841,126. Nov. 24, 1998.
- [63] Rosalind E Franklin and Raymond G Gosling. "Molecular configuration in sodium thymonucleate". In: *Nature* 171.4356 (1953), pp. 740–741.
- [64] Graham Fyffe, Xueming Yu, and Paul E. Debevec. "Single-shot photometric stereo by spectral multiplexing". In: *ICCP*. 2011.
- [65] Álvaro Gajardo and Hans-Georg Müller. "Cox Point Process Regression". In: *IEEE Trans. Inf. Theory* 68.2 (2022), pp. 1133–1156.
- [66] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang. "Single-shot compressed ultrafast photography at one hundred billion frames per second". In: *Nature* 516.7529 (2014), pp. 74–77.

- [67] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. “Single-photon sensitive light-in-flight imaging”. In: *Nat. Commun.* 6.6021 (2015).
- [68] R. C. Geary. “The Frequency Distribution of the Quotient of Two Normal Variates”. In: *Journal of the Royal Statistical Society* 93.3 (1930), pp. 442–446.
- [69] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Fredo Durand. “Deep joint demosaicking and denoising”. In: *Proc. ACM SIGGRAPH Asia*. 2016.
- [70] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.
- [71] Google. *Pixel 9 Pro*. https://store.google.com/ca/product/pixel_9_pro. 2024.
- [72] Francesco Gramuglia, Ming-Lo Wu, Claudio Bruschini, Myung-Jae Lee, and Edoardo Charbon. “A low-noise CMOS SPAD pixel with 12.1 ps SPTR and 3 ns dead time”. In: *IEEE J. Sel. Top. Quantum Electron.* 28.2: Optical Detectors (2021), pp. 1–9.
- [73] Hugh D Griffiths and Christopher J Baker. *An introduction to passive radar*. Artech House, 2022.
- [74] Hugh D Griffiths and NRW Long. “Television-based bistatic radar”. In: *Proc. Inst. Electr. Eng. F*. Vol. 133. 7. 1986, pp. 649–657.
- [75] Nicholas Griffiths Hugh Dand Willis. “Klein Heidelberg?The first modern bistatic radar system”. In: *IEEE Trans. Aerosp. Electron. Syst.* 46.4 (2010), pp. 1571–1588.
- [76] Sara Grollius, Andre Buchner, Manuel Ligges, and Anton Grabmaier. “Probability of unrecognized LiDAR interference for TCSPC LiDAR”. In: *IEEE Sens. J.* 22.13 (2022), pp. 12976–12986.
- [77] Anant Gupta, Atul Ingle, and Mohit Gupta. “Asynchronous single-photon 3D imaging”. In: *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV)*. 2019, pp. 7909–7918.
- [78] M Gupta and Shree K Nayar. “Micro Phase Shifting”. In: *Proc. CVPR*. Providence: IEEE, 2012.
- [79] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. “Phasor Imaging: A Generalization of Correlation-Based Time-of-Flight Imaging”. In: *ACM Trans. Graph. (SIGGRAPH)* 34.5 (2015), pp. 1–18.
- [80] Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, and Andreas Velten. “Compressive single-photon 3D cameras”. In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 17854–17864.
- [81] Robert H Hadfield. “Single-photon detectors for optical quantum information applications”. In: *Nature photonics* 3.12 (2009), pp. 696–705.
- [82] Edward E Hammer. “High frequency characteristics of fluorescent lamps up to 500 kHz”. In: *J. Illum. Eng. Soc.* 16.1 (1987), pp. 52–61.
- [83] Daniel Hampf, Ewan Schafer, Fabian Sproll, Toshimichi Otsubo, Paul Wagner, and Wolfgang Riede. “Satellite laser ranging at 100 kHz pulse repetition rate”. In: *CEAS Space J.* 11.4 (2019), pp. 363–370.

- [84] Glenn E Healey and Raghava Kondepudy. "Radiometric CCD camera calibration and noise estimation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.3 (1994), pp. 267–276.
- [85] Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. "Sub-picosecond photon-efficient 3D imaging using single-photon sensors". In: *Sci. Rep.* 8.1 (2018), p. 17726.
- [86] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. "Low-budget transient imaging using photonic mixer devices". In: *ACM Trans. Graph. (SIGGRAPH)* 32.4 (2013), pp. 1–10.
- [87] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajak, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, Jan Kautz, and Kari Pulli. "FlexISP: a flexible camera image processing framework". In: *Proc. SIGGRAPH Asia*. Shenzhen: ACM, 2014.
- [88] R Heintzmann, Q S Hanley, D Arndt-Jovin, and T M Jovin. "A dual path programmable array microscope (PAM): simultaneous acquisition of conjugate and non-conjugate images". In: *J. Microscopy* 204.2 (2001), pp. 119–135.
- [89] C Hernandez, G Vogiatzis, G J Brostow, B Stenger, and R Cipolla. "Non-rigid Photometric Stereo with Colored Lights". In: *Proc. ICCV*. Rio de Janeiro: IEEE, 2007.
- [90] Barmak Heshmat, Genevieve Gariepy, Jonathan Leach, Ramesh Raskar, and Daniele Faccio. "SPAD cameras for biomedical imaging: Promise and problems". In: *Conf. Lasers Electro-Opt. (CLEO)*. 2016, pp. 1–2.
- [91] Tomoki Hirata, Hironobu Murata, Hideaki Matsuda, Yojiro Tezuka, and Shiro Tsunai. "A 1-inch 17Mpixel 1000fps Block-Controlled Coded-Exposure Back-Illuminated Stacked CMOS Image Sensor for Computational Imaging and Adaptive Dynamic Range Control". In: *2021 IEEE International Solid-State Circuits Conference (ISSCC)*. Vol. 64. 2021, pp. 120–122.
- [92] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. "Video from a single coded exposure photograph using a learned over-complete dictionary". In: *Proc. ICCV*. Barcelona: IEEE, 2011, pp. 287–294.
- [93] Berthold Horn. *Robot vision*. 1986.
- [94] Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. "Passive Inter-Photon Imaging". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 8585–8595.
- [95] Atul Ingle, Andreas Velten, and Mohit Gupta. "High Flux Passive Imaging with Single-Photon Sensors". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 6753–6762.
- [96] Mark A Itzler, R Ben-Michael, C-F Hsu, Krystyna Slomkowski, Alberto Tosi, Sergio Cova, Franco Zappa, and Radu Ispasoiu. "Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications". In: *J. Mod. Opt.* 54.2-3 (2007), pp. 283–304.
- [97] Jinbeum Jang, Yoonjong Yoo, Jongheon Kim, and Joonki Paik. "Sensor-Based Auto-Focusing System Using Multi-Scale Feature Extraction and Phase Correlation Matching". In: *Sensors* 15.3 (2015).

- [98] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical Methods for Financial Markets*. Springer London, 2009.
- [99] Achuta Kadambi, Ayush Bhandari, Refael Whyte, Adrian Dorrington, and Ramesh Raskar. "Demultiplexing illumination via low cost sensing and nanosecond coding". In: *Proc. ICCP*. Santa Clara: IEEE, 2014.
- [100] Achuta Kadambi and Ramesh Raskar. "Rethinking machine vision time of flight with GHz heterodyning". In: *IEEE Access* 5 (2017), pp. 26211–26223.
- [101] Huifeng Ke, Navid Sarhangnejad, Rahul Gulve, Zhengfan Xia, Nikita Gusev, Nikola Katic, Kiriakos N Kutulakos, and Roman Genov. "Extending Image Sensor Dynamic Range by Scene-aware Pixelwise-adaptive Coded Exposure". In: 2019.
- [102] Maurice G Kendall. "On the analysis of oscillatory time-series". In: *J. R. Stat. Soc.* 108.1/2 (1945), pp. 93–141.
- [103] H Kim, B Wilburn, and M. Ben-Ezra. "Photometric stereo for dynamic surface orientations". In: *ECCV 2010, LNCS, vol. 6311*. Ed. by Kostas Daniilidis, Petros Maragos, and Nikos Paragios. Heidelberg: Springer, 2010, pp. 59–72.
- [104] Hanme Kim, Stefan Leutenegger, and Andrew J Davison. "Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera". In: *ECCV 2016, LNCS, vol. 9910*. Ed. by Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling. Cham: Springer, 2016, pp. 349–364.
- [105] Diederik P Kingma. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).
- [106] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco NC Wong, Jeffrey H Shapiro, and Vivek K Goyal. "First-photon imaging". In: *Sci.* 343.6166 (2014), pp. 58–61.
- [107] Ghulam Ahmed Kirmani. "Computational Time-resolved Imaging". PhD thesis. Massachusetts Institute of Technology, 2015.
- [108] Arash Komaee. "Maximum likelihood and minimum mean squared error estimations for measurement of light intensity". In: *Proc. Conf. on Information Sciences and Systems (CISS)*. 2010, pp. 1–6.
- [109] Hamed Komari Alaie and Hassan Farsi. "Passive sonar target detection using statistical classifier and adaptive threshold". In: *Appl. Sci.* 8.1 (2018), p. 61.
- [110] Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. "Passive Micron-scale Time-of-Flight with Sunlight Interferometry". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 4139–4149.
- [111] Hiroyuki Kubo, Suren Jayasuriya, Takafumi Iwaguchi, Takuya Funatomi, Yasuhiro Mukaigawa, and Srinivasa G Narasimhan. "Acquiring and characterizing plane-to-ray indirect light transport". In: *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2018, pp. 1–10.
- [112] Jayakanth Kunhoth, AbdelGhani Karkar, Somaya Al-Maadeed, and Abdulla Al-Ali. "Indoor positioning and wayfinding systems: a survey". In: *Hum.-centric Comput. Inf. Sci.* 10 (2020), pp. 1–41.

- [113] Heiner Kuschel. "Approaching 80 years of passive radar". In: *Int. Conf. Radar*. IEEE. 2013, pp. 213–217.
- [114] Robert Lange and Peter Seitz. "Solid-state time-of-flight range camera". In: *IEEE J. Quantum Electron.* 37.3 (2001), pp. 390–397.
- [115] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.
- [116] Jongho Lee, Atul Ingle, Jenu V Chacko, Kevin W Eliceiri, and Mohit Gupta. "CASPI: collaborative photon processing for active single-photon imaging". In: *Nat. Commun.* 14.1 (2023), p. 3158.
- [117] PA W Lewis and Gerald S Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Nav. Res. Logist. Q.* 26.3 (1979), pp. 403–413.
- [118] Fengqiang Li, Florian Willomitzer, Muralidhar Madabhushi Balaji, Prasanna Rangarajan, and Oliver Cossairt. "Exploiting wavelength diversity for high resolution time-of-flight 3D imaging". In: *IEEE Trans. Pattern Anal. Machine Intell.* 43.7 (2021), pp. 2193–2205.
- [119] Fengqiang Li, Florian Willomitzer, Prasanna Rangarajan, Mohit Gupta, Andreas Velten, and Oliver Cossairt. "SH-ToF: Micro resolution time-of-flight imaging with superheterodyne interferometry". In: *Proc. IEEE Int. Conf. on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–10.
- [120] You Li and Javier Ibanez-Guzman. "Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems". In: *IEEE Signal Process Mag.* 37.4 (2020), pp. 50–61.
- [121] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. "A 128×128 120 dB 15 μ s Latency Asynchronous Temporal Contrast Vision Sensor". In: *IEEE J. Solid-State Circuits* 43.2 (2008), pp. 566–576.
- [122] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. "A 128×128 120 dB 15 μ s Latency Asynchronous Temporal Contrast Vision Sensor". In: *IEEE Journal of Solid-State Circuits* 43.2 (2008), pp. 566–576.
- [123] Keh-Shin Lii and Murray Rosenblatt. "Estimation for almost periodic processes". In: *Ann. Stat.* 34.3 (2006), pp. 1115–1139.
- [124] Yi-Chun Lin, Yi-Ting Cheng, Tian Zhou, Radhika Ravi, Seyyed Meghdad Hasheminasab, John Evan Flatt, Cary Troy, and Ayman Habib. "Evaluation of UAV LiDAR for mapping coastal environments". In: *Remote Sens.* 11.24 (2019), p. 2893.
- [125] David B Lindell, Matthew O'Toole, and Gordon Wetzstein. "Single-photon 3D imaging with deep sensor fusion." In: *ACM Trans. Graph. (SIGGRAPH)* 37.4 (2018), p. 113.
- [126] David B Lindell, Matthew O'Toole, and Gordon Wetzstein. "Towards transient imaging at interactive rates with single-photon detectors". In: *Proc. IEEE Int. Conf. on Computational Photography (ICCP)*. 2018, pp. 1–8.
- [127] Ce Liu, William T Freeman, Richard Szeliski, and Sing Bing Kang. "Noise estimation from a single image". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 1. IEEE. 2006, pp. 901–908.

- [128] Fen Liu, Jing Liu, Yuqing Yin, Wenhan Wang, Donghai Hu, Pengpeng Chen, and Qiang Niu. "Survey on WiFi-based indoor positioning techniques". In: *IET Commun.* 14.9 (2020), pp. 1372–1383.
- [129] Jingdan Liu, Miguel Marquez, Yingming Lai, Heide Ibrahim, Katherine Légaré, Philippe Lassonde, Xianglei Liu, Michel Hehn, Stéphane Mangin, Grégory Malinowski, et al. "Swept coded aperture real-time femtophotography". In: *Nat. Commun.* 15.1 (2024), p. 1589.
- [130] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. "Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems". In: *Nat. Commun.* 11.1645 (2020).
- [131] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. "Non-line-of-sight imaging using phasor-field virtual wave optics". In: *Nature* 572.7771 (2019), pp. 620–623.
- [132] Zicheng Liu, Ying Shan, and Zhengyou Zhang. "Expressive expression mapping with ratio images". In: *Proc. SIGGRAPH*. Los Angeles: ACM, 2001.
- [133] N. R. Lomb. "Least-squares frequency analysis of unequally spaced data". In: *Astrophys. Space Sci.* 39.2 (1976), pp. 447–462.
- [134] Gregor Luetzenburg, Aart Kroon, and Anders A Bjørk. "Evaluation of the Apple iPhone 12 Pro LiDAR for an application in geosciences". In: *Sci. Rep.* 11.1 (2021), pp. 1–9.
- [135] Lumentum. *High-Power Q-Switched Diode-Pumped UV and Green Laser*. Lumentum Website. 2024. URL: <https://www.lumentum.com/en/products/laser-solid-state-q-switched-355-532-q-series>.
- [136] Y Luo and S Mirabbasi. "Always-on CMOS image sensor pixel design for pixel-wise binary coded exposure". In: *Int. Symposium on Circuits & Systems*. Baltimore: IEEE, 2017.
- [137] Yi Luo, Derek Ho, and Shahriar Mirabbasi. "Exposure-Programmable CMOS Pixel With Selective Charge Storage and Code Memory for Computational Imaging". In: *IEEE Trans. Circuits Syst.* 65.5 (2018), pp. 1555–1566.
- [138] Jonathan J Lynch, Harris P Moyer, James H Schaffner, Yakov Royter, Marko Sokolich, Brian Hughes, Yeong J Yoon, and Joel N Schulman. "Passive millimeter-wave imaging module with preamplified zero-bias detection". In: *IEEE Trans. Microwave Theory Tech.* 56.7 (2008), pp. 1592–1600.
- [139] Sizhuo Ma, Shantanu Gupta, Arin C Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. "Quanta burst photography". In: *ACM Trans. Graph. (SIGGRAPH)* 39.4 (2020), pp. 1–16.
- [140] O. Macchi and B. Picinbono. "Estimation and detection of weak optical signals". In: *IEEE Trans. Inf. Theory* 18.5 (1972), pp. 562–573.
- [141] Robert MacCurdy, Rich Gabrielson, Eric Spaulding, Alejandro Purgue, Kathryn Cortopassi, and Kurt Fristrup. "Automatic Animal Tracking Using Matched Filters and Time Difference of Arrival." In: *J. Commun.* 4.7 (2009), pp. 487–495.
- [142] Misha A. Mahowald and Carver Mead. "The Silicon Retina". In: *Scientific American* 264.5 (1991), pp. 76–83. ISSN: 00368733, 19467087. URL: <http://www.jstor.org/stable/24936904>.
- [143] Guoqiang Mao, Barış Fidan, and Brian DO Anderson. "Wireless sensor network localization techniques". In: *Comput. Networks* 51.10 (2007), pp. 2529–2553.

- [144] Jess Marcum. "A statistical theory of target detection by pulsed radar". In: *IRE Trans. Inf. Theory* 6.2 (1960), pp. 59–267.
- [145] Julien NP Martel, Lorenz K Mueller, Stephen J Carey, Piotr Dudek, and Gordon Wetzstein. "Neural sensors: Learning pixel exposures for hdr imaging and video compressive sensing with programmable sensors". In: *IEEE transactions on pattern analysis and machine intelligence* 42.7 (2020), pp. 1642–1653.
- [146] Luiz Eduardo Mendes Matheus, Alex Borges Vieira, Luiz FM Vieira, Marcos AM Vieira, and Omprakash Gnawali. "Visible light communication: concepts, applications and challenges". In: *IEEE Commun. Surv. Tutorials* 21.4 (2019), pp. 3204–3237.
- [147] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. "MC3D: Motion Contrast 3D Scanning". In: *Proc. ICCP*. Houston: IEEE, 2015.
- [148] J. E. Mazo and J. Salz. "On optical data communication via direct detection of light pulses". In: *Bell Syst. Tech. J.* 55.3 (1976), pp. 347–369.
- [149] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N Kutulakos. "Optimal structured light a la carte". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 6248–6257.
- [150] K Mitra, O S Cossairt, and A Veeraraghavan. "A Framework for Analysis of Computational Imaging Systems: Role of Signal Prior, Sensor Noise and Multiplexing". In: *IEEE T-PAMI* 36.10 (2014), pp. 1909–1921.
- [151] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications". In: *Optica* 7.4 (2020), pp. 346–354.
- [152] Paul Mos, Scott Lindner, Chao Zhang, Michael A Wayne, Tommaso Milanese, Claudio Bruschini, and Edoardo Charbon. "Piccolo gated: a CMOS 32x32 SPAD camera with all-solid-state nanosecond time gating and PCIe readout for single-photon time-domain DCS and near-infrared optical tomography". In: *Quantum Sensing and Nano Electronics and Photonics XX*. Vol. 12895. SPIE. 2024, pp. 39–45.
- [153] Eadweard Muybridge. *Animals in motion : an electro-photographic investigation of consecutive phases of animal progressive movements*. Chapman & Hall, 1902.
- [154] Hajime Nagahara, Changyin Zhou, Takuya Watanabe, Hiroshi Ishiguro, and Shree K Nayar. "Programmable aperture camera using LCoS". In: *European Conference on Computer Vision*. Springer. 2010, pp. 337–350.
- [155] Nikhil Naik, Achuta Kadambi, Christoph Rhemann, Shahram Izadi, Ramesh Raskar, and Sing Bing Kang. "A light transport model for mitigating multipath interference in time-of-flight sensors". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 73–81.
- [156] Srinivasa G Narasimhan and S.K Nayar. "Enhancing resolution along multiple imaging dimensions using assorted pixels". In: *IEEE T-PAMI* 27.4 (2005), pp. 518–530.
- [157] Srinivasa G. Narasimhan, Sanjeev J. Koppal, and Shuntaro Yamazaki. "Temporal Dithering of Illumination for Fast Active Vision". In: *ECCV 2008, LNCS, vol. 5305*. Ed. by David Forsyth, Philip Torr, and Andrew Zisserman. Heidelberg: Springer, 2008, pp. 830–844.

- [158] Nature. “Sound Ranging”. In: *Nat.* 104 (1919), pp. 278–280.
- [159] Shree Nayar and Tomoo Mitsunaga. “High dynamic range imaging: spatially varying pixel exposures”. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. Vol. 1. 2000, 472–479 vol.1.
- [160] Shree K Nayar and Vlad Branzoi. “Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures Over Space and Time.” In: 2003.
- [161] R A Newcombe, D Fox, and Steven Seitz. “DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time”. In: *Proc. CVPR*. Boston: IEEE, 2015.
- [162] Jerzy Neyman and Egon Sharpe Pearson. “IX. On the problem of the most efficient tests of statistical hypotheses”. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706 (1933), pp. 289–337.
- [163] R Ng, M Levoy, M Brédif, G Duval, Mark Horowitz, and Pat Hanrahan. “Light field photography with a hand-held plenoptic camera”. In: *Technical Report* (2005). URL: <http://www2.eecs.berkeley.edu/mylène/PI>
- [164] Cindy M Nguyen, Julien NP Martel, and Gordon Wetzstein. “Learning spatially varying pixel exposures for motion deblurring”. In: *2022 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2022, pp. 1–11.
- [165] Cristiano Niclass, Alexis Rochas, P-A Besse, and Edoardo Charbon. “Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes”. In: *IEEE J. Solid-State Circuits* 40.9 (2005), pp. 1847–1854.
- [166] K Noland et al. “The application of sampling theory to television frame rate requirements”. In: *BBC Research & Development White Paper* 282 (2014).
- [167] Makoto Nonoyama, Fumihiko Sakaue, and Jun Sato. “Multiplex Image Projection Using Multi-band Projectors”. In: *Workshop on Color and Photometry in Computer Vision*. Sydney: IEEE, 2013.
- [168] Jeremy L O’Brien. “Optical quantum computing”. In: *Science* 318.5856 (2007), pp. 1567–1570.
- [169] Matthew O’Toole. “Optical linear algebra for computational light transport”. PhD thesis. University of Toronto, June 2016.
- [170] Matthew O’Toole, Supreeth Achar, Srinivasa G Narasimhan, and Kiriakos N Kutulakos. “Homogeneous codes for energy-efficient illumination and imaging”. In: *Proc. SIGGRAPH*. Los Angeles: ACM, 2015.
- [171] Matthew O’Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. “Reconstructing transient images from single-photon sensors”. In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 1539–1547.
- [172] Matthew O’Toole, David B Lindell, and Gordon Wetzstein. “Confocal non-line-of-sight imaging based on the light-cone transform”. In: *Nat.* 555.7696 (2018), pp. 338–341.
- [173] Matthew O’Toole, John Mather, and Kiriakos N Kutulakos. “3D Shape and Indirect Appearance by Structured Light Transport”. In: *IEEE T-PAMI* 38.7 (2016), pp. 1298–1312.
- [174] Matthew O’Toole, Ramesh Raskar, and Kiriakos N Kutulakos. “Primal-dual coding to probe light transport.” In: *ACM Trans. Graph.* 31.4 (2012), pp. 39–1.

- [175] Amílcar Oliveira, Teresa Oliveira, Seijas Macías, and Antonio. “Distribution function for the ratio of two normal random variables”. In: *AIP Conference Proceedings*. Vol. 1648. 2015.
- [176] Alan V Oppenheim, John R Buck, and Ronald W Schafer. *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice Hall, 2001.
- [177] Bakhrom G Oripov, Dana S Rampini, Jason Allmaras, Matthew D Shaw, Sae Woo Nam, Boris Korzh, and Adam N McCaughan. “A superconducting nanowire single-photon camera with 400,000 pixels”. In: *Nat.* 622.7984 (2023), pp. 730–734.
- [178] Parth H Pathak, Xiaotao Feng, Pengfei Hu, and Prasant Mohapatra. “Visible light communication, networking, and sensing: A survey, potential and challenges”. In: *IEEE Commun. Surv. Tutorials* 17.4 (2015), pp. 2047–2077.
- [179] Agata M. Pawlikowska, Abderrahim Halimi, Robert A. Lamb, and Gerald S. Buller. “Single-photon three-dimensional imaging at up to 10 kilometers range”. In: *Opt. Express* 25.10 (2017), pp. 11919–11931.
- [180] Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. “Signal processing based pile-up compensation for gated single-photon avalanche diodes”. In: *arXiv preprint arXiv:1806.07437* (2018).
- [181] Pi Imaging Technology. *SPAD 512²*. Accessed: 2025-04-23. 2025. URL: <https://piimaging.com/product-spad512>
- [182] Julien Pilet, Christoph Strecha, and Pascal Fua. “Making Background Subtraction Robust to Sudden Illumination Changes”. In: *ECCV 2008, LNCS, vol. 5305*. Ed. by David Forsyth, Philip Torr, and Andrew Zisserman. Heidelberg: Springer, 2008, pp. 567–580.
- [183] Genady Pilyavsky, Philip Mauskopf, Nathan Smith, Edward Schroeder, Adrian Sinclair, Gerard T van Belle, Natalie Hinkel, and Paul Scowen. “Single-photon intensity interferometry (spiify): utilizing available telescopes”. In: *Monthly Notices of the Royal Astronomical Society* 467.3 (2017), pp. 3048–3055.
- [184] Mu Qiao, Ziyi Meng, Jiawei Ma, and Xin Yuan. “Deep learning for video compressive sensing”. In: *Apl Photonics* 5.3 (2020), p. 030801.
- [185] Yvain Queau, Roberto Mecca, Jean-Denis Durou, and Xavier Descombes. “Photometric stereo with only two images: A theoretical study and numerical resolution”. In: *Image and Vision Computing* 57 (2017), pp. 175–191.
- [186] Ankit Raghuram, Adithya Pediredla, Srinivasa G Narasimhan, Ioannis Gkioulekas, and Ashok Veeraraghavan. “STORM: Super-resolving transients by oversampled measurements”. In: *Proc. IEEE Int. Conf. on Computational Photography (ICCP)*. 2019, pp. 1–11.
- [187] Thinal Raj, Fazida Hanim Hashim, Aqilah Baseri Huddin, Mohd Faisal Ibrahim, and Aini Hussain. “A survey on LiDAR scanning mechanisms”. In: *Electron.* 9.5 (2020), p. 741.
- [188] Sabbir Rangwala. “The iPhone 12-LiDAR At Your Fingertips”. In: *Online (accessed July 2, 2022), Nov* (2020).
- [189] Scott M. Ransom, Stephen S. Eikenberry, and John Middleditch. “Fourier Techniques for Very Long Astrophysical Time-Series Analysis”. In: *Astron. J.* 124.3 (2002), pp. 1788–1809.
- [190] Joshua Rapp, Robin M A Dawson, and Vivek K Goyal. “Dithered depth imaging”. In: *Opt. Express* 28.23 (2020), pp. 35143–35157.

- [191] Joshua Rapp, Robin M A Dawson, and Vivek K Goyal. "Estimation From Quantized Gaussian Measurements: When and How to Use Dither". In: *IEEE Trans. Signal Process.* 67.13 (2019), pp. 3424–3438. eprint: 1811.06856.
- [192] Joshua Rapp, Robin MA Dawson, and Vivek K Goyal. "Improving lidar depth resolution with dither". In: *IEEE Int. Conf. on Image Processing (ICIP)*. IEEE. 2018, pp. 1553–1557.
- [193] Joshua Rapp and Vivek K Goyal. "A few photons among many: Unmixing signal and noise for photon-efficient active imaging". In: *IEEE Transactions on Computational Imaging* 3.3 (2017), pp. 445–459.
- [194] Joshua Rapp, Yanting Ma, Robin M A Dawson, and Vivek K Goyal. "Dead Time Compensation for High-Flux Ranging". In: *IEEE Trans. Signal Process.* 67.13 (2019), pp. 3471–3486.
- [195] Theodore S Rappaport, Jeffrey H Reed, and Brian D Woerner. "Position location using wireless communications on highways of the future". In: *IEEE Commun. Mag.* 34.10 (1996), pp. 33–41.
- [196] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. "Coded exposure photography: motion deblurring using fluttered shutter". In: *ACM TRANS. GRAPH* (2006), pp. 795–804.
- [197] Netanel Ratner and Yoav Y Schechner. "Illumination Multiplexing within Fundamental Limits". In: *Proc. CVPR*. IEEE, 2007.
- [198] Netanel Ratner, Yoav Y Schechner, and Felix Goldberg. "Optimal multiplexed sensing: bounds, conditions and a graph theory link". In: *Opt. Express* 15.25 (2007), pp. 17072–17092.
- [199] Rayleigh. "XXXI. Investigations in optics, with special reference to the spectroscope". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8.49 (1879), pp. 261–274.
- [200] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. "P2C2: Programmable pixel compressive camera for high speed imaging". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2011, pp. 329–336.
- [201] Vision Research. *Phantom TMX 7510*. <https://www.phantomhighspeed.com/products/cameras/tmx/7510>. 2021.
- [202] Simone Riccardo, Enrico Conca, Vincenzo Sesta, Andreas Velten, and Alberto Tosi. "Fast-gated 16×16 SPAD array with 16 on-chip 6 ps time-to-digital converters for non-line-of-sight imaging". In: *IEEE Sens. J.* 22.17 (2022), pp. 16874–16885.
- [203] Ricardo Roriz, Jorge Cabral, and Tiago Gomes. "Automotive LiDAR technology: A survey". In: *IEEE Trans. Intell. Transp. Syst.* 23.7 (2021), pp. 6282–6297.
- [204] S. M. Ross. *Stochastic Processes*. Wiley, 1983.
- [205] Diego Royo, Jorge García, Adolfo Muñoz, and Adrian Jarabo. "Non-line-of-sight transient rendering". In: *Comput. Graph.* 107 (2022), pp. 84–92.
- [206] Claude Rullière. *Femtosecond Laser Pulses*. Springer New York, 2005.
- [207] Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki. "Dense 3D Reconstruction from High Frame-Rate Video Using a Static Grid Pattern". In: *IEEE T-PAMI* 36.9 (2014), pp. 1733–1747.

- [208] J Salvi, J Pages, and J Batlle. "Pattern codification strategies in structured light systems". In: *Pattern Recognition* 37.4 (2004), pp. 827–849.
- [209] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. "A state of the art in structured light patterns for surface profilometry". In: *Pattern Recognition* 43.8 (2010), pp. 2666–2680.
- [210] Samsung. *Galaxy S24 Ultra*. <https://www.samsung.com/ca/smartphones/galaxy-s24-ultra/>. 2024.
- [211] Navid Sarhangnejad, Hyungjoong Lee, Nikola Katic, Matthew O'Toole, Kiriakos N Kutulakos, and Roman Genov. "CMOS Image Sensor Architecture for Primal-Dual Coding". In: *Int. Image Sensor Workshop*. Hiroshima: International Image Sensor Society, 2017.
- [212] Richard N Scarth. *Mirrors by the Sea: An Account of the Hythe Sound Mirror System Based on Contemporary Letters and Reports*. Hythe Civic Society, 1995.
- [213] Louis L Scharf and Cédric Demeure. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Prentice Hall, 1991.
- [214] Daniel Scharstein and Richard Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms". In: *Int. J. Computer Vision* 47.1 (2002), pp. 7–42.
- [215] Yoav Y Schechner, Shree K Nayar, and Peter N Belhumeur. "Multiplexing for optimal lighting". In: *IEEE T-PAMI* 29.8 (2007), pp. 1339–1354.
- [216] Johannes L Schonberger and Jan-Michael Frahm. "Structure-from-motion revisited". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 4104–4113.
- [217] Rudolf Schwarte, Zhanping Xu, Horst-Guenther Heinol, Joachim Olk, Ruediger Klein, Bernd Buxbaum, Helmut Fischer, and Juergen Schulte. "New electro-optical mixing and correlating sensor: Facilities and applications of the photonic mixer device (PMD)". In: *Proc. SPIE*. Vol. 3100. 1997, pp. 245–253.
- [218] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. "Motion adaptive deblurring with single-photon cameras". In: *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*. 2021, pp. 1944–1953.
- [219] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik P. A. Lensch. "Dual Photography". In: *ACM Trans. Graph. (SIGGRAPH)* 24.3 (2005), pp. 745–755.
- [220] Min-Woong Seo, Yuya Shirakawa, Yuriko Masuda, Yoshimasa Kawata, Keiichiro Kagawa, Keita Yasutomi, and Shoji Kawahito. "4.3 A programmable sub-nanosecond time-gated 4-tap lock-in pixel CMOS image sensor for real-time fluorescence lifetime imaging microscopy". In: *Proc. ISSCC*. San Francisco: IEEE, 2017.
- [221] Michael P Sheehan, Julián Tachella, and Mike E Davies. "A sketching framework for reduced data transfer in photon counting lidar". In: *IEEE Trans. Comput. Imaging* 7 (2021), pp. 989–1004.
- [222] Saad Mehmood Sheikh, Hafiz M Asif, Kaamran Raahemifar, and Fadi Al-Turjman. "Time difference of arrival based indoor positioning system using visible light communication". In: *IEEE Access* 9 (2021), pp. 52113–52124.

- [223] Mark Sheinin, Dorian Chan, Matthew O'Toole, and Srinivasa G Narasimhan. "Dual-shutter optical vibration sensing". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 16324–16333.
- [224] Mark Sheinin, Yoav Y Schechner, and Kiriakos N Kutulakos. "Computational imaging on the electric grid". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 6437–6446.
- [225] Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H Shapiro. "Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors". In: *IEEE Transactions on Computational Imaging* 1.2 (2015), pp. 112–125.
- [226] Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K Goyal, Franco NC Wong, and Jeffrey H Shapiro. "Photon-efficient imaging with a single-photon camera". In: *Nat. Commun.* 7.12046 (2016).
- [227] Shikhar Shrestha, Felix Heide, Wolfgang Heidrich, and Gordon Wetzstein. "Computational imaging with multi-camera time-of-flight systems". In: *Proc. SIGGRAPH*. Anaheim: ACM, 2016.
- [228] Anthony E Siegman. *Lasers*. Mill Valley, Calif: University Science Books, 1986.
- [229] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (2016), pp. 484–489.
- [230] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In: *Science* 362.6419 (2018), pp. 1140–1144.
- [231] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. "Implicit neural representations with periodic activation functions". In: *Advances in neural information processing systems* 33 (2020), pp. 7462–7473.
- [232] Donald L Snyder and Michael I Miller. *Random point processes in time and space*. Springer Science & Business Media, 2012.
- [233] Toshiki Sonoda, Hajime Nagahara, Kenta Endo, Yukinobu Sugiyama, and Rin-ichiro Taniguchi. "High-speed imaging using CMOS image sensor with quasi pixel-wise exposure". In: *Proc. ICCP*. Evanston: IEEE, 2016.
- [234] Richard C Sprinthall. "Basic statistical analysis". In: *(No Title)* (2011).
- [235] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo. "A Range Image Sensor Based on 10- μm Lock-In Pixels in 0.18- μm CMOS Imaging Technology". In: *IEEE Journal of Solid-State Circuits* 46.1 (2011), pp. 248–258.
- [236] Volker Strassen. "Gaussian elimination is not optimal". In: *Numerische mathematik* 13.4 (1969), pp. 354–356.
- [237] Santosh Subedi and Jae-Young Pyun. "A survey of smartphone-based indoor positioning system using RF-based wireless technologies". In: *Sensors* 20.24 (2020), p. 7230.

- [238] Qilin Sun, Xiong Dun, Yifan Peng, and Wolfgang Heidrich. "Depth and transient imaging with compressive SPAD array cameras". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 273–282.
- [239] Stefano Tavani, Andrea Billi, Amerigo Corradetti, Marco Mercuri, Alessandro Bosman, Marco Cuffaro, Thomas Seers, and Eugenio Carminati. "Smartphone assisted fieldwork: Towards the digital transition of geoscience fieldwork using LiDAR-equipped iPhones". In: *Earth Sci. Rev.* 227 (2022), p. 103969.
- [240] Yohann Tanderö and Stanley Osher. "On a mathematical theory of coded exposure". In: *Research in the Mathematical Sciences* 3 (2016), pp. 1–39.
- [241] Michael F Tompsett. *Charge transfer imaging devices*. US Patent 4,085,456. Apr. 18, 1978.
- [242] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. "Bundle adjustment: a modern synthesis". In: *Vision Algorithms: Theory and Practice: Proc. Int. Workshop on Vision Algorithms*. Springer. 2000, pp. 298–372.
- [243] Dobrosław Tsonev, Stefan Videv, and Harald Haas. "Light fidelity (Li-Fi): towards all-optical networking". In: *Broadband Access Communication Technologies VIII*. Vol. 9007. International Society for Optics and Photonics. SPIE, 2014, p. 900702.
- [244] William Sansome Tucker and E. T. Paris. "A selective hot-wire microphone". In: *Philos. Trans. R. Soc. London, Ser. A* 221 (1921), pp. 389–430.
- [245] Harry Urkowitz. "Energy detection of unknown deterministic signals". In: *Proc. IEEE* 55.4 (1967), pp. 523–531.
- [246] Sam Van der Jeught and Joris J J Dirckx. "Real-time structured light profilometry: a review". In: *Optics and Lasers in Engineering* 87 (2016), pp. 18–31.
- [247] Jacob T. VanderPlas. "Understanding the Lomb—Scargle Periodogram". In: *Astrophys. J. Suppl. Ser.* 236.1 (2018), pp. 1–28.
- [248] Edwin Vargas, Julien NP Martel, Gordon Wetzstein, and Henry Arguello. "Time-Multiplexed Coded Aperture Imaging: Learned Coded Aperture and Pixel Exposures for Compressive Imaging Systems". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 2692–2702.
- [249] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).
- [250] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Mounsi Bawendi, Diego Gutierrez, and Ramesh Raskar. "Femto-photography: Capturing and visualizing the propagation of light". In: *ACM Trans. Graph. (SIGGRAPH)* 32.4 (2013), pp. 1–8.
- [251] D. Vere-Jones. "On the Estimation of Frequency in Point-Process Data". In: *Journal of Applied Probability* 19 (1982), pp. 383–394. URL: <http://www.jstor.org/stable/3213577> (visited on 10/25/2022).

- [252] Federica Villa, Rudi Lussana, Danilo Bronzi, Simone Tisa, Alberto Tosi, Franco Zappa, Alberto Dalla Mora, Davide Contini, Daniel Durini, Sasha Weyers, et al. "CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight". In: *IEEE J. Sel. Top. Quantum Electron.* 20.6 (2014), pp. 364–373.
- [253] Federica Villa, Fabio Severini, Francesca Madonini, and Franco Zappa. "SPADs and SiPMs arrays for long-range high-speed light detection and ranging (LiDAR)". In: *Sensors* 21.11 (2021), p. 3839.
- [254] Jean Ville. "Etude critique de la notion de collectif". In: *Bull. Am. Math. Soc.* 45.11 (1939), p. 824.
- [255] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. "Phase-based video motion processing". In: *ACM Trans. Graph. (TOG)* 32.4 (2013), pp. 1–10.
- [256] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. "Riesz pyramids for fast phase-based video magnification". In: *Proc. IEEE Int. Conf. on Computational Photography (ICCP)*. IEEE. 2014, pp. 1–10.
- [257] Gordon Wan, Mark Horowitz, and Marc Levoy. *Applications of Multi-Bucket Sensors to Computational Photography*. Tech. rep. Stanford Computer Graphics Lab, 2012.
- [258] Gordon Wan, Xiangli Li, Gennadiy Agranov, Marc Levoy, and Mark Horowitz. "CMOS Image Sensors With Multi-Bucket Pixels for Computational Photography". In: *IEEE J. Solid-State Circuits* 47.4 (2012), pp. 1031–1042.
- [259] Liang Wang, Ruigang Yang, and James Davis. "BRDF invariant stereo using light transport constancy". In: *IEEE T-PAMI* 29.9 (2007), pp. 1616–1626.
- [260] Peiyuan Wang, Michael A Steindorfer, Franz Koidl, Georg Kirchner, and Erich Leitgeb. "Megahertz repetition rate satellite laser ranging demonstration at Graz observatory". In: *Opt. Lett.* 46.5 (2021), pp. 937–940.
- [261] Peng Wang, Jinyang Liang, and Lihong V Wang. "Single-shot ultrafast imaging attaining 70 trillion frames per second". In: *Nat. Commun.* 11.1 (2020), p. 2091.
- [262] Mian Wei, Sotiris Nousias, Rahul Gulve, David B Lindell, and Kiriakos N Kutulakos. "Passive Ultra-Wideband Single-Photon Imaging". In: *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV)*. 2023, pp. 8135–8146.
- [263] Gordon Wetzstein, Ivo Ihrke, and Wolfgang Heidrich. "On Plenoptic Multiplexing and Reconstruction". In: *Int. J. Computer Vision* 101.2 (2013), pp. 384–400.
- [264] Bennett S Wilburn and Moshe Ben-Ezra. "Time Interleaved Exposures And Multiplexed Illumination". 2015.
- [265] Arnold J. Wilkins and C Clark. "Modulation of light from fluorescent lamps". In: *Light. Res. Technol.* 22.2 (1990), pp. 103–109.
- [266] Ross E Williams. "Creating an acoustic synthetic aperture in the ocean". In: *J. Acoust. Soc. Am.* 60.1 (1976), pp. 60–73.
- [267] Robert J Woodham. "Photometric Method For Determining Surface Orientation From Multiple Images". In: *Opt. Eng.* 19.1 (1980).

- [268] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. "Eulerian video magnification for revealing subtle changes in the world". In: *ACM Trans. Graph.* 31.4 (July 2012).
- [269] Andreas Wuttig. "Optimal transformations for optical multiplex measurements in the presence of photon noise". In: *Applied Optics* 44.14 (2005), pp. 2710–2719.
- [270] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. "A Theory of Fermat Paths for Non-Line-Of-Sight Shape Reconstruction". In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 6793–6802.
- [271] Guochang Xu and Yan Xu. *GPS*. Springer, 2007.
- [272] Wei Xu, Ee Chien Chang, Leoug Keong Kwoh, Hock Lim, Wang Cheng, and Alice Heng. "Phase-unwrapping of SAR interferogram with multi-frequency or multi-baseline". In: *Proceedings of IGARSS'94-1994 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 2. IEEE. 1994, pp. 730–732.
- [273] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. "Generalized Assorted Pixel Camera: Postcapture Control of Resolution, Dynamic Range, and Spectrum". In: *IEEE-TIP* 19.9 (2010), pp. 2241–2253.
- [274] Takuya Yoda, Hajime Nagahara, Rin-ichiro Taniguchi, Keiichiro Kagawa, Keita Yasutomi, and Shoji Kawahito. "The Dynamic Photometric Stereo Method Using a Multi-Tap CMOS Image Sensor". In: *Sensors* 18.3 (2018).
- [275] Michitaka Yoshida, Toshiki Sonoda, Hajime Nagahara, Kenta Endo, Yukinobu Sugiyama, and Rin-ichiro Taniguchi. "High-speed imaging using cmos image sensor with quasi pixel-wise exposure". In: *IEEE Transactions on Computational Imaging* 6 (2019), pp. 463–476.
- [276] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. "Image deblurring with blurred/noisy image pairs". In: *ACM Trans. Graph. (SIGGRAPH)*. Vol. 26. 3. 2007, pp. 1–10.
- [277] Larry Yujiri, Hiroshi H Agravante, Mike Biedenbender, G Samuel Dow, Martin R Flannery, Steven W Fornaca, Bruce I Hauss, Ronald L Johnson, Roger T Kuroda, Karen Jordan, et al. "Passive millimeter-wave camera". In: *Passive Millimeter-Wave Imaging Technology*. Vol. 3064. SPIE. 1997, pp. 15–22.
- [278] Larry Yujiri, Merit Shoucri, and Philip Moffa. "Passive millimeter wave imaging". In: *IEEE Trans. Microwave Theory Tech.* 4.3 (2003), pp. 39–50.
- [279] G. Udny Yule. "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers". In: *Philos. Trans. R. Soc. Lond., A* 226 (1927), pp. 267–298.
- [280] Franco Zappa, Simone Tisa, Alberto Tosi, and Sergio Cova. "Principles and features of single-photon avalanche diode arrays". In: *Sens. Actuators A: Phys.* 140.1 (2007), pp. 103–112.
- [281] Jie Zhang, Ralph Etienne-Cummings, Sang Chin, Tao Xiong, and Trac Tran. "Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure". In: *Opt. Express* 24.8 (2016), pp. 9013–9024.

- [282] Zhongyuan Zhang, Tao Li, Chris Ding, and Xiangsun Zhang. "Binary matrix factorization with applications". In: *Seventh IEEE international conference on data mining (ICDM 2007)*. IEEE. 2007, pp. 391–400.
- [283] Yajing Zheng, Lingxiao Zheng, Zhaoifei Yu, Boxin Shi, Yonghong Tian, and Tiejun Huang. "High-speed image reconstruction through short-term plasticity for spiking cameras". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 6358–6367.
- [284] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. "Retina-like visual image reconstruction via spiking neural model". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 1438–1446.
- [285] Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitan, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini, Daniele Faccio, and Edoardo Charbon. "Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation". In: *Sci. Rep.* 10.20986 (2020).
- [286] David Zimmerman. "Tucker's Acoustical Mirrors: Aircraft Detection Before Radar". In: *War Soc.* 15.1 (1997), pp. 73–99.
- [287] H. Jay Zwally and Per Gloersen. "Passive microwave images of the polar regions and research applications". In: *Polar Rec.* 18.116 (1977), pp. 431–450.