

Experimental Design under Low Signal Conditions

by

Shao Ran Huang

B. S., Chinese University of Hong Kong, 2014

Sc. M., Brown University, 2016

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Physics at Brown University

Providence, Rhode Island

May 2020

© Copyright 2020 by Shao Ran Huang

This dissertation by Shao Ran Huang is accepted in its present form by
the Department of Physics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____

Rashid Zia, Advisor

Recommended to the Graduate Council

Date _____

James M. Valles, Ph.D., Reader

Date _____

Alexander Zaslavsky, Ph.D., Reader

Approved by the Graduate Council

Date _____

Andrew G. Campbell, Ph.D.
Dean of the Graduate School

Vita

Shao Ran Huang (Matthew) was born in 1992 in Hong Kong. He went to the Chinese University of Hong Kong as an undergraduate in 2010 and received a Bachelor of Science in Physics. During his undergraduate career, he worked with Professor Buford Price and Dr Ryan Bay to improve and engineer the IceCube laser dust loggers to investigate the Antarctic dust profile during the last glacial period, allowing reconstruction of paleoclimate records in great detail. He was subsequently accepted into the doctoral program at the Brown University Department of Physics in 2014. In 2016 he received his M.S. in physics along with a Honorable Mention for Excellence as a Graduate Teaching Assistant. That summer, he joined the Zia lab where his research focused on designing optimal experiments under low signal conditions, and its application to spectroscopy, imaging and education research.

Publications

1. **M S.R. Huang**, Y. Wang, S. Cuff, R. Zia, Lifetime Measurement and Dispersion Engineering of High Contrast Gratings (In preparation)
2. **M S.R. Huang**, Juan David Lizarazo Ferro, R. Zia, Application of hardware binning in improving imaging an spectroscopy with CCD (To be submitted)
3. **M S.R. Huang**, Y. Wang, S M. Brown, B. Sandstede, R. Zia, Learning from our mistakes - rethinking graduate admissions (To be submitted)
4. Li, W., **Huang, M S. R.**, Yadavalli, K.S., Lizarazo, D., Zhou, Y., Alexander Zaslavsky, Padture, N., Zia, R.. Direct Characterization of Carrier Diffusion in Halide-Perovskite Thin Films using Transient Photoluminescence Imaging. ACS Photonics (2019)
5. S. Cuff, F. Dubois, **M S.R. Huang**, D. Li, X. Letartre, R. Zia, P. Viktorovitch, and H.S. Nguyen. Tailoring the Local Density of Optical States and directionality of light emission by symmetry-breaking. IEEE Journal of Selected Topics in Quantum Electronics 25.3 (2019): 1-7.

Acknowledgements

These five years of graduate studies at Brown have been both very long and very short. It felt so long when it was late at night in the cleanroom and I was alone trying a variation of a fabrication recipe for the fifth time; it felt so short during the moments of inspiration and when I was with people I love and friends I treasure. I could have never gone so far and achieved so much without the support of my friends, my family, my advisor, and my significant other.

First of all, I would like to thank my advisor Professor Rashid Zia. He is one of the most intelligent and inspiring people I have met in my life. Most importantly, he taught and showed me the way a researcher should think and solve problems, which is much more valuable than knowledge that is easily accessible. Although he is sometimes very demanding, every decision and action of his is always made and done with the well-being of others in mind. His teachings made me who I am today and I am really grateful to have him as my advisor.

I am also really grateful and fortunate to have my lab mates and friends with me during my graduate studies. Wenhao has been my great friend since the beginning and endured the toughest time in the lab together with me. I really enjoyed my time (especially biking and deep chats) with David, and I will be always grateful for his immense help on refining the thesis. Yang has always been very thought-provoking and supportive. Zekun has been my greatest ping pong buddy and a very good friend. I am also grateful to have Ken, John, Martin, Peter, DongQing, Jonathan and many more, as my friends at Brown. They made my life at Brown much more colorful.

I would also like to thank Dr. Sébastien Cuff, Mr Michael Jibitsky, Professor Domenico Pacifici, Professor Alexander Zaslavsky and Professor James Valles for the insightful discussions and for being of great help along the way.

Most importantly, I must also thank Cheryl, my dad, my mum and my sister for their love and support. It is their sacrifice and unlimited support that allowed me to achieve what I wanted to achieve. I will always be grateful to them.

This work is dedicated to my family.

Contents

List of Tables	xi
List of Figures	xii
1 Introduction	1
1.1 Background	1
1.2 Motivation	2
1.3 Outline	6
2 Engineering of LDOS and Directionality of Light Emission	8
2.1 Background	8
2.2 Theoretical Background	9
2.3 Numerical Simulations	13
2.3.1 Coupling of Modes	13
2.3.2 Effects on the Local Density of Optical States	15
2.3.3 Designing Dispersion	17
2.4 Experimental Realization	19
2.4.1 Fabrication	19
2.4.2 Energy Momentum Spectroscopy	19
2.4.3 Comparison with Simulation and Discussion	22

2.5	Investigation of HCG with Other Dispersions	27
2.5.1	Dirac Cone	27
2.5.2	Bound State in the Continuum	31
2.5.3	Flat Band with Bound State in the Continuum	32
2.6	Lifetime measurement of emitters in HCG	35
2.7	Conclusion	43
2.8	Acknowledgement	43
3	Hardware Binning Applied to Image Denoising and Spectroscopy	45
3.1	Introduction	45
3.2	Common Photon Detectors	46
3.2.1	Charge-Coupled Device	46
3.2.2	Complementary Metal-Oxide Semiconductor Imager	46
3.2.3	CCD vs CMOS	47
3.3	Noises and Configurations in CCD	47
3.3.1	CCD Subtypes	52
3.4	Characterizing a CCD camera	54
3.4.1	Background Count and Read Noise	54
3.4.2	Dark Current	55
3.4.3	Photon Transfer Curve	55
3.5	Hardware binning and its application	61
3.5.1	Introduction	61
3.5.2	Prerequisite: A noise model for analysis	62
3.6	Pixel Intensity Reconstruction	63
3.6.1	Derivation of the Likelihood Function	64
3.6.2	Mathematical Details of the EM Algorithm Applied to Denoising	65

3.6.3	Application of EM algorithm in pixel intensity reconstruction	67
3.7	Spectrum Localization	76
3.7.1	Weighted Least-Squares Regression of a Gaussian Spectrum	76
3.7.2	Numerical Fitting of Mixture of two Gaussian Peaks	78
3.8	Conclusion	80
3.9	Acknowledgement	80
4	Learning from Our Mistakes - Expanding Reviews of Graduate Admission Deci-	
	sions	81
4.1	Introduction	81
4.2	Background	82
4.3	Method	83
4.3.1	Data Source	83
4.3.2	GradART	84
4.3.3	Additional data acquisition tools	86
4.4	Sample application	88
4.4.1	Comparison between honorable mention and awardees	90
4.4.2	Matching students to the Carnegie Classification	93
4.5	Conclusion	103
4.6	Acknowledgement	104
5	Conclusions	105
	Bibliography	107

List of Tables

2.1	Table summarizing the lifetime fits results	38
5.1	Table showing summary statistics of the performances of students from the wScholar results ran with the GRFP data set	89
5.2	Table showing summary statistics of the performances of students from different classes of baccalaureate institution from the wScholar results ran with the GRFP data set	94
5.3	Table showing summary statistics of the performances of students from different classes of graduate institution from the wScholar results ran with the GRFP data set	99

List of Figures

1.1	Effect of oxygen flow rate on stiochiometry of VO_2 fabricated by reactive sputtering of vanadium	4
1.2	Hysteresis in deposition rate of reactive sputtering of vanadium due to target poisoning	5
2.1	Notation for the phenomenological analysis of HCG and two examples of them. Figure adapted from Nguyen et al. [1].	10
2.2	Effect of the double-period design on dispersion. Figure reprinted from Nguyen et al. [1].	12
2.3	Energy momentum spectrum of two high contrast gratings. The arrows above the grating drawings symbolize the incoming y-polarized plane wave. One with vertical mirror symmetry and another with symmetry breaking due to the partial etch. One can see the symmetry breaking mixes the two previously orthogonal modes. The intensity profile of the E_y at four different points are also shown on the right.	14
2.4	Band diagrams and 1D LDOS for a) vertically symmetric HCG and b) symmetry-broken HCG.	16
2.5	Photoluminescence of the silicon emitters - note that the huge drop in intensity at long wavelength is due to the sharp decrease in quantum efficiency of the InGaAs camera.	18

2.6	a) Schematic of the investigated structure. The nominal values in fabrication are as follows: $a = 335$ nm, $FF = 0.375$, $\alpha = 0.1$. b) Scanning electron microscope image of the fabricated sample.	18
2.7	Setup for conducting energy momentum spectroscopy. Obj: Objective, DM: Dichroic mirror, SP: Short pass filter, LP: Long pass filter, BL: Bertrand lens, TL: Tube lens, FM: Flip mirror	20
2.8	Left: Energy momentum spectrum of the unpatterned sample. Middle: Energy momentum spectrum of the patterned grating sample. Right: Energy momentum spectrum of the patterned sample normalized by the unpatterned sample in log scale. .	21
2.9	Reflectivity of the patterned sample at 532 nm calculated from RCWA.	22
2.10	Reflectivity of the unpatterned sample at 532 nm calculated from RCWA.	23
2.11	a) Experimental and simulated energy momentum spectrum of the device. b) Field (E_y) distribution in the grating for the four different bands at $k_x=0$	24
2.12	Center: experimental and simulated energy momentum spectra. Left and right: experimental and calculated 1D LDOS.	25
2.13	a) Energy momentum spectrum plotted in 3D as a function of intensity, momentum, and energy. b) Normalized PL intensity as a function of k_x summed over each whole band.	26
2.14	Structure of the devices that exhibit Dirac dispersions and bound state in the continuum dispersion	28
2.15	Measured energy momentum spectrum of the devices designed to exhibit Dirac cone dispersion, with period = 260 nm, FF from 0.38 to 0.42. The red and yellow lines at 1264 nm and 1270 nm respectively are plotted for visual aid.	29
2.16	Simulated reflectivity spectrum of the devices designed to exhibit Dirac cone dispersion, with period = 260 nm, FF from 0.38 to 0.42.	30

2.17	Simulated reflectivity spectrum of the devices at $FF = 0.50$, period from 580 nm to 760 nm, designed to exhibit bound state in the continuum dispersion.	31
2.18	Measured energy momentum spectra of the devices at $FF = 0.50$, period from 580 nm to 760 nm, designed to exhibit bound state in the continuum dispersion.	32
2.19	Simulated reflectivity spectrum of the devices at $FF = 0.42$, P varying from 310 nm to 370 nm, designed to exhibit flat BIC dispersion.	33
2.20	Measured energy momentum spectrum of the devices at $FF = 0.42$, P varying from 310 nm to 370 nm, designed to exhibit flat BIC dispersion.	34
2.21	Simulated reflectivity spectrum of the devices at $P = 340$ nm, FF varying from 0.38 to 0.50, designed to exhibit flat BIC dispersion.	34
2.22	Measured energy momentum spectrum of the devices at $P = 340$ nm, FF varying from 0.38 to 0.50, designed to exhibit flat BIC dispersion.	35
2.23	Setup for conducting lifetime measurement. Obj: Objective, DM: Dichroic mirror, SP: Short pass filter, OC: Optical chopper, LP: Long pass filter, TL: Tube lens, FM: Flip mirror, FC: Fiber coupler	37
2.24	Control device PL decay fitted to mono-exponential	38
2.25	Control device PL decay fitted to bi-exponential	39
2.26	BIC P 620 device PL decay fitted to mono-exponential	39
2.27	BIC P 620 device PL decay fitted to bi-exponential	40
2.28	BIC P 640 device PL decay fitted to mono-exponential	40
2.29	BIC P 640 device PL decay fitted to bi-exponential	41
2.30	Dirac 1300 P 260 device PL decay fitted to mono-exponential	41
2.31	Dirac 1300 P 260 device PL decay fitted to bi-exponential	42
3.1	Dependence of dark current on temperature. Adapted from [2].	49

3.2	A standard specification sheet of a scientific CCD from Princeton Instruments. Adapted from [3].	51
3.3	Baseline of PIXIS at 100 kHz	55
3.4	Baseline of PIXIS at 2 MHz	56
3.5	Baseline SD of PIXIS at 100 kHz	56
3.6	Baseline SD of PIXIS at 2 MHz	57
3.7	Temporal variation in baseline for 2 MHz. 8 pixels are selected and 30000 frames are taken. The average of 1000 frames is plotted as a function of time.	57
3.8	Histogram of baseline for four different pixels on PIXIS, overlaid with Gaussian fits .	58
3.9	Histogram of dark current for PIXIS	58
3.10	Example Photon Transfer Curve, by Janesick et al. [4].	60
3.11	Simple illustration showing a Gaussian ground truth spectrum, plotted with single pixel and two pixel binned measurements. Both are corrupted first with Poisson noise and then readout noise of $SD = 50$, implemented with numpy [5].	67
3.12	Verification that the EM algorithm converges to the maximum likelihood estimation in the case of two pixel: Left figures show the explicit calculation of the likelihood function for the measurement. Right figures show the predicted ground truth value for each pixel as a function of iteration steps of the EM algorithm. Black line indicates the maximum likelihood estimation. a) and b) show the results for two random experiments.	69
3.13	Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 2 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 2 pixels.	72

3.14	Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 4 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 4 pixels.	73
3.15	Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 8 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 8 pixels.	73
3.16	Example spectra at the particular read noise and exposure time distribution that combined measurement defeats single pixel measurement.	74
3.17	Cramér-Rao lower bound computed over a range of single-pixel exposure times t , fluxes f_1 and f_2 , and readout noise σ . The values for f_1 are constant along the rows, and the values for f_2 are the fractions of f_1 shown on the top.	75
3.18	Comparison between the Cramér-Rao lower bound computed and the expected variance when all time is spent on single pixel measurement. In this particular example, $f_1 = 10$, $f_2 = 7$, and $\sigma = 2$	76
3.19	Density plot: Uncertainty in peak localization in binned pixel measurement. Contour plot: Ratio of the uncertainty in peak localization in binned pixel measurement to that in single pixel measurement. Number lower than 1 means binned measurement gives lower uncertainty. Red line: Ratio of uncertainty = 1.	77
3.20	RMSE for peak localization for single pixel, double pixel, and quadruple pixel binned measurements. Left: Peak width being a random number from 2 to 16. Right: Peak width being a random number from 1 to 4.	79

4.1	The four main components of GradART: a) NSC: Expands admission data by adding information on non-matriculated students (as discussed in the text, the ratio of applications to matriculated students is approximately twenty; the ratio of students who struggled vs completed successfully is based on [6, 7]. b) wScholar: Allows programmatic search of publication records for all students. c) Awards Tracker: Retrieves fellowship and grant records from NIH and NSF. d) Thesis Finder: Allows quick access to the PhD theses of graduated students for additional analysis.	85
4.2	Showing the actual results from wScholar and Awards Tracker: illustrating the two extreme levels. The top part shows the academic achievements of an anonymous individual, while the bottom part shows aggregate statistics for chemistry students from three anonymous schools. Both individual level and institution level views can be generated, allowing small and large scale decisions to be scrutinized.	87
4.3	Showing the normalized histogram for number of papers for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.13 and 0.09 respectively. The mean values are 3.38 for awardees and 3.23 for honorable mention students. . .	90
4.4	Showing the normalized histogram for number of first authored papers for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.28 and 0.86 respectively.	91
4.5	Showing the normalized histogram for impact factors for first authored paper summed for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.37 and 0.18 respectively.	92
4.6	Showing the normalized histogram for number of citations for first authored papers summed for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.15 and 0.79 respectively.	92
4.7	Showing the total number of students, average number of first authored papers and average number of citations for each major from each category of undergraduate schools.	95

4.8	Showing the normalized histogram for number of papers for students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.48 and 0.63 respectively.	96
4.9	Showing the normalized histogram for number of first authored papers for students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.04 and 0.03 respectively. The respective means for R1 and Baccalaureate Colleges are 1.26 and 1.38 respectively.	96
4.10	Showing the normalized histogram for sum of impact factors of first authored papers by students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.46 and 0.69 respectively.	97
4.11	Showing the normalized histogram for sum of citations of first authored papers by students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.44 and 0.33 respectively.	97
4.12	Showing the total number of students, average number of first authored papers and average number of citations for each major from each category of graduate schools. .	100
4.13	Showing the normalized histogram for number of papers for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are $2e-6$ and $9e-7$ respectively. The mean values are 3.3 for R1 and 2.3 for R2 students respectively.	101
4.14	Showing the normalized histogram for number of first authored papers for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are 0.01 and 0.03 respectively. The mean values are 1.3 for R1 and 1.1 for R2 students respectively.	101
4.15	Showing the normalized histogram for impact factors for first authored paper summed for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are $1e-4$ and $6e-7$ respectively. The mean values are 6.7 for R1 and 4.2 for R2 students respectively.	102

4.16 Showing the normalized histogram for number of citations for first authored papers summed for awardees and honorable mentions. Mann-Whitney U test and t-test p- values are 2e-3 and 1e-7 respectively. The mean values are 39 for R1 and 21 for R2 students respectively.	102
--	-----

Chapter 1

Introduction

1.1 Background

Design of experiments (DoE) investigates the optimal experimental methods to perform a given measurement. It aims to extract the greatest amount of information with the least uncertainty possible about the target quantity with constraints on the available resources like measurement time. As experimentation is a central part of science, DoE is an important consideration for many researchers and has been widely discussed in literature [8–12].

In some situations, DoE can be hugely beneficial. A popular example is the weighing problem [11], which considers the optimal manner of measuring n weights using a chemical balance. The balance shows the difference in weights between objects in the left and the right pans and each measurement has an independent normally distributed error $\epsilon \sim N(0, \sigma^2)$. The simplest method would be to weight each object successively. Denoting y_i and x_i to be the observations and the actual weights respectively, the observations y_i would be $\sim N(x_i, \sigma^2)$. Another less apparent way would be to weigh such that information from all weights can be obtained at every measurement. One way to do so is to configure the measurements such that the configuration vectors, where an entry of 1 (-1) means the weight is on the left (right) pan, form the rows of a Hadamard matrix. A Hadamard matrix is

a square matrix composed of 1 and -1 with mutually orthogonal rows and each pair of rows have matching (mismatched) entities in half (the other half) of their columns; interested readers may refer to [13]. The estimations y_i would then $\sim N(x_i, \sigma^2/n)$ with the same amount of observations. To achieve the same precision with the first method, n observations for each object will be required, leading to n^2 observations in total.

Another illustration of the importance of experimental design is exhibited in machine learning. Experiments which test the performance of a machine learning model trained with a specific set of hyperparameters can be very costly in terms of computational power and time. An example was the famous convolutional neural network AlexNet taking five to six days to train with two state-of-the-art GPUs in 2012 [14]. Techniques like Bayesian Optimization [15] were therefore developed to inform the best next experimental location to probe in the hyperparameter phase space, and have been applied in different real world experiments [16, 17].

These examples illustrate how DoE can improve the measurement accuracy with the same resources. This is of particular importance under low-signal conditions, for instance, in the case of spectroscopy of weak emitters and particle localization experiments.

1.2 Motivation

Our interest in investigating experimental design grew after challenging experiments with vanadium dioxide (VO_2). Our aim was to find a fabrication recipe to grow stoichiometric VO_2 on a conductive thin film for electrically switching the phase-change material. VO_2 is an interesting phase-change material as it exhibits the metal-insulator transition when excited thermally, optically or electrically. The structural change is accompanied by a considerable change in refractive index, rendering the material a great candidate for optical switching. VO_2 can be fabricated with many methods. Sputtering is one of the most common techniques used to grow stoichiometric VO_2 . Within sputtering,

there are also options of reactive sputtering of vanadium metal in oxygen and RF sputtering of oxide targets (commonly with V_2O_5 , V_2O_3). For both avenues, the growth condition for VO_2 is very critical and requires precise proportions of oxygen and argon gases in the sputtering chamber, as vanadium exhibits many other different oxidation states including V_2O_5 , V_2O_3 , V_3O_7 and V_6O_{13} . The presence of other phases weakens the contrast brought about by the metal-insulator transition in the thin film and has to be avoided.

It turned out that two important considerations about the experiment were not taken into account, making the goal of fabricating stoichiometric VO_2 very difficult to achieve.

1. The control of oxygen flow was not precise enough for the deposition of pure VO_2

After an extensive search over the fabrication condition, it was found that the optimal window for fabricating stoichiometric VO_2 was between 4.7 and 5 sccm (standard cubic centimeters) of O_2 flow. The quality of VO_2 was quantified by measuring the order of magnitude of resistance change of the heat induced metal insulator transition. As shown in Fig 1.1, both oxygen rich and oxygen deficient samples were obtained in that particular range. It was subsequently found that the accuracy of the oxygen mass flow control in the system was only 0.2 sccm. That explained the variations in quality of VO_2 observed even though the designated oxygen flow rate was kept the same. It severely restricted the precise control of the process conditions, rendering fabrication of pure VO_2 very difficult.

2. Choice of sputtering target and substrate was not optimal

Reactive sputtering of vanadium metal was chosen, because it was the common choice in older literature. It was not realized that relatively newer papers use mostly V_2O_5 as the sputtering target because the sputtering condition can be much more stable. The former method suffers from the target poisoning effect, which is due to the oxidation of the surface of the vanadium metal sputtering target. As the surface is oxidized, sputtering yield and deposition rate decrease, which in turn further reduce the use of oxygen gas in the chamber, leading to even greater oxidation on the surface. This makes the plasma content and the degree of sputtering target oxidation dynamic and unstable. Fig 1.2 shows the target poisoning effect. As the oxygen gas flow increases, the

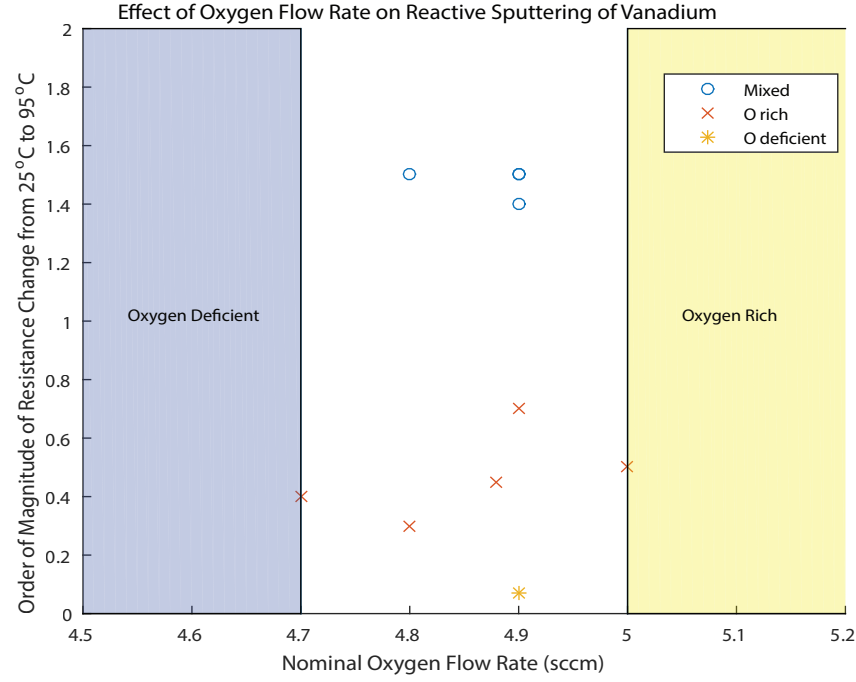


Figure 1.1: Effect of oxygen flow rate on stoichiometry of VO_2 fabricated by reactive sputtering of vanadium

deposition rate initially increases as there is more plasma sputtering the material. But it soon drops significantly as the surface of the target is oxidized. As the oxygen flow rate decreases from the maximum, the deposition rate only slightly increases, as the surface of the sputtering target is still highly oxidized, leading to a low sputtering yield. This leads to the observed hysteresis. As the oxygen flow rate further decreases, the degree of oxidation further decreases and deposition rate increases, using up more oxygen gas in the chamber and eventually inducing a runaway effect which returns the target to the highly metallic state. This shows that the deposition condition is sensitive to changes in the oxygen content in the plasma and is unstable. This makes reactive sputtering of vanadium particularly challenging.

As the ultimate goal was to fabricate stoichiometric VO_2 on a conductive substrate, testing fabrication on borosilicate glass was not effective. It was subsequently found that the ideal fabrication

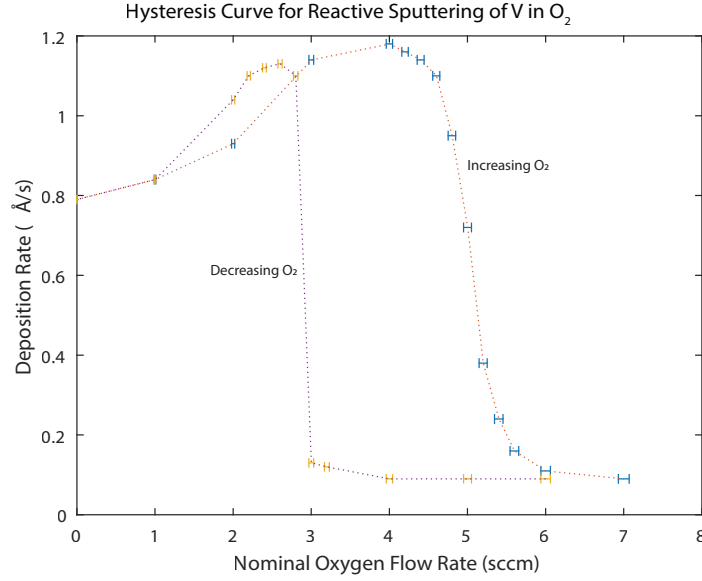


Figure 1.2: Hysteresis in deposition rate of reactive sputtering of vanadium due to target poisoning condition of VO_2 also depends on the substrate. Optimizing on borosilicate glass only gives a very rough estimate of the ideal fabrication condition on other substrates. This showed that a researcher should be knowledgeable of the properties of the material he works with and has a well informed plan before conducting experiments.

In view of these, several changes have been made to the experiments. (1) A more advanced mass flow controller (Alicat MCE) with an accuracy of better than 0.02 sccm was installed, providing an order of magnitude improvement in control accuracy. (2) V_2O_5 sputtering target was used instead of the vanadium metal sputtering target, making the oxygen content in the chamber much more stable. (3) Changing the deposition substrates to be silicon, which is much closer to the final substrate choice doped poly-silicon. After these changes, stoichiometric VO_2 was optimized and successfully fabricated on silicon. The results of the subsequent experiments with VO_2 do not align with the central theme of the thesis and are therefore not discussed. However, it proved to be an experiment that could have benefited from a better DoE by intentionally taking into account the constraints of resources as well as the uncertainties brought about by experimental apparatus and

measurements.

1.3 Outline

In this thesis, DoE under low-signal conditions is discussed in three different contexts. In particular, the thesis is structured as follows:

In the second chapter, we present a demonstration of enhancement in photoluminescence from weak silicon defects emitters by high contrast gratings. We then show the characterization of an array of photonic devices with a variety of geometrical parameters which enhances light emission into particular wavelengths and directions by modifying the local density of states. We discuss the measurement of the engineered dispersions, including Dirac cone, bound state in the continuum, and flat bands and PL lifetime measurements. It was originally thought that due to the inherent noisy nature of InGaAs detectors and the weak emission from silicon, intelligent experimental designs would be required. However, it turned out that long measurement time was sufficient to achieve the desired signal-to-noise ratio.

The low-signal experimental conditions encountered promoted us to investigate the optimal design of experiments. Therefore, in the third chapter, we scrutinize the way spectroscopy and imaging are done. After reviewing the essentials of CCDs, we illustrate the characterization of a CCD camera to obtain its noise statistics. We then apply the EM algorithm to denoising and use the Cramér-Rao lower bound to investigate the optimal measurement scheme. After that, we analyze spectra where there is significant spatial correlation between neighbouring pixels, and illustrate that improvements in localization accuracy can be achieved with hardware binning.

In the fourth chapter, we apply the concept of experimental design to tackle a broader scientific challenge. Graduate admission is a process complicated with inherently large variances and limited data. With a usually long completion time of graduate studies, feedback information acquisition is very slow. In this work, we therefore strive to improve the data acquisition process for a more

thorough analysis. We report a method and present software tools for obtaining and analyzing academic outcome data for graduate students. We then apply the software to the NSF Graduate Research Fellowships data set as an illustration, hoping that it can encourage and facilitate a more data-driven approach to examining graduate admission as well as other related decisions in hiring and recruitment.

Chapter 2

Engineering of LDOS and Directionality of Light Emission

2.1 Background

Ever since Purcell's work in the 1940s [18], engineering of optical environment has been an active topic of research. It was shown that spontaneous emission is not only affected by the underlying intrinsic electronic levels of the system, but is also significantly influenced by the density of optical states. The Purcell effect asserts that the enhancement in spontaneous emission rates scales with the quality factor Q and the inverse of the mode volume V . A few select examples of the many ways by which to modify the local electromagnetic environment include Drexhage's work on modifying the fluorescence decay time of europium molecules by placing them in close proximity to mirrors [19]; Björk's work on modifying the spontaneous emission rate and radiation pattern of quantum wells with one-dimensional dielectric microcavities [20]; and Englund's work on controlling the spontaneous emission rate of single quantum dots with a single-defect cavity in a 2D photonic crystal [21]. Such work opened the exciting window of allowing researchers to manipulate local

optical environment to control emission of light.

Traditionally, efforts to modify spontaneous emission have been relatively more successful in two kinds of resonators: photonic crystals and metal nano-particles [22]. For metal nano-particles there is the advantage of being able to modify the cavity volume below the diffraction limit by confinement through plasmon oscillations, and with cavities formed between two metal nano-particles there have been reports of confinement on the order of $10^{-6}(\lambda_0/n)^3$ (where λ_0 is the free space wavelength of the transition and n the refractive index of the medium inside the cavity). However, they have the disadvantage that coupling to electron oscillations in metals unavoidably brings about significant energy loss [23].

In photonic crystals there is the advantage that their dielectric structures can incur in very low loss, while there is the drawback that the minimum mode volume is generally restricted by the diffraction limit. Efforts to manipulate the spontaneous emission rates with photonic crystals are still ongoing with novel concepts like heterostructure photonic nanocavities achieving quality factors as high as 2.5×10^6 [24].

The advancement of sub-wavelength lithography has made possible the fabrication of photonic crystals made with high-contrast gratings that have length scales small enough to eliminate the presence of higher order diffraction modes. High contrast gratings are usually composed of high-index sub-wavelength gratings in low-index media [25], and enable the control of light emission in manners as will be shown in this thesis.

2.2 Theoretical Background

An intuitive approach for designing high-index contrast sub-wavelength gratings with broken symmetry was recently proposed by our collaborators Nguyen et al. [1]. This allows both the local density of optical states and directionality of light to be manipulated by varying the geometrical parameters of high contrast gratings.

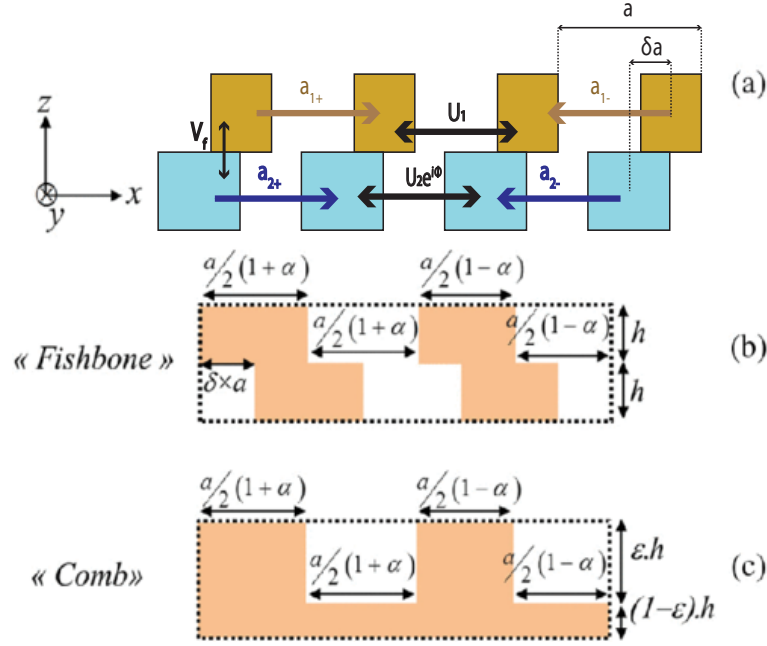


Figure 2.1: Notation for the phenomenological analysis of HCG and two examples of them. Figure adapted from Nguyen et al. [1].

This approach is based on the introduction of slight geometrical perturbations to the structure of regular gratings so as to produce structures with corrugations as shown in Figure 2.1. These structures can be analysed either with numerical methods such as rigorous coupled wave analysis (RCWA) or with phenomenological models. In this section, we quickly review the phenomenological model as proposed by Nguyen et al. [1] which offers a useful construct to understand the relevant interactions that allow the method to work. This guided the design of structures that enhance the emission from silicon defects which are studied later in this chapter.

The model consists of two interacting grating structures of the same periodicity as shown in Fig 2.1a. Both layers have backward- and forward-propagating fundamental zero-order modes and their amplitudes are described by (a_{1+}, a_{2+}) and (a_{1-}, a_{2-}) as sketched in that same figure. Under the coupled mode theory formalism, the interaction of the electromagnetic waves in the two layers can be described by a Hermitian 4x4 Hamiltonian as described in eq. 2.1 and which is represented

in the ordered basis formed by $[a_{1+}, a_{1-}, a_{2+}, a_{2-}]$. In this model U_1 and U_2 are the diffractive coupling rates between the forward and backward waves in the two layers, which describe the effects of reflection due to the refractive index oscillations; the factors β_1 and β_2 are diffractive coupling rates resulting from the evanescent part of the guided modes of the other layer; the phase $e^{\pm i\phi}$ is due to the first-order diffraction phase shift at the X point between the two layers due to the lateral offset; ω_1, ω_2 and v_1, v_2 are the energies and the group velocities of the propagating waves in the two layers respectively; finally, V_f is the evanescent coupling rate between waves of different layers travelling in the same direction.

$$H = \begin{bmatrix} \omega_1 + v_1 k_x & U_1 + \beta_1 U_2 e^{i\phi} & V_f & 0 \\ U_1 + \beta_1 U_2 e^{-i\phi} & \omega_1 - v_1 k_x & 0 & V_f \\ V_f & 0 & \omega_2 + v_2 k_x & U_2 e^{i\phi} + \beta_2 U_1 \\ 0 & V_f & U_2 e^{-i\phi} + \beta_2 U_1 & \omega_2 - v_2 k_x \end{bmatrix} \quad (2.1)$$

By diagonalization of the Hamiltonian, the eigenmodes of the high contrast grating can be obtained. By breaking the vertical mirror symmetry, the even and odd eigenmodes are hybridized, resulting in exotic dispersion characteristics. Note that this hybridization requires spectral overlapping of even and odd modes, which occurs when the diffractive coupling rates U_1 and U_2 exceed the evanescent coupling rate V_f , a condition usually satisfied in high index contrast structures [1].

Note that guided modes cannot be detected in the far field if they are below the light-line. The designs of the dispersion have been all below the light line to avoid radiative losses. To bring these modes above the light-line while minimizing radiative loss, an approach is to use a double-period perturbation, as illustrated in the Fig 2.1 by the parameter α . This technique is commonly known as “Brillouin-zone folding” [26]. This small double-period perturbation can change the period of the structure to be two times larger, making the first Brillouin zone two times smaller. This effectively folds the first Brillouin zone and brings the dispersion previously at the edge of the first Brillouin zone to the Γ point, as illustrated in Fig 2.2.

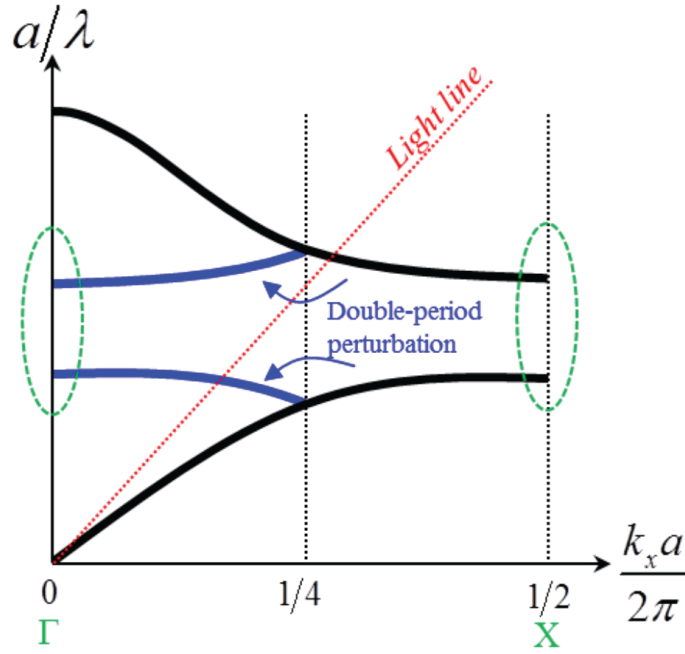


Figure 2.2: Effect of the double-period design on dispersion. Figure reprinted from Nguyen et al. [1].

Of particular interest is the comb structure shown in Fig 2.1c because of its relatively standard fabrication process. In this case one of the layers is a grating, and the other a homogeneous layer of equal index of refraction so that both U_2 and β_1 equal zero. In this structure the geometrical parameters a and α can be modified so as to obtain Dirac cones and flat band dispersions [1].

We applied this concept and experimentally demonstrated this engineering of dispersion on weak silicon emitters, showing how all-dielectric symmetry-broken high contrast gratings can be designed to simultaneously control the directionality of light emission and the local density of optical states, and enhance weak light emission from silicon into specific directions.

2.3 Numerical Simulations

2.3.1 Coupling of Modes

The dispersion engineering enabled by high contrast gratings depends on the coupling of photonic eigenmodes. By varying the various parameters of the gratings, including height, fill factor, and period, the mixing of the photonic modes can be controlled and their dispersion can be tweaked into particularly desired shapes.

The physics behind coupling modes in high contrast gratings can be understood with the following illustration. Consider an amorphous silicon grating as in Fig 2.3a. The photonic eigenmodes can be obtained by rigorous coupled wave analysis. Setting the horizontal direction to be x and the out-of-plane direction to be y , plane waves with a y polarization are sent in at different incident angles. A small absorption in the silicon is configured ($k = 10^{-6}$) and the intensities of plane waves reflected by and transmitted through the structure are measured to calculate the absorption. By repeating this simulation for plane waves with different energy and incident angle, the absorption spectrum as a function of energy and momentum in x direction can be obtained.

Note that by reciprocity theorem, the source and the detector can be inverted without affecting the physical phenomenon. More specifically, the Green dyadic is symmetric in source and detector [27]. Therefore, the absorption spectrum can also yield information about the far-field radiation of emitters in the fabricated structure.

In the absorption spectrum in Fig 2.3a, the two modes are the even and odd waveguided eigenmodes in the gratings. Note that they are orthogonal to each other and they cross without mixing. An interesting phenomenon happens when the vertical symmetry is broken - when the high contrast grating is not fully etched, leaving 50 nm of silicon at the bottom of the grating. As seen in the absorption spectrum in Fig 2.3b, the two originally orthogonal modes now couple to each other, resulting in a gap in the photonic band.

By tuning the geometric parameters of the structure, including etch depth, period of the grating

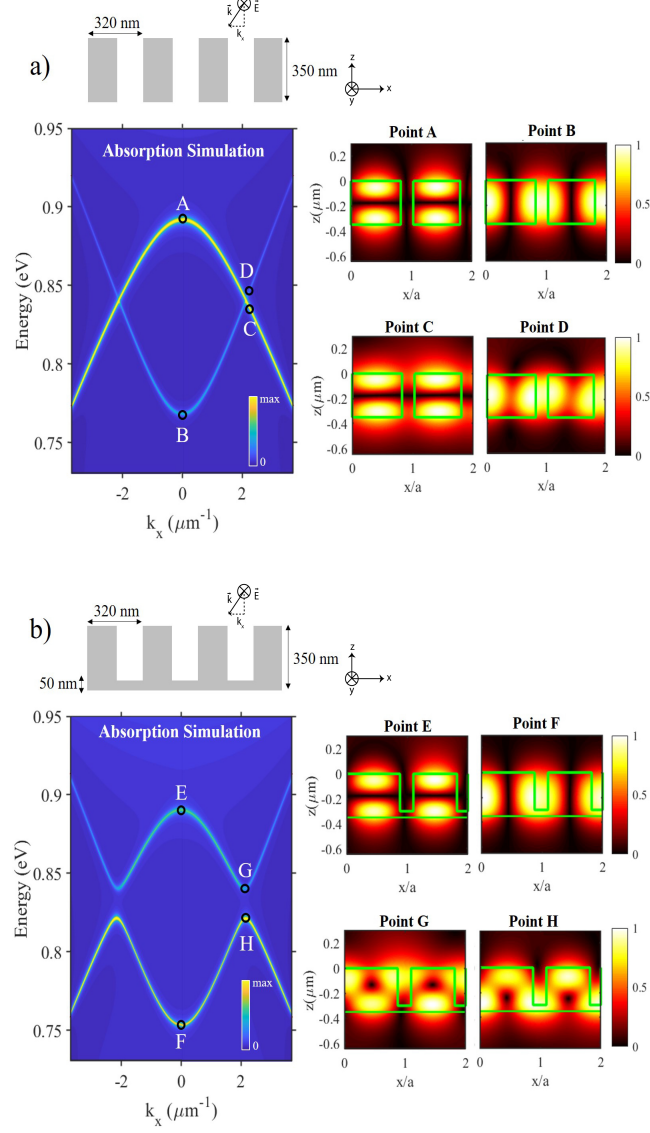


Figure 2.3: Energy momentum spectrum of two high contrast gratings. The arrows above the grating drawings symbolize the incoming y-polarized plane wave. One with vertical mirror symmetry and another with symmetry breaking due to the partial etch. One can see the symmetry breaking mixes the two previously orthogonal modes. The intensity profile of the E_y at four different points are also shown on the right.

and fill factor, the dispersion of the structure can be manipulated to produce varying dispersion relationships. This method therefore offers the possibility to engineer the directionality of light at specific wavelengths.

Note that a double period perturbation, as discussed before, is also used in the subsequently investigated structures so that the unit cell is doubled in size, effectively folding the modes from the edge to the center of the first Brillouin zone. This allows the waveguide modes to be above the light-line and be accessible through free space illumination and detection without introducing significant loss mechanisms.

2.3.2 Effects on the Local Density of Optical States

As the local density of optical states is directly dependent on the group velocity of light, which in turn depends on the curvature of the energy momentum dispersion, the emission rate of nearby emitters can also be controlled and directed into desired modes.

To investigate this effect, our collaborator Dr. Sébastien Cueff calculated the band structure with a commercial software implementation of the plane wave expansion method (Bandsolve, RSoft), and the 1D LDOS in k_x direction computed by summing the number of modes with the dispersion data at ω over a range $d\omega$ for the two structures. The results are shown in Fig 2.4.

At momentum close to the band gap opening due to the coupling of the two modes, the dispersion has a near zero group velocity (as $V_g = \frac{d\omega}{dk}$) which results in a very large local density of states, strongly enhancing emission into this particular momentum and wavelength. This strong enhancement is verified in the measurement of the energy momentum spectrum in the subsequent sections.

This example illustrates how LDOS can be specifically enhanced at particular wavelength and angle by modifying the various parameters of a high contrast grating. By designing structures with flat dispersion at desired wavelength / angle, emission can be strongly enhanced into that particular mode. This can be particularly useful in cases in which emitters are very weak (like single photon

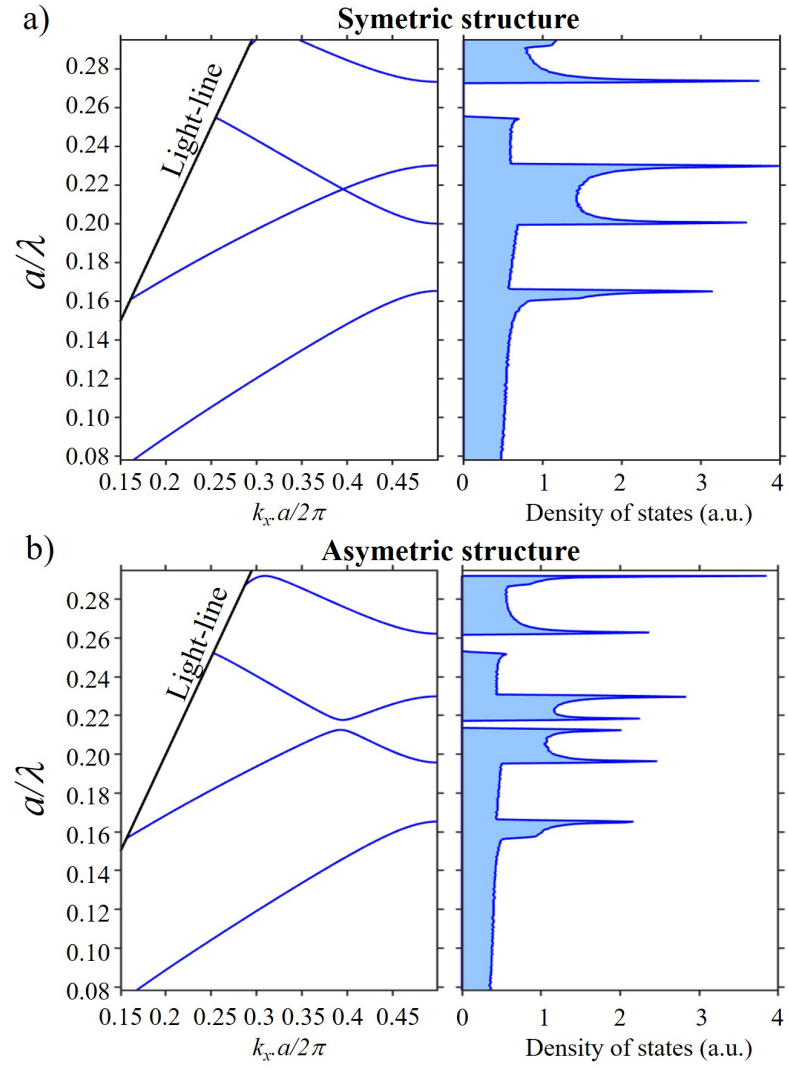


Figure 2.4: Band diagrams and 1D LDOS for a) vertically symmetric HCG and b) symmetry-broken HCG.

emitters: NV centers in diamond) and it is desired to couple light from the emitters into useful optical modes (like coupling into an optical fiber for quantum communication purposes).

2.3.3 Designing Dispersion

Simulations for various different parameters were carried out by our collaborator Dr. Sébastien Cueff to achieve a range of specific dispersions. Rigorous coupled-wave analysis (RCWA) is used to find the modes of the structures. TE polarized light at different angles and wavelengths were used as the incident light and reflectivities were measured. The parameters of the structure including period, fill factor, and thicknesses of the amorphous silicon layer, are varied to achieve specifically desired dispersions. Note that RCWA could not directly obtain information of the waveguide optical modes below the light line as they do not interact with fields from the far-field. Only modes which couple to the far-fields (modes above the light line) can be found. These modes can be folded from the boundary to the center by the double period design approach discussed in the previous text.

Initial efforts have been trying to focus on a simple illustration of enhancement of light emission at the flat regions. Silicon was chosen because it is an indirect bandgap material and is a very poor emitter at room temperature. Previous literature reported that defects in crystalline and amorphous silicon produce weak light emission below the bandgap between $1.3\ \mu\text{m}$ and $1.6\ \mu\text{m}$ [28]. Therefore, a wavelength range in this region was investigated. Fig 2.5 shows the photoluminescence of these emitters pumped by 532 nm laser.

To experimentally demonstrate the coupling of light and the enhancement of emission into specifically desired modes, a high contrast grating structure with different types of dispersion features was chosen for detailed experimentation and analysis. The design is illustrated as in Fig 2.6a. It consists of several layers of thin films: 100 nm of crystalline silicon on $2\ \mu\text{m}$ of SiO_2 on a silicon substrate, followed by 100 nm of Y_2O_3 as an etch stop layer (which in subsequent experiments serves as the host of erbium ions), and finally the 350 nm thick amorphous silicon grating. The period of the structure is 335 nm, fill factor is 0.375, and the perturbation factor is 0.1.

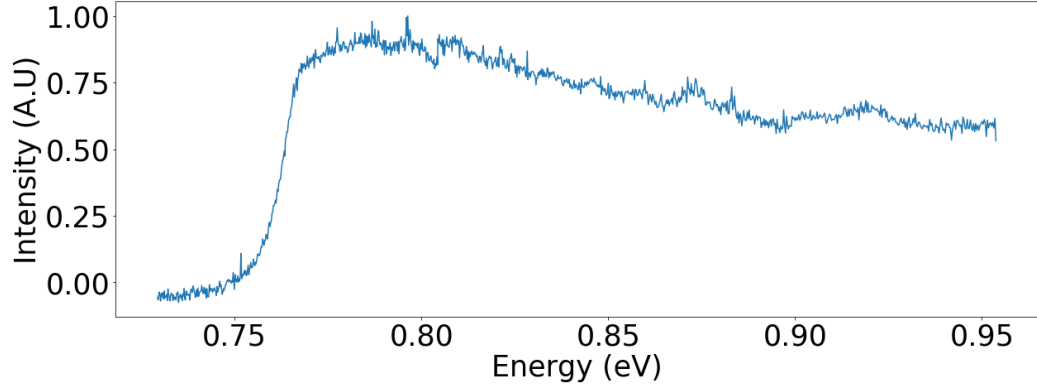


Figure 2.5: Photoluminescence of the silicon emitters - note that the huge drop in intensity at long wavelength is due to the sharp decrease in quantum efficiency of the InGaAs camera.

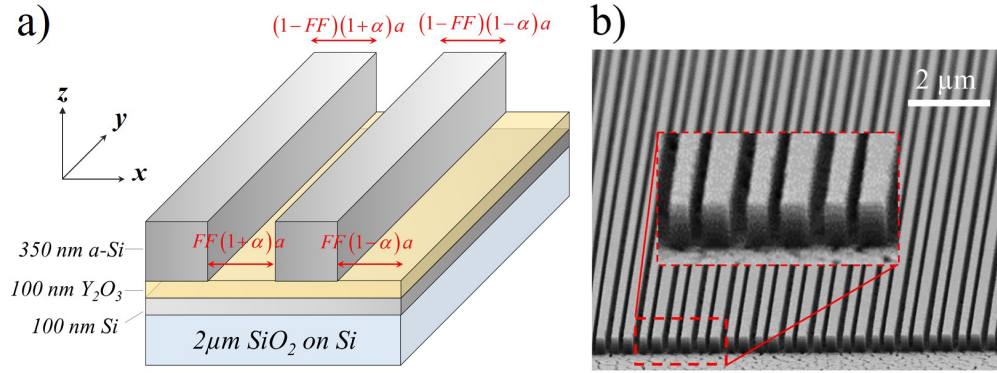


Figure 2.6: a) Schematic of the investigated structure. The nominal values in fabrication are as follows: $a = 335\text{ nm}$, $FF = 0.375$, $\alpha = 0.1$. b) Scanning electron microscope image of the fabricated sample.

2.4 Experimental Realization

2.4.1 Fabrication

The fabrication starts with a commercial silicon on insulator substrate. Using electron beam evaporation, 100 nm of Y_2O_3 is then deposited by electron beam evaporation. Because the deposition was done in vacuum, the deposited thin film was oxygen deficient. The Y_2O_3 layer was then annealed at 900°C under a flux of 0.5 lpm of O_2 for 1 hour in order to achieve the desired stoichiometry and crystallinity. A layer of amorphous silicon of 350 nm thick was then deposited by plasma enhanced chemical vapor deposition (PECVD) using SiH_4 as precursor and helium as the plasma gas. The power of the RF driver was fixed at 25W and the pressure used was 2 Torr, with a substrate heating of 300°C. A 100 nm layer of hydrogen silsesquioxane (HSQ) was then spun and baked at 80°C for 4 minutes on the sample acting as the electron beam resist. For patterning, electron beam lithography was used to expose the resist to create a pattern for etching the amorphous silicon with inductively coupled reactive ion etching (ICP-RIE) using Cl_2 and Ar. The scanning electron microscope image of the fabricated sample is shown in Fig 2.6b.

2.4.2 Energy Momentum Spectroscopy

The experimental setup we built to do energy momentum spectroscopy is illustrated in Fig 2.7. To excite the weak emitters in silicon, a diode pumped, frequency-doubled Nd:YVO₄ laser at 532 nm (Verdi Coherent) was used to optically pump the emitters. An inverted microscope (Nikon Eclipse Ti) was used to measure the photoluminescence (PL) of the emitters. An air objective (Nikon Plan Apo, 60x, 0.95 NA) was used to focus the pump light and collect the emitted light. The resulting NIR emission was then dispersed by an aberration-corrected Schmidt-Czerny-Turner spectrometer (Princeton Instruments, IsoPlane SCT-320), and subsequently measured by a high-resolution 2D InGaAs focal plane array (Princeton Instruments, PIoNIR 640).

To measure the dispersion relation of the structure, energy momentum spectra were measured

using a 250 mm focal length Bertrand lens to image the back focal plane of the objective onto the camera. As discussed in [29], positioning the Bertrand lens between the objective and tube lens simplifies the optical alignment process and significantly improves Fourier imaging quality. A polarizer was placed between the Bertrand lens and the objective to only collect TE-polarized light. The gratings were aligned perpendicular to the slits of the spectrograph, imaging the dispersion along k_x while $k_y=0$. All measurements were done at room temperature.

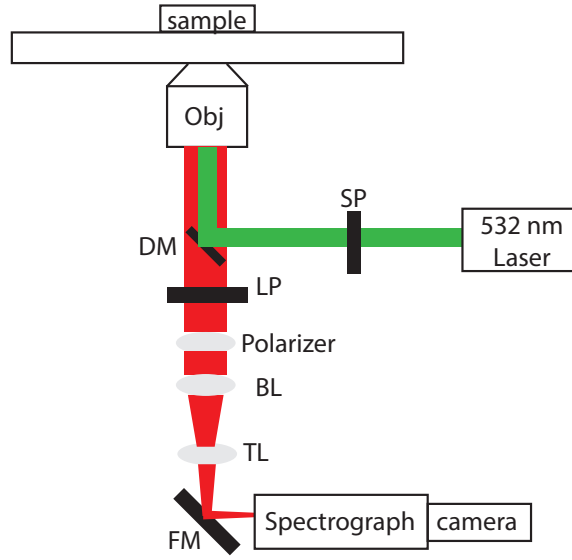


Figure 2.7: Setup for conducting energy momentum spectroscopy. Obj: Objective, DM: Dichroic mirror, SP: Short pass filter, LP: Long pass filter, BL: Bertrand lens, TL: Tube lens, FM: Flip mirror

Two samples with the same deposition layers, one of them patterned and the other one unpatterned, were then compared. Their energy momentum spectra were measured under equal experimental conditions of pump power and exposure time and the results are illustrated in the left and middle graphs in Fig 2.8. It was observed that the PL of the unpatterned sample was very weak with wide fringes that resembled signals from a poor Fabry-Perot cavity formed by the silicon layers, this was probably due to the absorption and roughness of the layers. For the patterned sample, however, the PL signals were much stronger and very pronounced in shape as bands similar to that

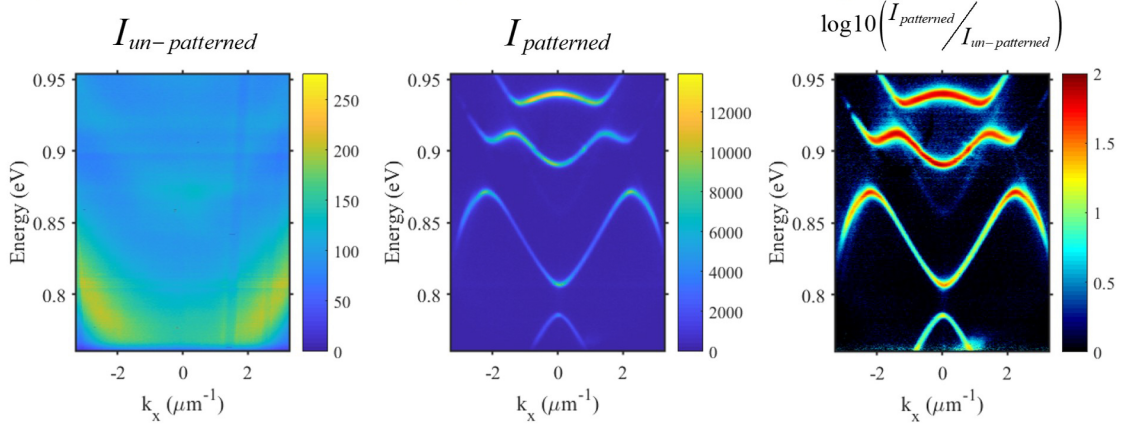


Figure 2.8: Left: Energy momentum spectrum of the unpatterned sample. Middle: Energy momentum spectrum of the patterned grating sample. Right: Energy momentum spectrum of the patterned sample normalized by the unpatterned sample in log scale.

of the absorption spectra obtained from the RCWA simulations. The right figure in Fig 2.8 shows the logarithm of the ratio between the two measurements with a difference that could be up to two orders of magnitude. It was also observed that the intensity along the same band also varied, with the maxima being close to the flat regions of the bands in agreement with the increased LDOS at regions with small group velocity.

Note that the actual enhancement in PL was smaller, as the intensity of the pumping laser reaching the emitters was not strictly the same in both cases. The extinction coefficient of the amorphous silicon layer deposited was 0.38 (as measured by ellipsometry), leading to an attenuation coefficient of $8.98 \mu m^{-1}$. With a thickness of 350 nm, that would correspond to an absorption of 95.7%. However, there was also the complication of the reflection of the pumping laser by the grating. Fig 2.9 and Fig 2.10 show the reflectivities of the patterned and unpatterned sample calculated at 532 nm. The patterned sample had a much higher reflectivity than the unpatterned sample. A full quantitative comparison would require integrating over all angles of the pump laser, taking into the consideration that part of the amorphous silicon is replaced by air in the patterned sample (defined by the fill factor). Another experiment which involves modifying the incident laser

power so as to result in the same amount of optical power at the location of the emitters would allow a better quantitative comparison between the magnitude of the PL intensities and arriving at a more accurate enhancement factor. However, since this project focuses on the engineering of dispersion, this complication was not considered.

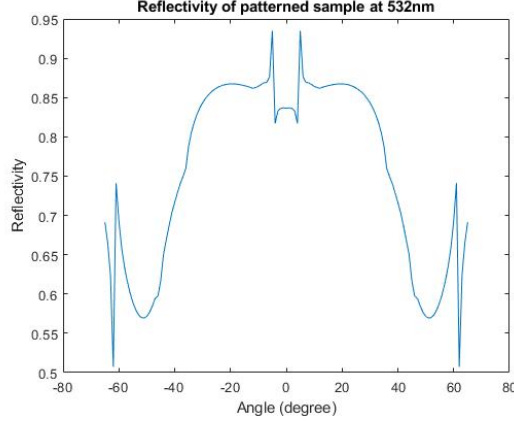


Figure 2.9: Reflectivity of the patterned sample at 532 nm calculated from RCWA.

2.4.3 Comparison with Simulation and Discussion

Energy Momentum Spectrum

Fig 2.11a shows the experimental data (in the red dotted box) together with the simulated absorption spectrum and Fig 2.11b shows the field distribution intensity E_y for the four bands at $k_x = 0$. The emission from the four measured bands are very similar to the simulated spectrum, with no significant PL outside of the bands. There is a good agreement between the experimental and simulated data. Note that, however, this agreement is obtained with slightly modified parameters (period $a = 329.5\text{nm}$ instead of the nominal value 335nm , fill factor $FF = 0.355$ instead of the nominal value 0.375 and perturbation $\alpha = 0.125$ instead of the nominal value 0.1). This is expected to be due to the small fabrication mismatch with respect to the design and inaccuracy in refractive indices measured by ellipsometry.

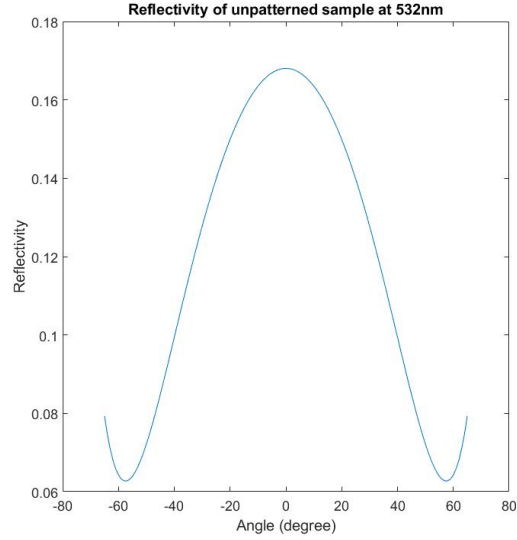


Figure 2.10: Reflectivity of the unpatterned sample at 532 nm calculated from RCWA.

LDOS

To obtain an experimental estimation of the 1D LDOS as a function of wavelength, one can sum up the angular contributions at each specific wavelength. Fig 2.12 compares the experimentally estimated and the simulated 1D LDOS (obtained by summing up all the modes located at ω over a $d\omega$ window). It can be observed that there is a good qualitative agreement between the simulated and experimental results. As the LDOS is strongly enhanced as the group velocity V_g (which corresponds to the slope $\frac{d\omega}{dk}$) of the dispersion approaches zero, emission is strongly enhanced at the extrema of the photonic bands, as observed in the right graph Fig 2.8. This illustrates a method to selectively enhance emission into specific wavelengths. By carefully tuning the parameters of the grating, one can couple emitters to a grating with strong LDOS at a desired wavelength to strongly enhance emission into that particular wavelength / mode.

In both experiment and simulation, the $k_y = 0$ contribution was selected. Experimentally this was achieved by closing the slit, while theoretically only the modes around $k_y = 0$ were summed. Since the periodicity in refractive indices was only in the x direction, the density of states was

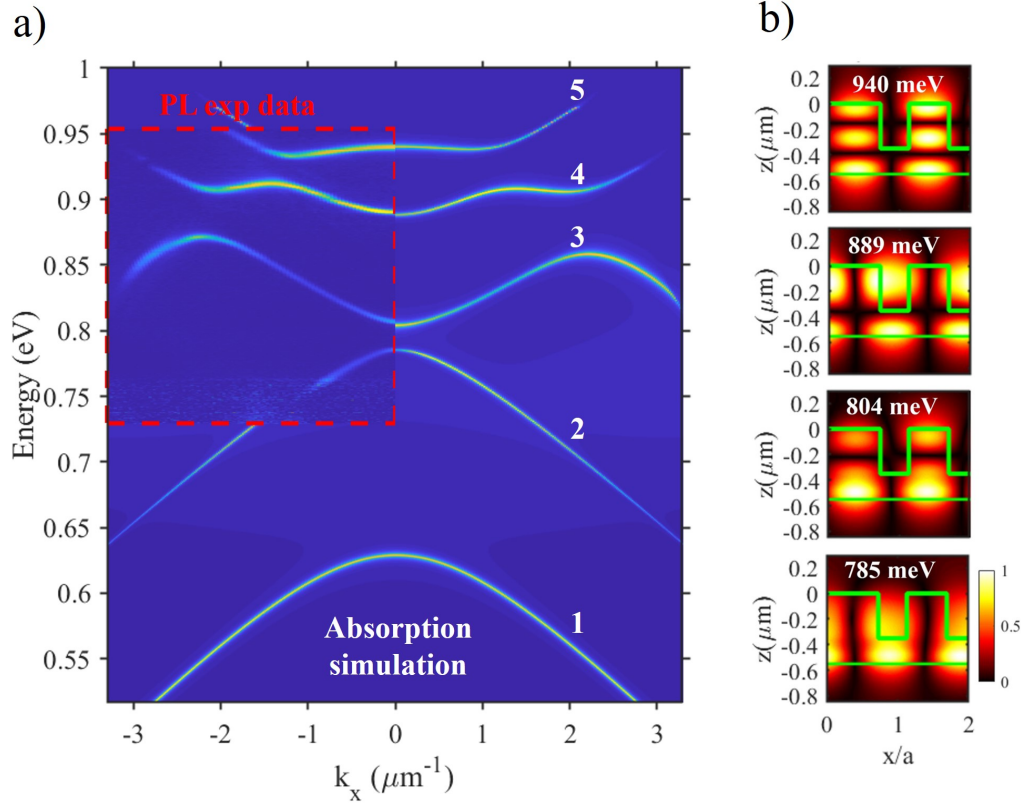


Figure 2.11: a) Experimental and simulated energy momentum spectrum of the device. b) Field (E_y) distribution in the grating for the four different bands at $k_x=0$

only significantly affected in k_x direction. To modify the density of states in both k_x and k_y , the structures could be designed in 2 dimensions. An example would be a 2-dimensional pillar array.

Directionality

Fig 2.13a shows a 3D plot of the energy momentum spectrum. One can clearly differentiate the different angular distribution of emission for each photonic band. As seen in graphs showing the intensity summed for each band in Fig 2.13b, different modes have significantly different directionality because of their differences in dispersion. For example, the flat-band has relatively similar contribution from $k_x = -1\mu\text{m}^{-1}$ to $k_x = 1\mu\text{m}^{-1}$, while the M-band channels most of the energy to

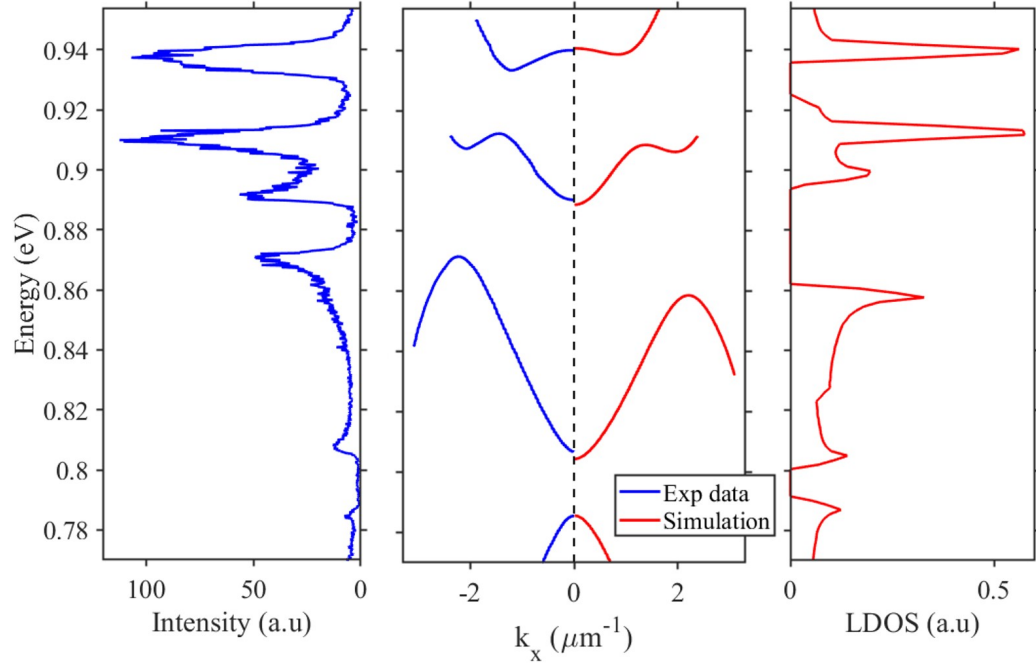


Figure 2.12: Center: experimental and simulated energy momentum spectra. Left and right: experimental and calculated 1D LDOS.

high angles (around $k_x = 2\mu\text{m}^{-1}$ and $k_x = -2\mu\text{m}^{-1}$), and the parabolic band tightly concentrates emission into very small angles (around $k_x = 0$). This again illustrates the freedom allowed by this grating manipulation in channeling energy into specific directions, in addition to the discussed wavelength.

One should however note that this ability to engineer emission of light into particular wavelengths and directions is not without limitation. As the origin of this dispersion engineering is from coupling of the modes in the high contrast gratings, not any kind of arbitrarily shaped dispersion can be achieved. For example, isolated point-like dispersions would be very difficult to achieve, as it requires engineering all other points (k_x) of the dispersion band inaccessible to the far field (creating a point at the Γ point $k_x = 0$ is relatively easier because of the possibility of creating a symmetry protected mode).

In any case, in addition to the freedom in engineering the structures, there is also the freedom in choosing an emitter. By choosing directional emitters at particular wavelengths, there is a very exciting opportunity in achieving emission in a desired wavelength and direction.

As a small change in geometrical parameters / refractive indices can induce a huge change in dispersion, another interesting avenue of research would be to incorporate tunable materials like phase change materials or liquid crystals into this design, achieving light emission with electrically controlled emission wavelength and direction.

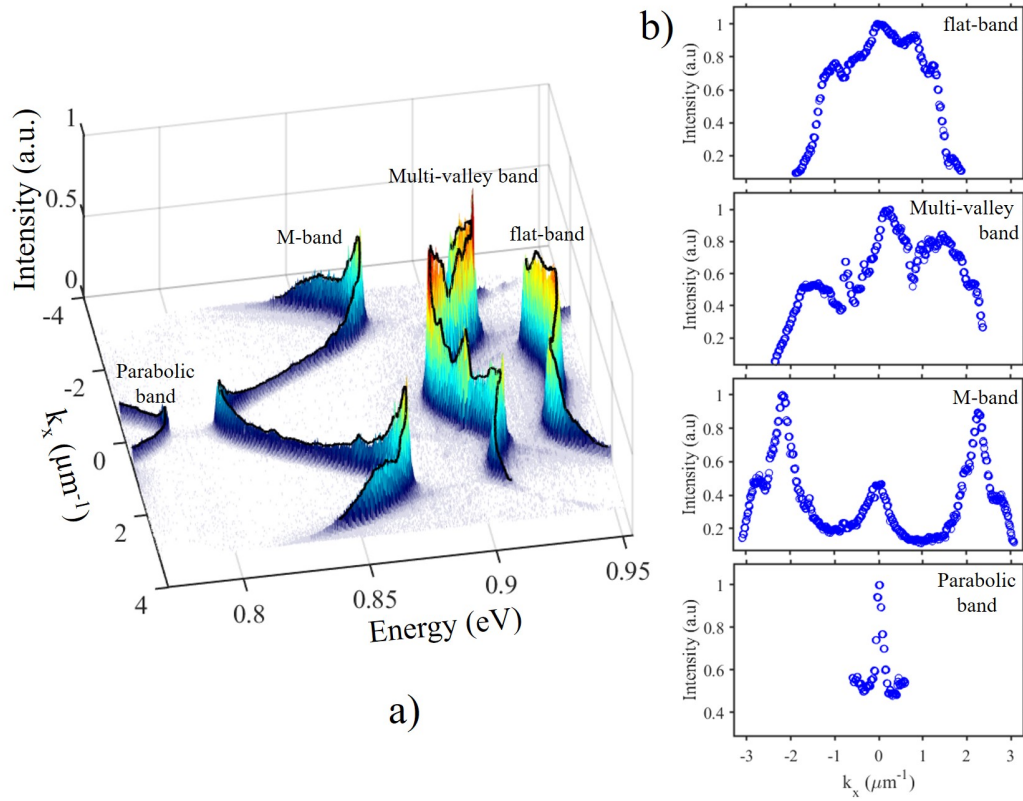


Figure 2.13: a) Energy momentum spectrum plotted in 3D as a function of intensity, momentum, and energy. b) Normalized PL intensity as a function of k_x summed over each whole band.

2.5 Investigation of HCG with Other Dispersions

Recently, several types of exotic photonic dispersions have been extensively studied: bound state in the continuum [30], Dirac cone [31, 32] and flat band [33, 34]. To investigate the dispersion engineering enabled by parameter tuning, simulations were done to obtain the reflectivity spectrum under different geometrical parameters of the structure. We then fabricated and investigated the devices that lead to these interesting dispersions.

Due to the low loss of optical fiber at $1.5\mu\text{m}$, near infrared emission of erbium is of great technological importance. It is therefore interesting to investigate the enhancement of emission from erbium with these exotic dispersions. In the subsequent experiments, the samples have undergone similar fabrication and experimental conditions, except that erbium doped yttrium oxide is used as the etch stop instead of just yttrium oxide.

2.5.1 Dirac Cone

Fig 2.14 shows the structure of the devices that exhibit Dirac cone dispersions. Note that the thickness of the amorphous silicon was modified to be 300 nm instead of 350 nm in our previous device. By simulating the structures with RCWA with the same fill factor (0.38) as the previously discussed device and changing the period, dispersions resembling Dirac cones at around 1.26 μm are observed. Fig 2.16 shows a panel of RCWA result for devices with period = 260 nm and varying fill factor. It can be observed that the two bands are initially slightly separated at $\text{FF} = 0.38$. As FF increases, the two bands slowly merge. The two bands also slightly shift towards shorter wavelengths, with the center being at 1270 nm at $\text{FF}=0.38$ to being at 1264.5 nm at $\text{FF}=0.43$. At around $\text{FF}=0.42$ or $\text{FF}=0.43$, the two bands completely merge to form Dirac cones.

Fig 2.15 shows the measured energy momentum dispersion of the five devices, with red and yellow lines plotted at 1264nm and 1270nm respectively for visual aid. There are two observations that warrant discussion: (1.) The separation of the bands according to the RCWA simulations is at most

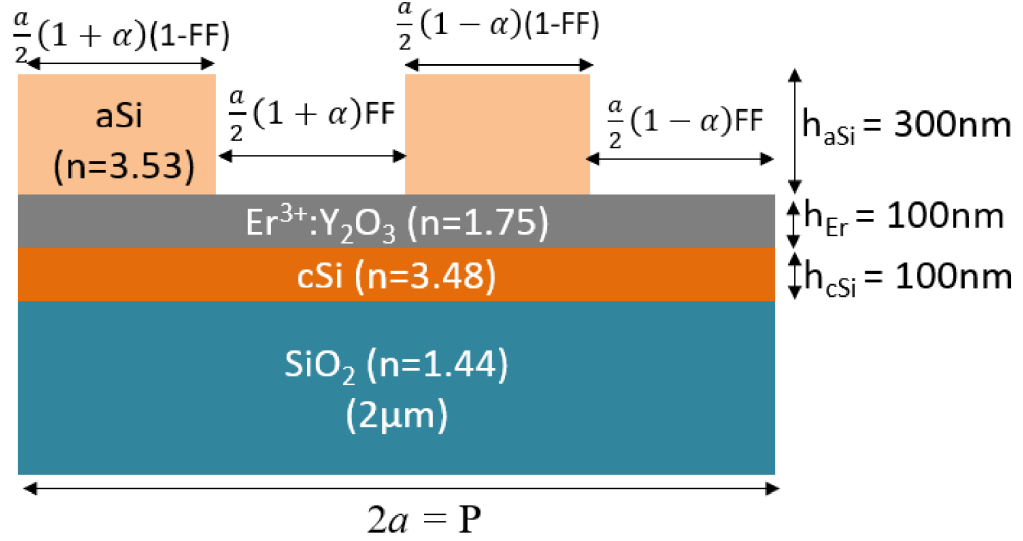


Figure 2.14: Structure of the devices that exhibit Dirac dispersions and bound state in the continuum dispersion

1.5 nm. Since the slit width used was $200\ \mu\text{m}$ and the pixel size is $20\ \mu\text{m}$, the spectral resolution is worse than 3 nm [35], so the gap cannot be distinguished. (2.) The centers of all the Dirac cones are shorter than 1264 nm (above the red line). This can be due to fabrication imperfections with respect to the designed parameters. It was indeed observed that the centers slightly blue-shifts as FF increases, agreeing with the prediction in the simulation.

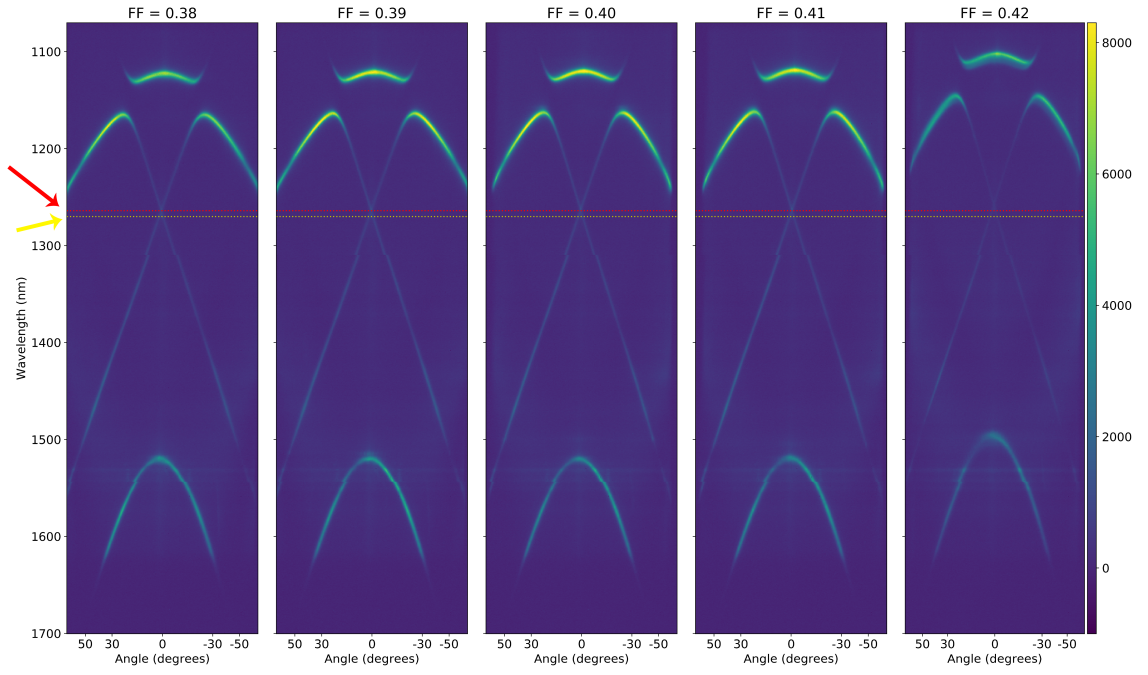


Figure 2.15: Measured energy momentum spectrum of the devices designed to exhibit Dirac cone dispersion, with period = 260 nm, FF from 0.38 to 0.42. The red and yellow lines at 1264 nm and 1270 nm respectively are plotted for visual aid.

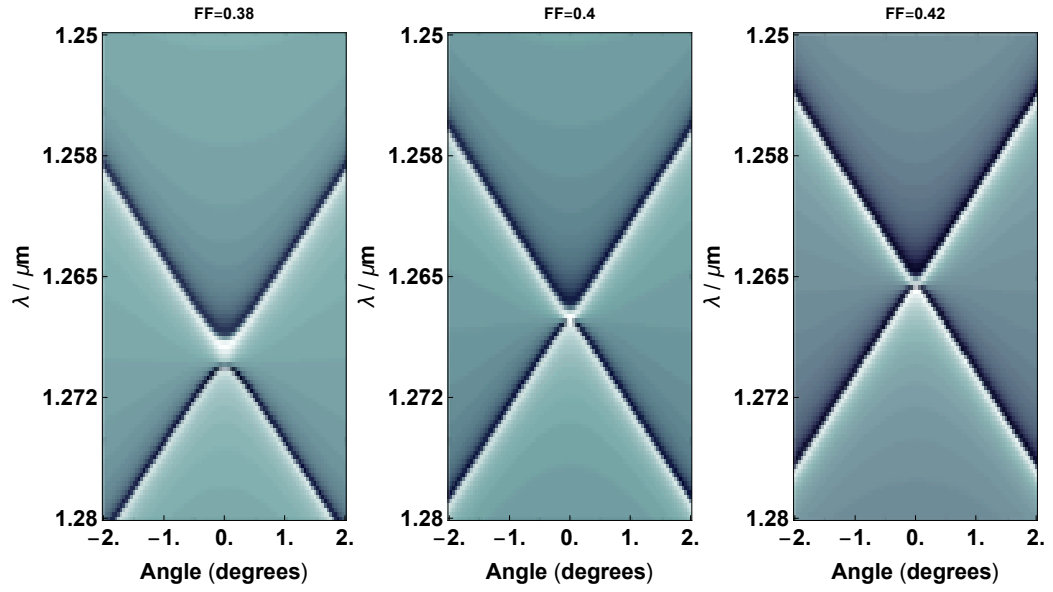


Figure 2.16: Simulated reflectivity spectrum of the devices designed to exhibit Dirac cone dispersion, with period = 260 nm, FF from 0.38 to 0.42.

2.5.2 Bound State in the Continuum

Interestingly, simply by changing the fill factor to 0.50, the dispersion can be converted to that of a bound state in the continuum. Fig 2.17 shows the reflectivity simulation graphs as the period is changed from 580 nm to 760 nm. It can be observed that the bands red-shift as the period is increased. Fig 2.18 shows the measured energy momentum spectra of these devices. It can be observed that the photonic bands have vanishing PL intensity at the tip of the bands, even when they coincide with the erbium emission lines at around 1545 nm, verifying that the designed bound states in the continuum do not couple to the far field.

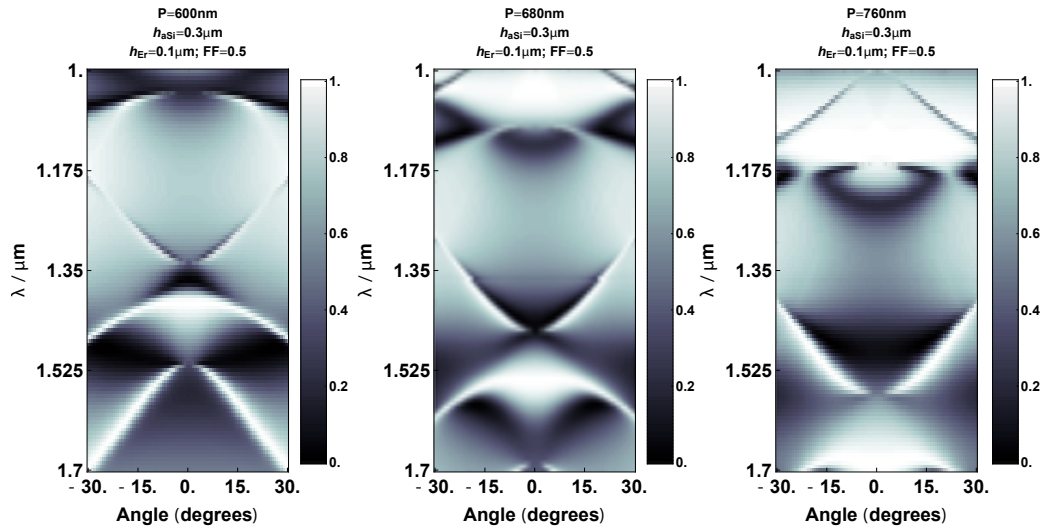


Figure 2.17: Simulated reflectivity spectrum of the devices at $\text{FF} = 0.50$, period from 580 nm to 760 nm, designed to exhibit bound state in the continuum dispersion.

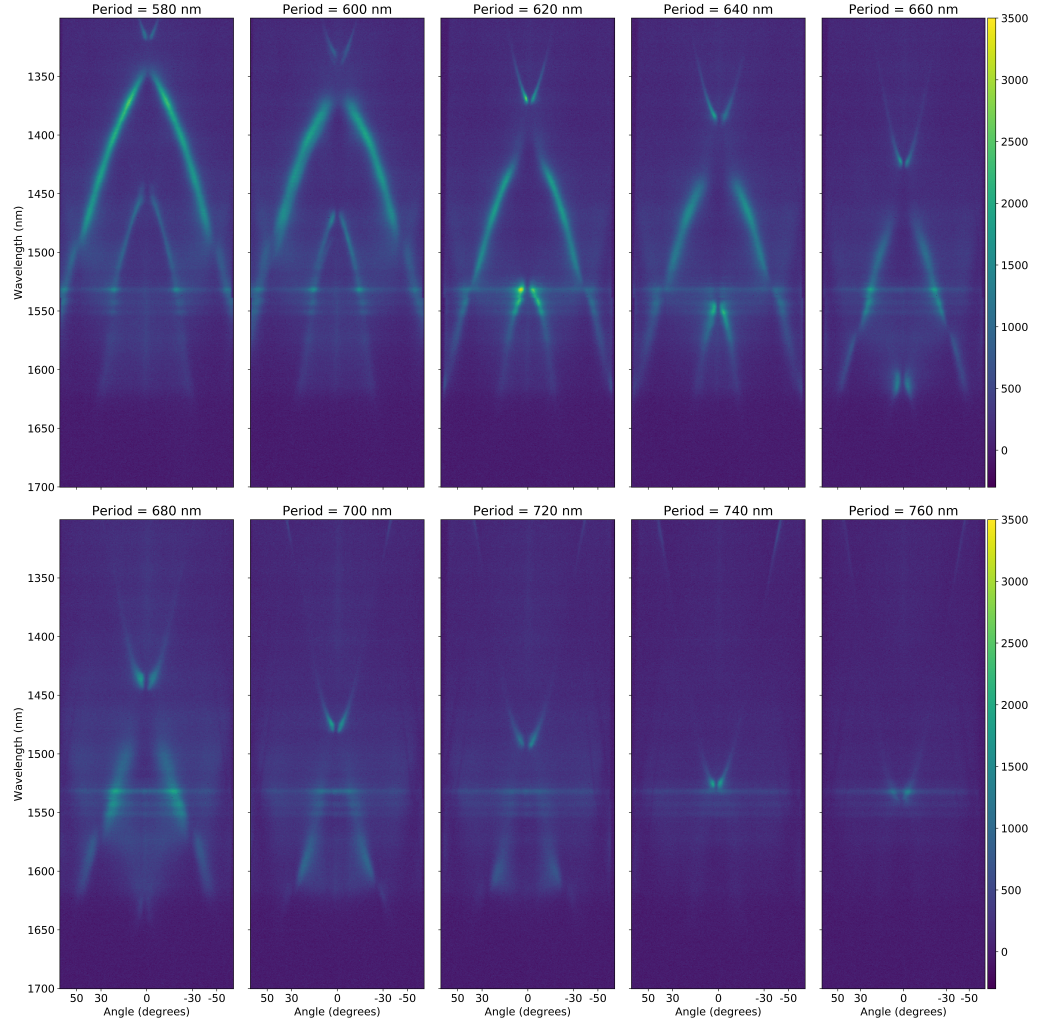


Figure 2.18: Measured energy momentum spectra of the devices at $FF = 0.50$, period from 580 nm to 760 nm, designed to exhibit bound state in the continuum dispersion.

2.5.3 Flat Band with Bound State in the Continuum

Fig 2.19 and Fig 2.20 shows the simulated reflectivity and experimental energy momentum spectra of devices with $FF = 0.42$, $P = 310$ nm to 370 nm. It can be observed that as P increases, the bands red-shift. $P = 340$ nm corresponds to the dispersion with flat band closest to the $1.5 \mu\text{m}$ erbium line. So additional simulation at $P=340\text{nm}$ is done, varying the FF from 0.38 to 0.50. Fig 2.21 and Fig 2.22 show the simulated and experimental energy momentum spectra of the devices respectively.

It can be observed that as FF increases, the band gets flatter. The $FF = 0.50$ energy momentum spectrum also exhibits strongest enhancement due to the flatness of the band and its alignment with an erbium line.

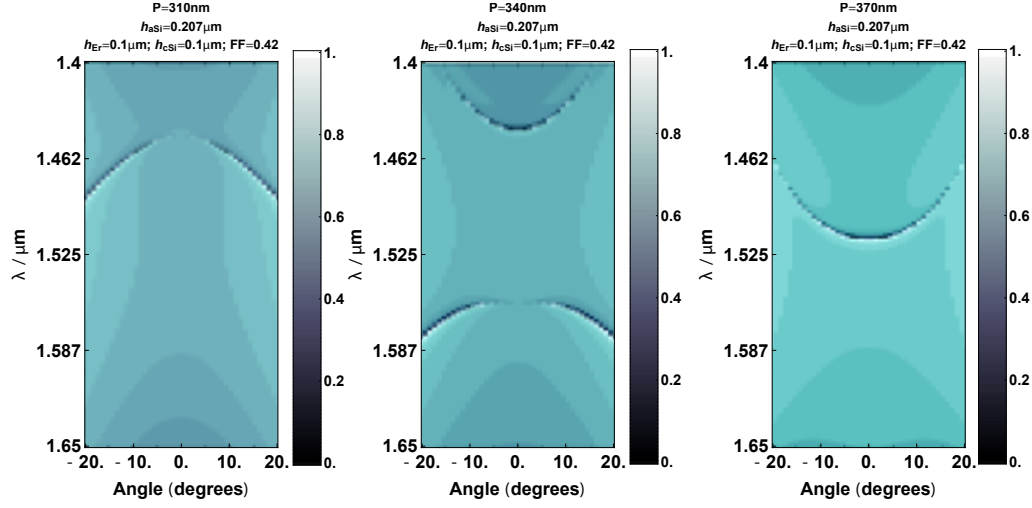


Figure 2.19: Simulated reflectivity spectrum of the devices at $FF = 0.42$, P varying from 310 nm to 370 nm, designed to exhibit flat BIC dispersion.

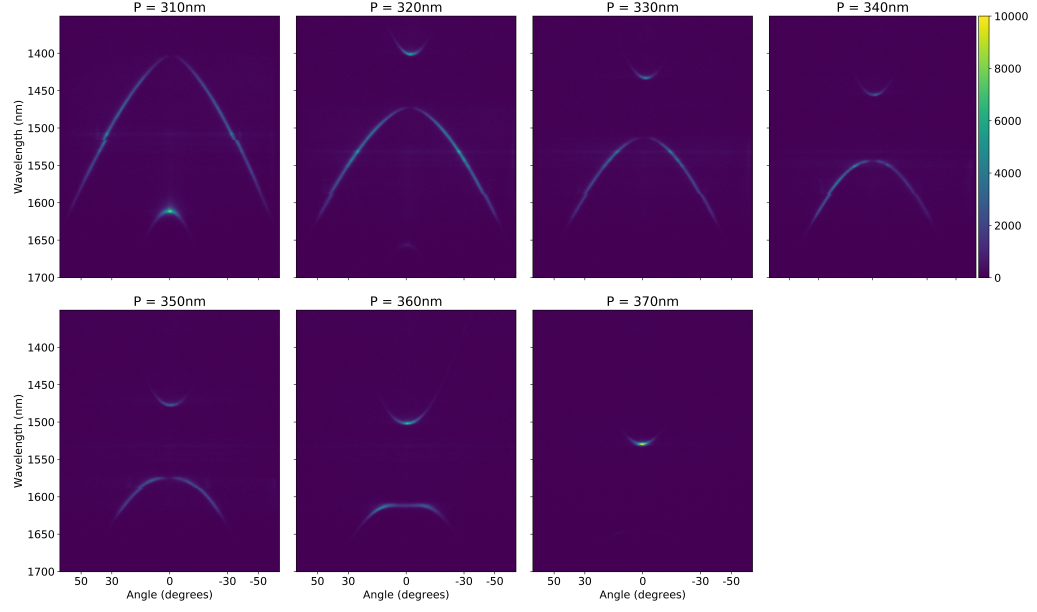


Figure 2.20: Measured energy momentum spectrum of the devices at $FF = 0.42$, P varying from 310 nm to 370 nm, designed to exhibit flat BIC dispersion.

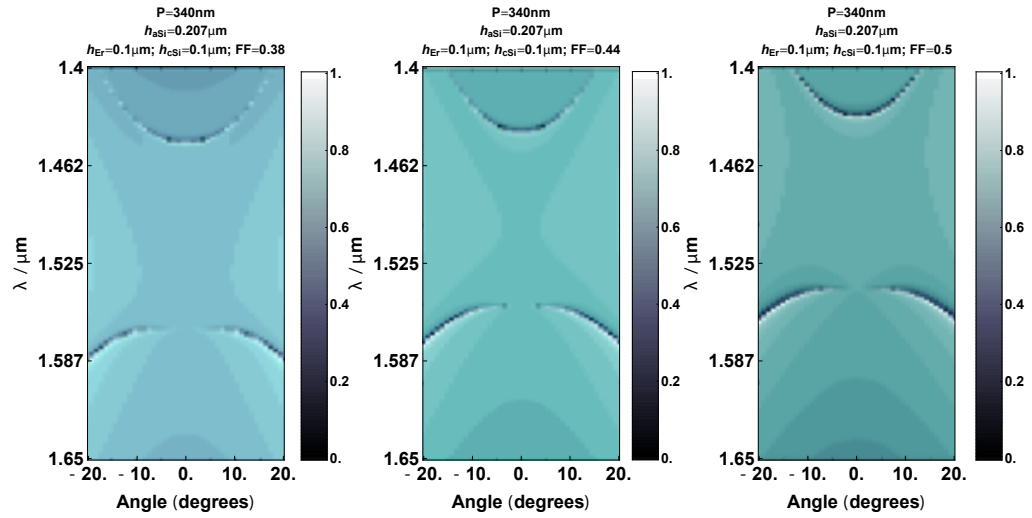


Figure 2.21: Simulated reflectivity spectrum of the devices at $P = 340\text{nm}$, FF varying from 0.38 to 0.50, designed to exhibit flat BIC dispersion.

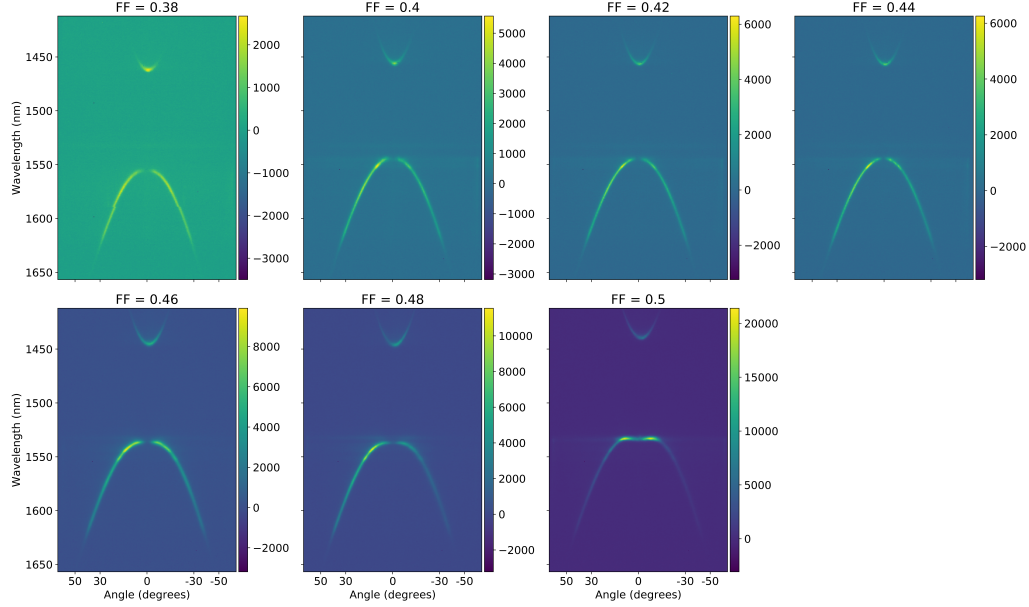


Figure 2.22: Measured energy momentum spectrum of the devices at $P = 340$ nm, FF varying from 0.38 to 0.50, designed to exhibit flat BIC dispersion.

2.6 Lifetime measurement of emitters in HCG

Due to the modification of the LDOS, it is expected that the excited state lifetimes of emitters should also be significantly affected. We therefore built a lifetime measurement setup to do time-correlated single photon counting measurements to investigate the change of emitters' lifetime by the structures.

The setup we built to measure lifetime of the emitters is illustrated in Fig 2.23. The sample was again excited by 532 nm laser illumination. The photoluminescence was fiber-coupled into an InGaAs single photon avalanche diode (SPAD, Micro Photon Devices), which generates a TTL pulse whenever a photon is detected. The TTL signals are then sent to a multi-channel analyzer (Stanford Research System SR430) which bins the TTL pulses with respect to the trigger. The optical chopper was operated at 40 Hz and acts as the source of the trigger. The trigger signal is sent to a pulse generator which creates a delay to best position the time bins with respect to the onset of the decay

signal, and is then sent to the multi-channel analyzer for triggering the start of the bins.

As the SPAD is sensitive to near infrared emission, there is high background noise when the bias voltage is applied. It therefore has to be gated periodically only for a short duration at a high frequency to achieve a high signal to noise ratio. After each trigger event due to the detection of a photon, the SPAD has a hold off time before it can be ready for the detection of the next photon. It is important to have a low probability of detecting a photon during this recovery time, as the arrival of a photon at this time would not be detected. This is known as the photon pile-up effect and must be avoided, as it would distort the shape of a decay curve and make the average measured lifetime appear shorter. Generally, a rule of thumb for time correlated single photon counting experiment is to have the count rate no more than 1 - 10 % of the gate frequency [36], imposing a constraint on the signal intensity.

On the other hand, as the lifetime of Er^{3+} at near infrared wavelength is on the order of ms, there is another constraint on the frequency of laser repetition rate. The optical chopper frequency is therefore set at 40 Hz. To acquire enough data points for sufficient signal to noise ratio, a decay experiment for a device therefore takes up to 8 hours.

We measured the lifetime of emitters on a region unpatterened to act as the control. Three devices were selected for the lifetime measurements. They are, respectively, devices designed for:

1. Bound state in continuum, with period = 620 nm, fill factor = 0.50.
2. Bound state in continuum, with period = 640 nm, fill factor = 0.50.
3. Dirac cone at 1300 nm, with period = 260 nm, fill factor = 0.40.

The binned PL decay curves are fitted to exponential decays to reveal the lifetime of the excited emitters. Mono-exponential fits were first conducted by fitting the decay to $BG + Be^{-t/\tau}$, where t is the time between the decay event and the onset of the laser (40Hz due to the optical chopping) and BG is the background found by taking the median of the binned intensity at long time (>18ms). However, it can be observed that the decay curves are not well fitted, by noting that in all decays

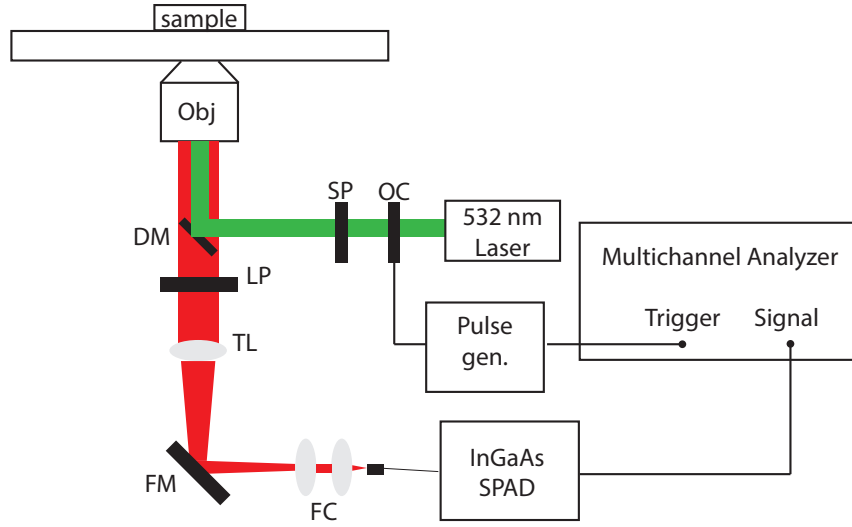


Figure 2.23: Setup for conducting lifetime measurement. Obj: Objective, DM: Dichroic mirror, SP: Short pass filter, OC: Optical chopper, LP: Long pass filter, TL: Tube lens, FM: Flip mirror, FC: Fiber coupler

the actual data decays faster than the fit at short time, and decays slower than the fit at long time. This reveals that the decay dynamics cannot be simply described by a single exponential.

We then attempted to try bi-exponential fits, by fitting the decay intensity to $A + Be^{-t/\tau_1} + Ce^{-t/\tau_2}$. The fits are more satisfactory and follow closely the trend of the decay curves. Table 2.1 summarizes the results of the fits. The lifetimes measured fall within the range of lifetimes of erbium doped yttrium in literature [37, 38]. The lifetime variations are all within 23% from the control device's lifetime. The bi-exponential decay behavior may be caused by photoluminescence not only from erbium but also from crystalline silicon. It is also possible that the bi-exponential decay is due solely to the erbium from the yttrium oxide layer as it has been previously observed in cases without the presence of silicon [38], in this case the bi-exponential decay may be explained by the meta-stable energy levels in erbium. To avoid this possible effect, in-band pumping of erbium at 1480 nm can be used to directly excite the electrons to the $^4I_{13/2}$ level. Also note that the emitters would experience different LDOS in the different parts of the structure.

To fully make sense of the measured lifetimes, one could carry out finite difference time domain simulations to approximate the local density of states experienced by the emitters proximal to the gratings. This can be done by placing unit dipoles at different locations of the emitter layer (yttrium oxide) and measuring the power radiated by the dipoles. As the emitters are isotropic in the thin film with no preferred directions, the simulations would have to be done for dipoles in all three directions and then summed.

	Mono-exponential fit: B (ms)	Bi-exponential fits: B, C (ms)
Device		
Control	3.47 ± 0.02	$4.03 \pm 0.06, 1.28 \pm 0.03$
BIC P620 FF0.50	4.00 ± 0.01	$4.93 \pm 0.04, 1.41 \pm 0.02$
BIC P640 FF0.50	3.27 ± 0.01	$3.89 \pm 0.05, 1.34 \pm 0.03$
Dirac P260 FF0.40	3.48 ± 0.01	$4.91 \pm 0.04, 1.45 \pm 0.02$

Table 2.1: Table summarizing the lifetime fits results

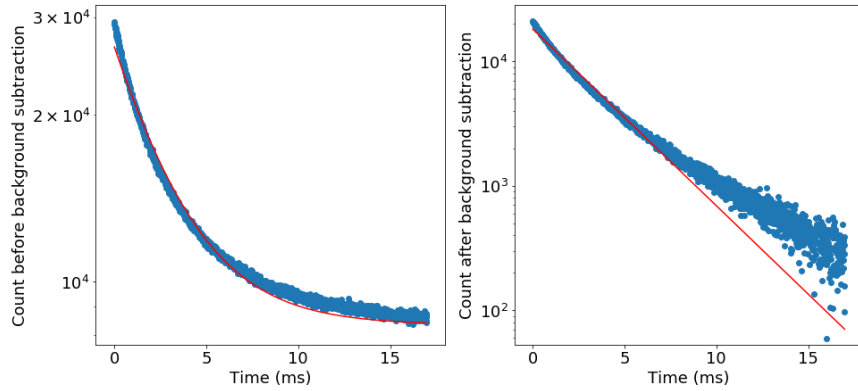


Figure 2.24: Control device PL decay fitted to mono-exponential

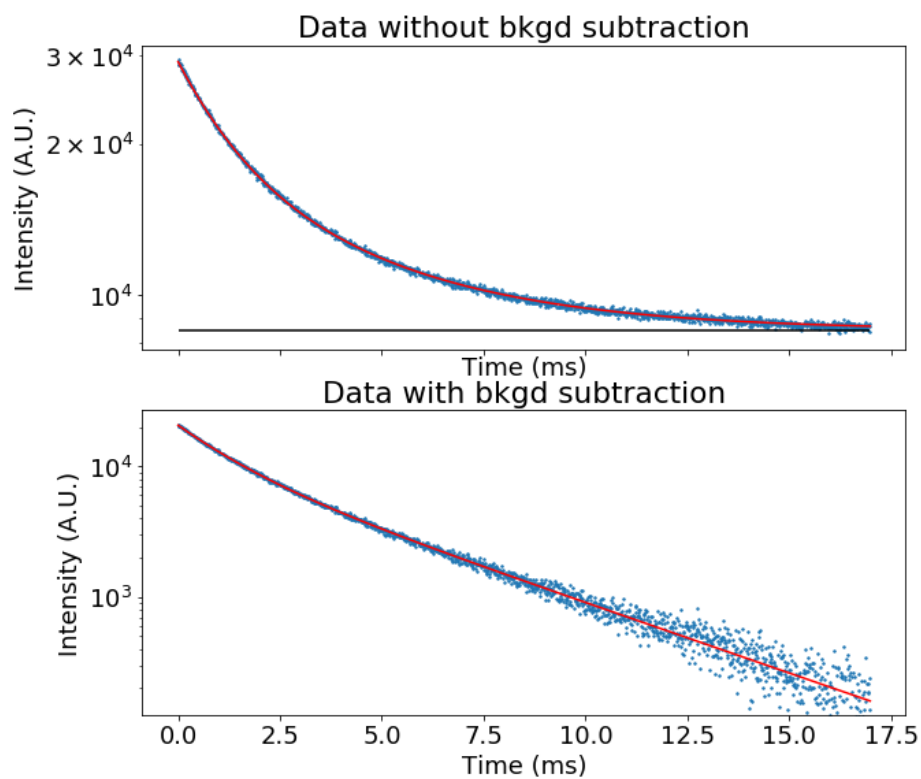


Figure 2.25: Control device PL decay fitted to bi-exponential

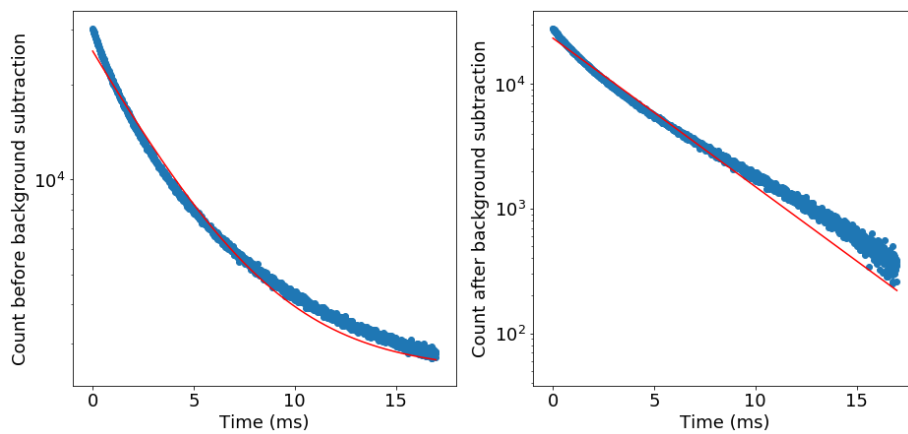


Figure 2.26: BIC P 620 device PL decay fitted to mono-exponential

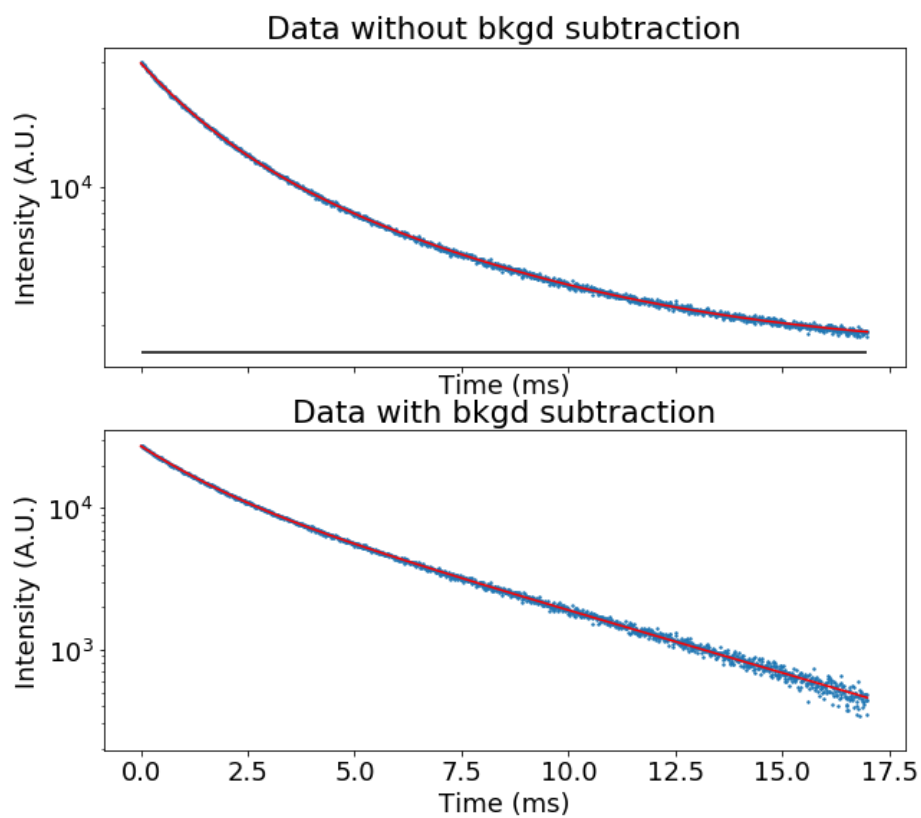


Figure 2.27: BIC P 620 device PL decay fitted to bi-exponential

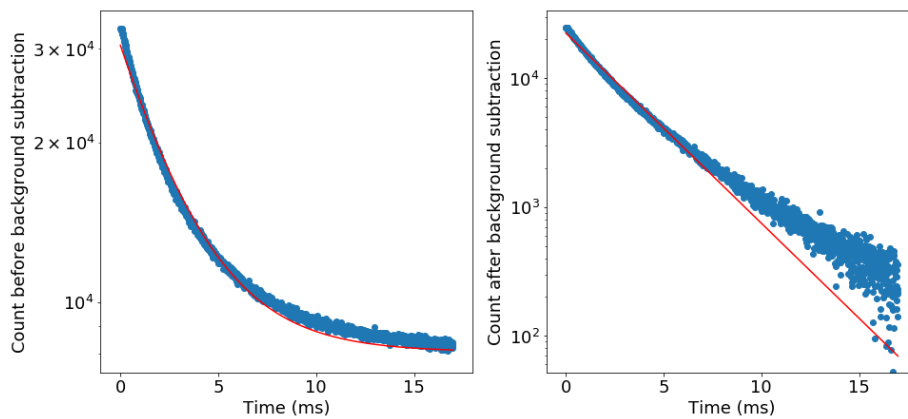


Figure 2.28: BIC P 640 device PL decay fitted to mono-exponential

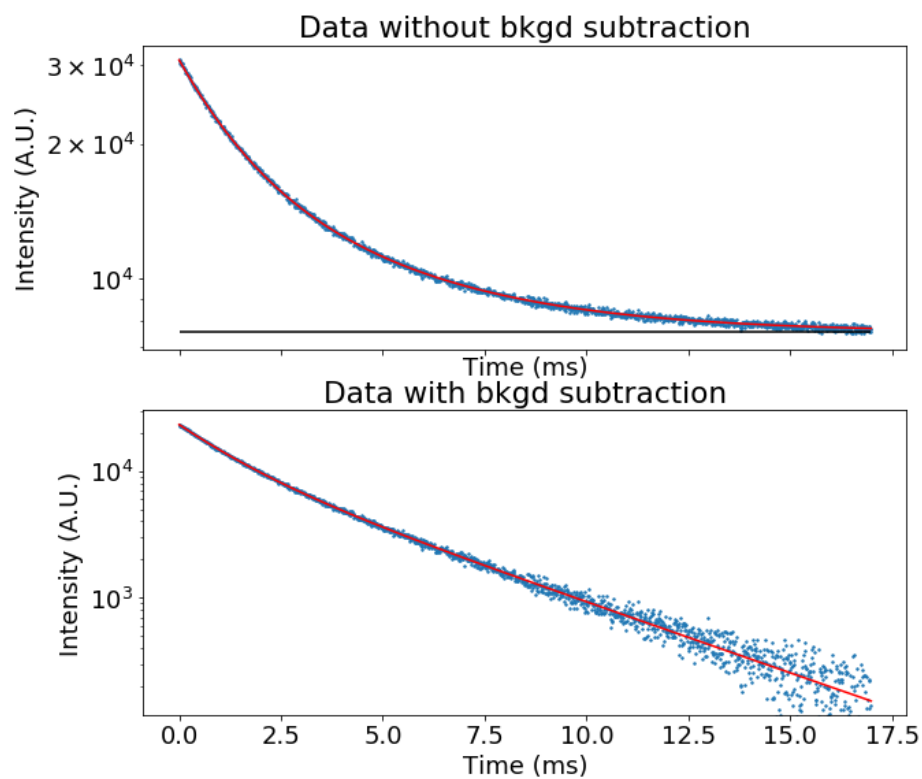


Figure 2.29: BIC P 640 device PL decay fitted to bi-exponential

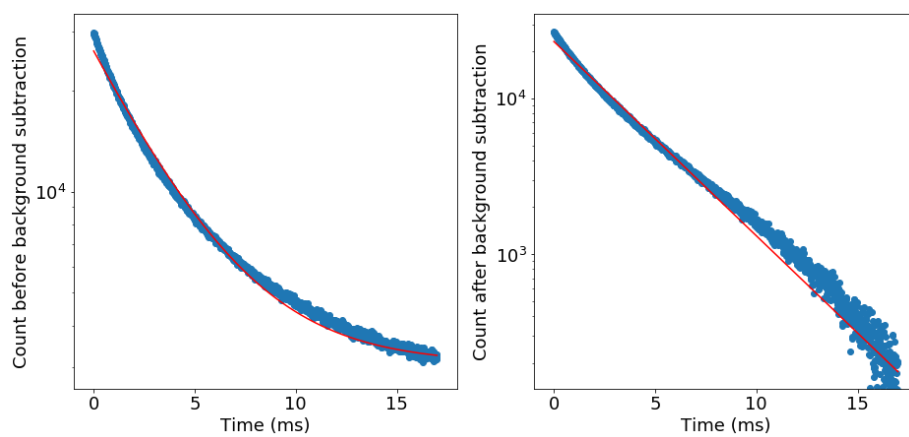


Figure 2.30: Dirac 1300 P 260 device PL decay fitted to mono-exponential

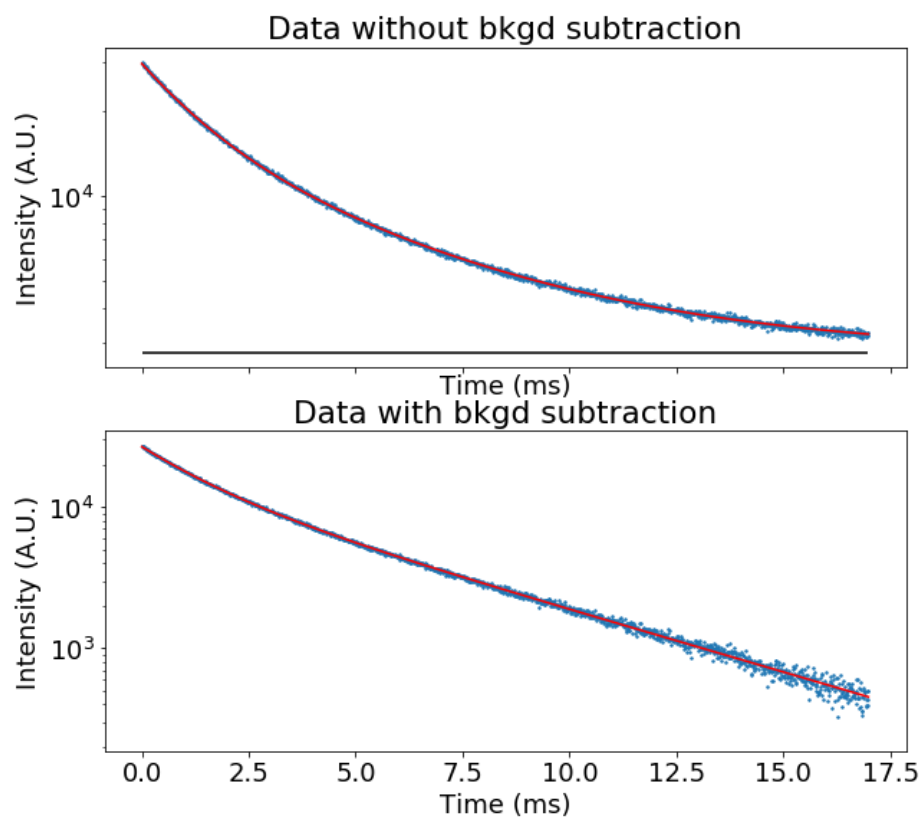


Figure 2.31: Dirac 1300 P 260 device PL decay fitted to bi-exponential

2.7 Conclusion

In this part of the thesis, we discussed the work on engineering the dispersion to enhance emission into particularly desired wavelength and direction. We then proceeded to investigate devices designed to exhibit exotic dispersions including Dirac dispersion, bound state in the continuum, and flat bands. These were experimentally realized and characterized, both by energy momentum spectroscopy and photoluminescence lifetime measurements. A clearer understanding of the measured lifetimes could come from FDTD simulations of the devices and the corresponding LDOS calculation.

Originally, it was anticipated that coupled with the strong background noise associated with the InGaAs detectors, the low intensity of the weak emitters would pose an experiment challenge, requiring special experimental design. In the end, it was found that simply having a long measurement time was sufficient. Due to this long measurement time, we partially automated the energy momentum spectrum measurement such that data could be continuously collected to efficiently characterize the devices. Unfortunately, the presence of the Bertrand lens complicates the auto-focus procedure. With the introduction of a lens flipper, it is believed that the whole measurement can be fully automated.

The weak signals and long measurement times we encountered in these experiments prompted us to re-evaluate the experimental design choices under low signal conditions in spectroscopy and more generally in imaging, which we will examine in the next chapter.

2.8 Acknowledgement

The simulations, fabrication, and experimental measurements described in this chapter are the result of the collaboration with Dr. Sébastien Cuffe's lab from CNRS (Centre national de la recherche scientifique) in France. At Brown we built the experimental setup and carried out the experiments for measuring the energy-momentum spectra and the photoluminescence lifetimes; we were also responsible for the deposition, annealing and characterization of the yttrium oxide layers. In the

subsequent part of the collaboration, we also contributed to the incorporation of erbium ions in the yttrium oxide layer to act as emitters. Dr. Sébastien Cuffe's lab carried out the RCWA simulations to guide the design of the devices and conducted the plasma enhanced chemical vapor deposition, electron beam lithography, and inductively coupled reactive ion etching to fabricate the grating structures.

Chapter 3

Hardware Binning Applied to Image Denoising and Spectroscopy

3.1 Introduction

The low-signal experimental conditions encountered in the previous work prompted us to think about the question: Are we designing our experiments in the optimal way? Did we make a conscious choice in conducting each part of the experiment?

As the most common experiments that we conduct are spectroscopy and imaging, we proceeded to scrutinize the way spectroscopy and imaging are done. To inform the optimal way of measurement, one must understand the experimental equipment and the noise sources involved. In the subsequent parts of the chapter, we focus on a very common photodetector, a charge-coupled device (CCD). After understanding the characteristics of different noise sources in CCDs, we investigate the use of hardware binning in spectroscopy and localization experiments under low signal conditions. The use of standard single pixel measurement scheme is challenged with statistical tools. We also discuss the use of numerical techniques in pixel intensity reconstruction and peak localization.

3.2 Common Photon Detectors

3.2.1 Charge-Coupled Device

Charge-coupled devices, commonly abbreviated as CCDs, are one of the most commonly used tools to measure light. It is effectively an array of silicon photo-diodes, each operating as an isolated photon detector. When a photon is absorbed by the detector, an electron-hole pair is generated. The various levels of doping and voltages applied to electrodes in the CCD create potential wells that attract and store electrons generated by photons. Electrons therefore accumulate throughout the exposure time in these potential wells. During this period, the shutter of the camera is open and allows photons to reach the detector. After this duration, the shutter is then closed to stop additional photons from reaching the detector. During read-out, the charges from each element are shifted to a read-out register by a series of clock voltages applied to the electrodes. They are then sensed as voltages, amplified, and fed to other electronics for storage or further analysis.

3.2.2 Complementary Metal-Oxide Semiconductor Imager

Similar to a CCD, a Complementary Metal-Oxide Semiconductor (CMOS) imager is composed of arrays of pixels. However, each one of them has its own electron to voltage converter and amplifier. The voltage is then fed into a multiplexer and transmitted off-chip for storage or further analysis. The presence of individual converters and amplifiers eliminates the need of ancillary circuitry as in the case of CCD. However, the trade-off is that the optical characteristics are usually not as good as CCDs; in particular, the dark current is generally higher, and both linearity and accuracy are worse. The fact that electrons are sensed and amplified by different amplifiers also introduces more significant fixed pattern noise in CMOS, compared to the case of one or two converters / amplifiers in the case of CCD. Also, CMOS imagers usually are more susceptible to fringing, leading to erroneous fringes / oscillations observed in the image / spectrum, because of the reflections between the dense metal wiring layers and the silicon surface.

3.2.3 CCD vs CMOS

CCD cameras have been the standard for scientific research imaging for decades, but, they suffer from a major drawback, that is, the speed of read-out. As the number of read-out amplifiers on a CCD is traditionally very small (mostly just one or two), the rate of charge-to-voltage conversion poses a significant bottleneck to its speed and noise characteristics. CMOS, on the other hand, has an amplifier for each pixel. This four to six order of magnitude difference in number of converters and amplifiers between CCD and CMOS therefore makes high-end CMOS cameras less noisy and faster than high-end CCDs [39].

That being said, CCDs have been the dominant scientific imaging cameras for spectroscopy, because of the far superior images with the available fabrication technology. Higher uniformity and smaller features are required for high-end scientific CMOS image sensors. Only recently did lithography technology advance to a stage that allows high-end scientific CMOS at reasonable costs to become better than high-end scientific CCDs in terms of noise and speed [39]. As a result, many labs still use CCDs or other variants like EMCCD or ICCD. Also, in the case of spectroscopy of homogeneous samples, the ability to hardware bin a thousand of pixels gives CCD a clear advantage over CMOS, as hardware binning is not possible in CMOS. Therefore, it is still very valuable to investigate the noise sources in a CCD and build a noise model to design well-informed ways of using CCD cameras to take measurements.

3.3 Noises and Configurations in CCD

Measurement is always a battle between signal and noise. One must therefore understand the sources and characteristics of noise in the equipment in order to investigate ways to obtain measurements with high signal-to-noise ratios.

The quantities that are of particular interest for a CCD are briefly discussed below:

1. Dark Current and Deepest Cooling Temperature

Dark current is due to the thermal generation of electron-hole pairs. Even when there is no light reaching the camera, there are still accumulated electrons in the pixels due to this dark current, giving it this name. As it is a thermal process, the generation rate is highly dependent on the temperature. The magnitude of dark current roughly follows Arrhenius law [2] as shown in Fig. 3.1. Due to this exponential inverse dependence, a good way to mitigate this thermal generation is to cool the camera. That is why scientific CCDs are usually deeply cooled to below -20°C . In most cases, dark currents are orders of magnitude smaller than other noise sources. For typical scientific grade CCD camera, the dark current per second is two to four orders of magnitude smaller than the system read noise. So in most cases they can be safely neglected. In cameras that are less cooled or having very long exposures (e.g. astronomical observations), the obtained image is usually subtracted by the expected rate of dark current multiplied by the exposure time to remove the contribution of dark current. The expected rate of dark current is pixel dependent but is usually relatively stable. The values can be determined beforehand by having a long and dark exposure. Since the thermal generation of carriers is a Poisson process, even after subtracting this number, there is still a component of Poisson shot noise that has a variance equal to the magnitude of carriers generated by the dark current.

2. CCD Format

The number of pixels, together with the magnification, determines the field of view. The size of pixels determines the upper bound of resolution (but for optical wavelengths, usually the diffraction limit is the limiting bound). Fill factor is the proportion of area that is sensitive to light versus the total area of the chip. Presence of electronics on the chip can reduce the fill factor, but the use of micro-lenses can help restore this loss.

3. Full Well

Full well describes the maximum number of electrons that can be stored in a pixel (potential well). When the number of electrons in a pixel exceed the full well, it can spill over into the readout circuitry or the substrate, resulting in spurious lines or leakage of charge into surrounding pixels.

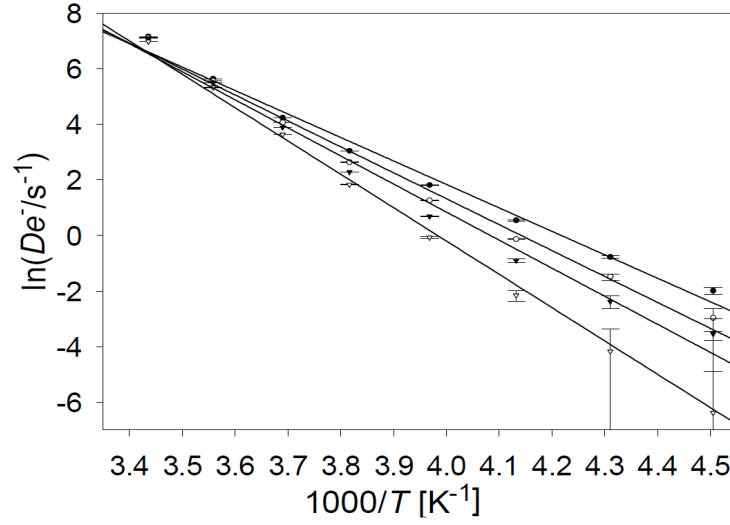


Figure 3.1: Dependence of dark current on temperature. Adapted from [2].

This is called blooming, and is particularly common in images which have both very bright and very dim parts.

4. ADC Speed / Bits

This describes the analog-to-digital conversion speed, in other words, how fast the voltages from a pixel (derived from number of electrons) can be converted into digital signal for further processing. This is the bottleneck in the speed of CCD cameras. Typical conversion speeds are in the order of 100kHz to 10MHz. For 100kHz and a 1024 x 1024 camera, a full read-out requires around 10 s. The number of bits determines the maximum number allowed by the ADC. An n -bit ADC is able to represent 2^n states.

5. System Read Noise

This is the noise introduced when the electrons stored in a pixel are read out. It is highly dependent on the architecture of the CCD and is an inherent property of the camera. It is a result of different noise sources which originate from amplifying and converting the electrons into voltages and digitizing them. It is usually inversely proportional to the read out speed - the faster the read out, the higher this noise is. It is typically considered to be Gaussian distributed [40–42]. A relevant

concept regarding read noise is hardware binning, which effectively turns a group of pixels into “super-pixels”. This can be done by modifying the usual clock voltages such that the electrons from these pixels are shifted into a same group when read out. As the number of read-out operations is decreased, the effective read noise also decreases.

6. Vertical Shift Speed

This refers to how fast charges in a pixel can be shifted vertically to the read-out register. This directly affects how fast each image can be read out.

7. Non-linearity

Linearity refers to whether the number of detected photons (usually referred to as Analog Digital Unit (ADU) or counts) scale linearly with the number of incident photons onto the CCD camera. CCDs are usually extremely linear unless there are specific anti-blooming electronics in the chip [43].

8. Photon Shot Noise

Photon shot noise is due to the intrinsic particle nature of light. It follows a Poisson distribution, so the variance equals to the mean of the distribution. The only way to combat this noise is to increase the photon flux as the signal-to-noise ratio scales with the square root of the signal.

9. Fixed Pattern Noise

This is sometimes considered to be a noise when considering all the pixels as a whole. This is due to the slight difference in sensitivities in pixels of the sensor. It can be combated with proper calibration technique including flat fielding and wavelength intensity calibration when a camera is coupled to a spectrometer.

As an example, Fig. 3.2 shows a specification sheet from Princeton Instruments, a well-known vendor of scientific instruments. Relevant values for the discussed parameters can be referenced from it.

For scientific CCD cameras, the noise sources that dominate are read noise and photon shot noise in most cases [44–48]. One can therefore build a model that incorporates these two noise sources and devise ways to more intelligently take measurements with CCD cameras to gain more information

SPECIFICATIONS

	 PIXIS: 1024BR_eXcelon	PIXIS: 1024BR	 PIXIS: 1024B_eXcelon	PIXIS: 1024B/BUV	PIXIS: 1024F
Features	Back-illuminated, deep depletion CCD with eXcelon technology. Highest QE in the UV and the NIR. No etaloning.	Back-illuminated, deep depletion CCD. High QE in the NIR and no etaloning.	Back-illuminated CCD with eXcelon technology. Highest QE in the visible and high QE in the NIR. Extremely low etaloning. 5x - 100x lower dark charge than the BR.	Back-illuminated CCD. Highest sensitivity in the visible region. Special BUV version offers the highest sensitivity in the UV region.	Front-illuminated CCD. Affordable technology for moderate light level applications. No etaloning.
CCD Image Sensor	Princeton Instruments' proprietary CCD with eXcelon technology, grade 1, NIMO	e2v CCD47-10 back-illuminated deep depletion, grade 1, NIMO	Princeton Instruments' proprietary CCD with eXcelon technology, grade 1, AIMO	e2v CCD47-10 back-illuminated, grade 1, AIMO	e2v CCD47-10 front-illuminated, grade 1, AIMO
Dark current @ -70°C (e-/p/sec)	0.02 (typical) 0.07 (max)	0.02 (typical) 0.07 (max)	0.0004 (typical) 0.001 (max)	0.0004 (typical) 0.001 (max)	0.0002 (typical) 0.0007 (max)
CCD UV coating	Optional UV coating (not needed for BUV version)				
CCD format	1024 x 1024 imaging pixels; 13 x 13 μm pixels; 100% fill factor				
Imaging area	13.3 x 13.3 mm (optically centered)				
Lens mount	Adjustable C-mount with integral 25mm shutter; spectrometer adapter available				
Deepest cooling temperature	-90°C typical; -70°C guaranteed, specified at ambient temperature of +20°C				
Thermostating precision	$\pm 0.05^\circ\text{C}$				
Cooling method	Thermoelectric air or liquid cooling (CoolCUBE II liquid circulator available)				
Full well:	100 ke- (typical), 60 ke- (min)				
Output node	250 ke- (typical), 220 ke- (min)				
ADC speed/bits	100kHz/16-bit and 2MHz/16-bit				
System read noise @100 kHz	3.0 e- rms (typical), 5 e- rms (max)				
@2 MHz	9.0 e- rms (typical), 15 e- rms (max)				
Vertical shift speed	< 3.2 $\mu\text{sec}/\text{row}$ to 18 $\mu\text{sec}/\text{row}$ (programmable)				
Non-linearity	<1% @ 100 kHz				
Software selectable gains	1, 2, 4 e-/ADU (typical); available at all speeds				
Operating systems supported	Windows 8/7 (64-bit) and Linux (64-bit), Windows 8/7/XP (32-bit)				
Data interface	USB2.0 (5m interface cable provided); Optional Fiberoptic interface is available for remote operation				
I/O signals	Two MCX connectors for programmable frame readout, shutter, trigger in				
Operating environment	+5 to +30°C non-condensing				
Certification	CE				
Dimensions / Weight	16.59 cm (6.53") x 11.81 cm (4.65") x 11.38 cm (4.48") (L x W x H) / 2.27 kg (5 lbs)				

All specifications subject to change

FRAME RATE

Binning	Readout Time	
	@ 2 MHz	@ 100 kHz
	1 x 1	0.58 sec
	2 x 2	0.28 sec
	8 x 8	0.14 sec

Figure 3.2: A standard specification sheet of a scientific CCD from Princeton Instruments. Adapted from [3].

within a shorter time.

3.3.1 CCD Subtypes

There have been efforts trying to combat the long readout time drawback of CCDs, creating variants from the traditional simple full frame camera, such as inter-line and frame transfer camera.

Full Frame Camera

The simplest configuration is the full frame camera. The whole pixel array is used to receive photons and convert them to electrons. There is usually either a mechanical shutter to block light when charge is being read out or the experiment is done with pulsed light source such that no light is incident on the sensor when charges are being read out. Otherwise, smearing will occur, meaning that photons still hit the sensor when the charges are being read out. There will therefore be ghost images of objects at positions that they should not be. For high speed and dim applications, this can be a problem as the read-out time can be in the order of second to tens of seconds, meaning that the image can be severely corrupted while it is being read out.

The major problem of full frame camera is that it requires a long read-out time. Due to this long overhead, unless an intelligent de-smearing algorithm is applied, the duty cycle of the camera is strongly restricted, leading to challenges in its usage in high speed and low-light applications. The advantage, however, is that it is simple in the hardware sense and fully utilizes the whole pixel array.

Inter-line Camera

Interline cameras were designed to help solve the problem of long readout time overhead. They consist of an array of masked pixels immediately next to an array of unmasked pixels. After the exposure, charges are shifted into the masked pixels right away. The vertical shift to the readout register and readout is then done in the masked pixels. This allows the next exposure to be done in parallel to the shifting and readout, greatly increasing the maximum duty cycle to near 100%.

The problem of interline cameras is that the fill factor of the pixel array is seriously lowered, since approximately half of the CCD is now masked and unable to detect photons. There have been attempts to combat this issue by introducing micro-lenses that direct light originally hitting the masked area to the unmasked area, and the effective fill factor can be restored to be close to 100%. However, this also suffers from several drawbacks. Firstly, micro-lenses are most effective when the light is incident in the normal direction onto the pixels. The relatively low NA of these micro-lenses restricts the collection of light from larger entrance angles. Secondly, micro-lenses are commonly made with a polymer that has a poor UV transmission, leading to a low quantum efficiency of the sensor at short wavelengths.

Frame Transfer camera

Frame transfer cameras are another popular option in CCD. They are similar to interline cameras in that there is also a masked area which is not sensitive to incoming photons in them. However, the masked area is not immediately next to the light-sensitive pixels. Instead, it is located between the readout register and the unmasked pixels. This means that the fill factor of the sensor is unaffected and remains 100% without the use of micro-lenses. After each exposure, charges are quickly shifted to the masked area, serving as a charge storage region. This duration is usually short (in the order of millisecond, depending on the camera [49]). After shifting the charges to storage, exposure can be continued and readout can be carried out in parallel, so the exposure duty cycle can still be close to 100%.

The drawback of frame transfer cameras, is that only half of the camera is light-sensitive. This raises the cost of the whole sensor as a much larger chip is required to make a CCD with the same light sensitive area.

3.4 Characterizing a CCD camera

To investigate the optimal way to measure with CCDs, one must have a good understanding of the noise sources of the sensor. As an illustrative example, we carried out characterization measurements on a Princeton Instruments camera PIXIS 1024B.

3.4.1 Background Count and Read Noise

To begin with, one should take many dark frames (with shutter closed or with no light incident on the sensor) to understand the behavior of pixel background and readout noise.

At an ADC speed of 2 MHz 30000 frames were taken, and 9000 frames were taken with the ADC at speed of 100 kHz. The mean and standard deviation of every single pixel over all the frames was then calculated and plotted. Histograms showing the occurrences of each pixel intensity are also plotted in Fig 3.3. Note that the readout time is of the order of seconds to ten seconds (<0.6 s for 2 MHz and <10 s for 100 kHz), so the dark current effect is minimal (it is on the order of 0.002 ADU / pixel / s, see subsequent section).

Fig. 3.3 and Fig. 3.4 show the mean of the measurements for 100 kHz and 2 MHz ADC speed respectively. They are the baselines of 1024 x 1024 pixels of the camera. The variations in baseline are ≈ 3 counts and within 60 counts respectively. Even though the variation in baseline for 2 MHz is relatively large, there is not much temporal variation in baseline for individual pixels, as shown in Fig. 3.7 which shows the temporal variation in baseline for 8 select pixels at 2 MHz. This means that the counts detected can be safely subtracted by its background values to obtain the actual signal incident on the pixel.

Fig. 3.5 and Fig. 3.6 show the standard deviations of the measurements for 100 kHz and 2 MHz ADC speed respectively. They can be regarded as the read noise of 1024 x 1024 pixels of the camera. There appears to be some well defined structures in the density plots for the sensor, probably corresponding either to electronics on the chip or uneven cooling of the sensor.

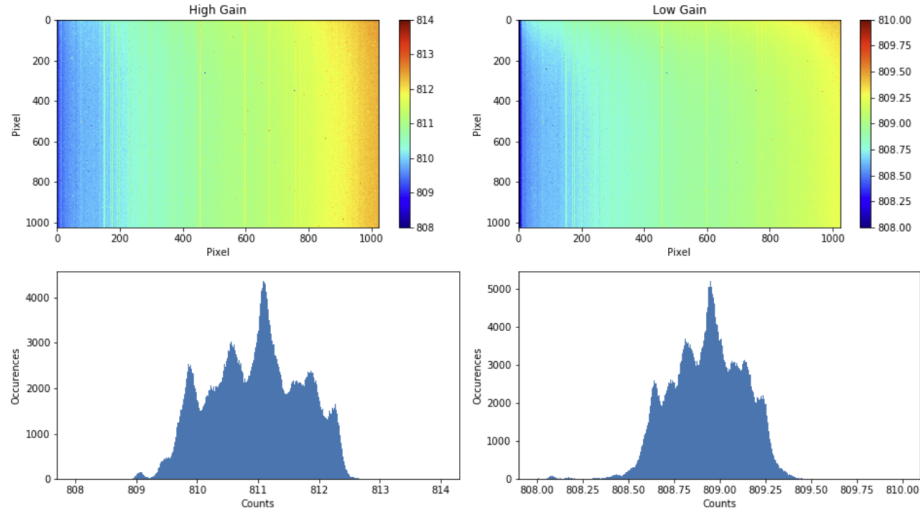


Figure 3.3: Baseline of PIXIS at 100 kHz

To investigate the normality of readout noise, which is a common assumption in literature [42, 44–48], we took 30000 frames for 4 different pixels at high gain and 2 MHz ADC speed. Their resulting histograms are shown in Fig. 3.8, together with Gaussian fits to the occurrences. The graphs can be seen to be pretty close to Gaussian, though there are higher occurrences at certain intensities.

3.4.2 Dark Current

To measure the dark current of the CCDs, a long exposure time is needed due to its usual low value. We conducted an experiment on the PIXIS camera by hardware binning 1024 pixels together and taking spectrum for 3 hours at -70°C . The resultant measurements are illustrated as a histogram in Fig 3.9. The median dark current is 1.99 ADU per 1024 pixels per second, corresponding to approximately 0.002 ADU per pixel per second.

3.4.3 Photon Transfer Curve

A photon transfer curve is a graph of noise vs signal which shows the dominance of different noise sources at different photon flux levels [50]. The most accurate way to characterize the camera

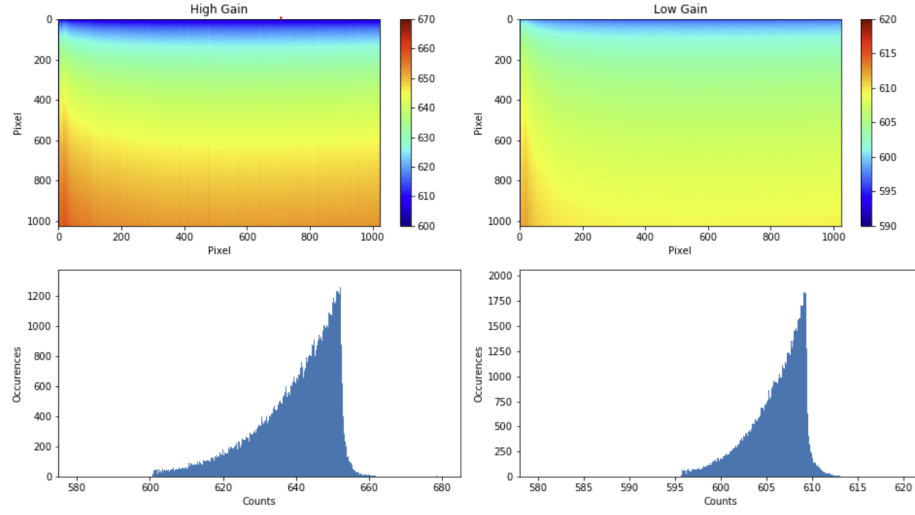


Figure 3.4: Baseline of PIXIS at 2 MHz

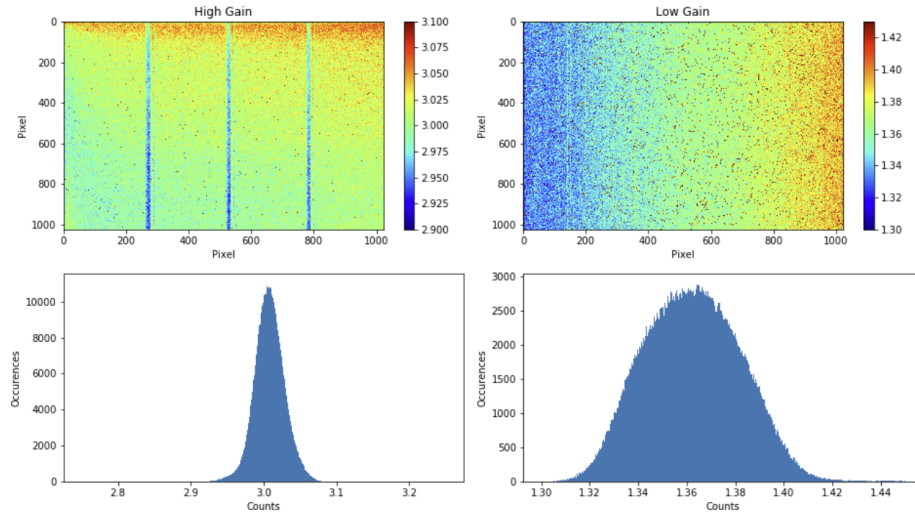


Figure 3.5: Baseline SD of PIXIS at 100 kHz

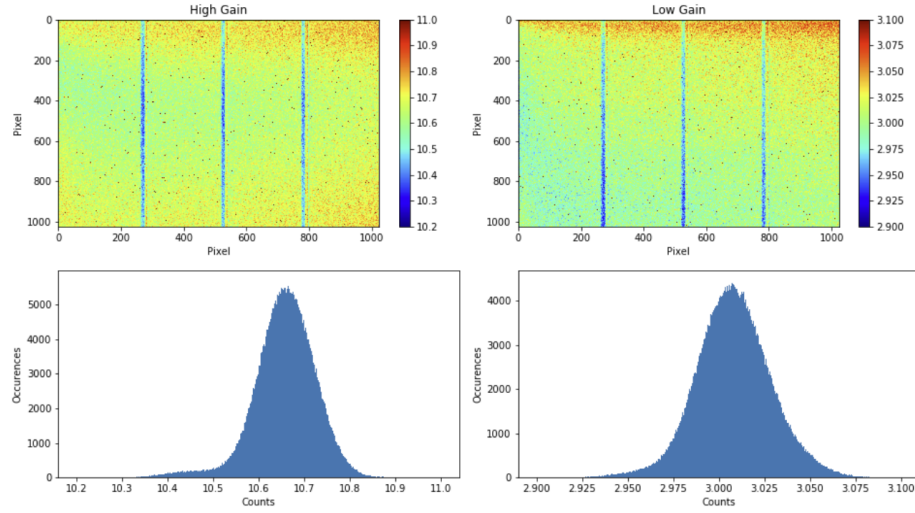


Figure 3.6: Baseline SD of PIXIS at 2 MHz

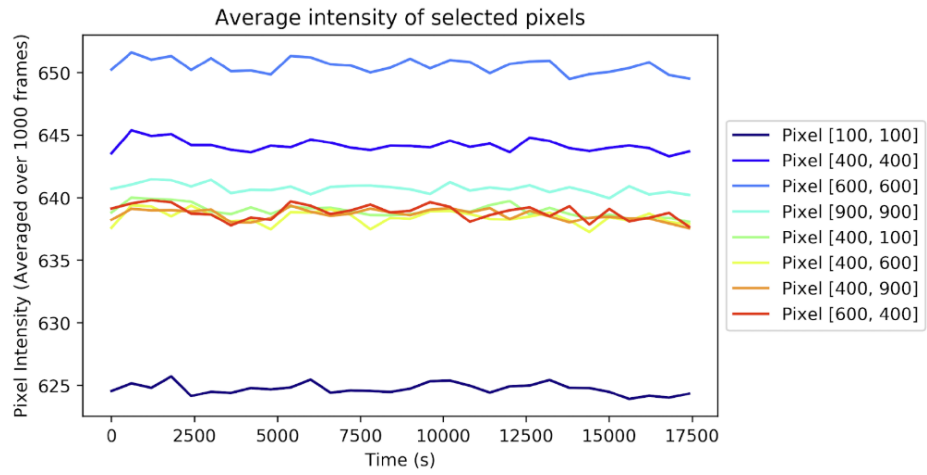


Figure 3.7: Temporal variation in baseline for 2 MHz. 8 pixels are selected and 30000 frames are taken. The average of 1000 frames is plotted as a function of time.

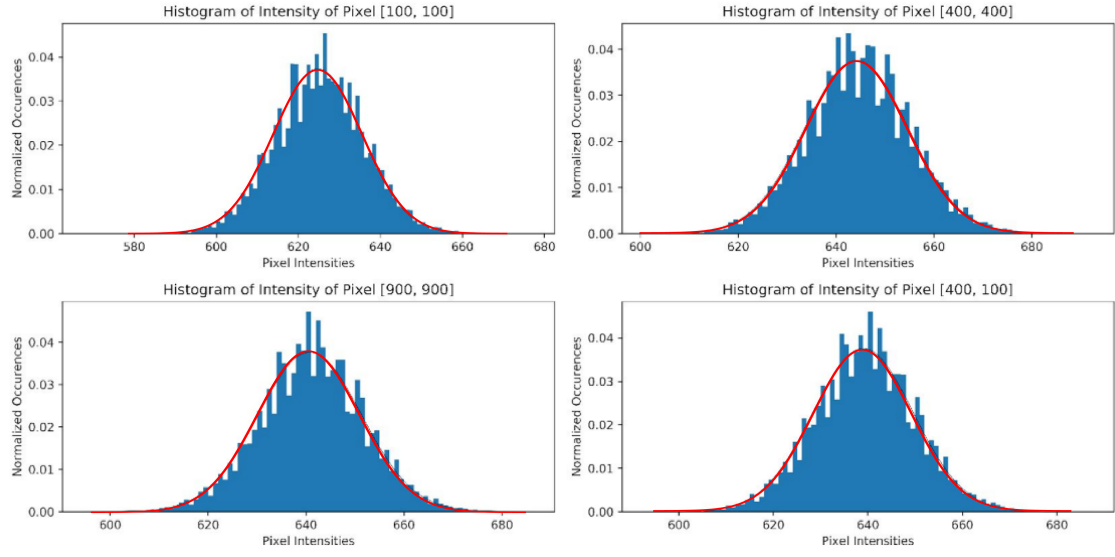


Figure 3.8: Histogram of baseline for four different pixels on PIXIS, overlaid with Gaussian fits

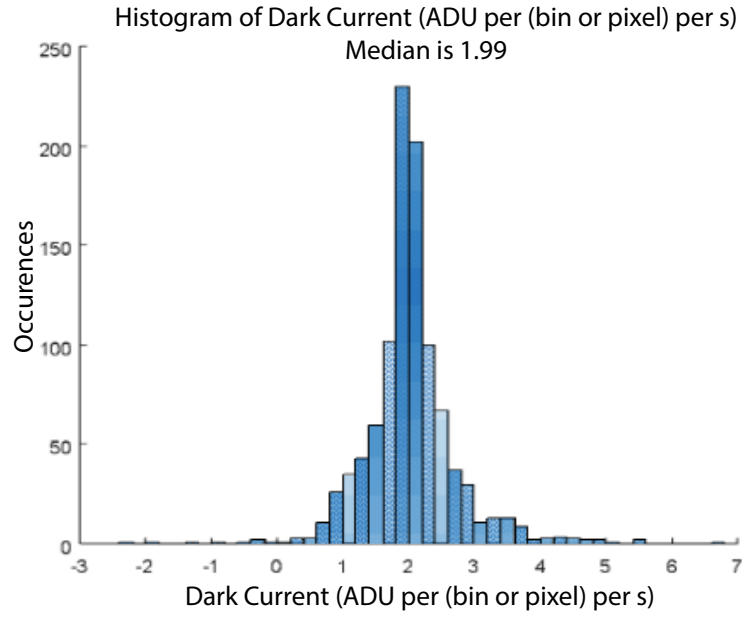


Figure 3.9: Histogram of dark current for PIXIS

would be to construct a photon transfer curve for every single pixel by measuring the variance of each pixel at different illumination levels. However, by having similar pixels and the same readouts by construct, CCDs have relatively uniform noise statistics and sensitivity across all pixels. For efficiency, a simplification that is usually adopted is to construct one single photon transfer curve for the whole CCD, and treat different pixels equally so that they act as different experiments. Under a uniform illumination, the noise at a certain illumination can then be easily estimated just with two exposures, as illustrated in the following discussion.

Following the treatment in [50], the total noise in CCDs can be obtained by the quadrature sum of read noise σ_R , shot noise $\frac{S}{K_{\text{ADC}}(e^-/\text{DN})}$ (where $K_{\text{ADC}}(e^-/\text{DN})$ is the electron to digital number sensitivity at the ADC and S is the signal level) and $P_N S$ is the fixed pattern noise (which originates from the difference in sensitivity of pixels, therefore scales linearly with the signal).

$$\sigma_{\text{Total}} = \sqrt{\sigma_R^2 + \frac{S}{K_{\text{ADC}}(e^-/\text{DN})} + P_N S} \quad (3.1)$$

By taking the difference between two frames, therefore, the sum of read noise and shot noise can be obtained:

$$\begin{aligned} \sigma_{\Delta} &= \sqrt{\frac{\sum_{i=1}^{N_{\text{Pix}}} (\text{DN1}_i - \text{DN2}_i)^2}{N_{\text{Pix}}}} \\ &= (\sigma_R^2 + \frac{S}{K_{\text{ADC}}(e^-/\text{DN})}) \times \sqrt{2} \end{aligned} \quad (3.2)$$

By varying either the exposure time or the illumination strength, plotting the noise as a function of signal gives us the photon transfer curve. Fig. 3.10 shows an example photon transfer curve from [4]. There are roughly four regimes that can be distinguished: (1.) When the signal is very weak, the noise is dominated by read noise. This is called the read noise floor as the noise does not actually go to zero as signal goes to zero, leading to a horizontal line in noise versus signal. (2.) When signal is further increased, it is then dominated by the signal shot noise, which has a standard deviation proportional to the square root of the signal and the sensitivity. That is why the slope of the shot noise dominated part is approximately half on a log-log plot. (3.) When signal is further increased, fixed pattern noise dominates. Note that fixed pattern noise originates from the

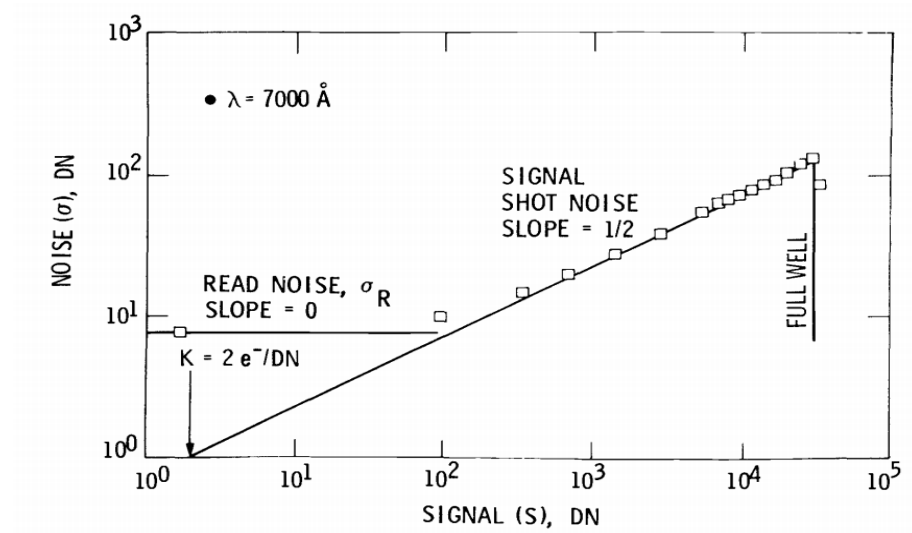


Figure 3.10: Example Photon Transfer Curve, by Janesick et al. [4].

difference in sensitivities of the pixels, and therefore its standard deviation scales with the signal itself, leading to a slope close to one. (4.) When signal is further increased, some pixels start to saturate, meaning that the number of electrons in each pixel reaches the full well. The detected count therefore clips at the maximum number (defined by the well capacity of the physical pixel or the output node), making the standard deviation across pixels decrease. The noise (variance) will eventually go to zero as all pixels become saturated.

Note that in practical applications when we consider values in each pixel, fixed pattern noise is not actually a noise. It is simply a slight difference in gain of individual pixels that can be usually ignored for CCD when proper calibration, like flat field or wavelength calibration when coupled to a spectrometer, is carried out.

3.5 Hardware binning and its application

3.5.1 Introduction

For low signals, read noise becomes comparable to the intrinsic shot noise in magnitude and becomes an important noise source. From the previous discussion, we understood that it arises due to the read out process of values in a pixel. One way to mitigate such read noise is to minimize the number of read outs. This can be done by the technique of hardware binning.

Hardware binning is a technique to combine the reading of pixels into reading operations on super-pixels. This involves a modification of the charge transfer clock voltages so that carriers from several (or more) pixels are grouped to a physical pixel before read out. As a result of grouping of k pixels, the number of times of read out is decreased k -fold, thereby avoiding the addition of readout noise by k -fold.

The readout noise during hardware binning is more or less similar, because the same read out process is involved. It is generally considered (and experimentally verified [51]) to be independent of the signal and binning condition [4, 52, 53].

One should be aware of the difference between hardware binning and software binning. Hardware binning (also known as analog summation) is done by physically grouping the charges of one or more pixels together before reading out and digitizing. It decreases the number of read outs required. Software binning (also known as digital summation), on the other hand, involves reading out and digitizing each pixel one by one, then storing and summing the digitized intensities. It therefore requires the same number of read outs as single pixel measurements. As a result, only hardware binning can actually reduce the readout noise and enhance read out speed. Note that in the case of a CMOS camera, each pixel has its own readout and amplifier, so hardware binning is not possible.

Even though read noise can be reduced through hardware binning, there are two trade-offs: Spatial resolution is lowered as a result, and if the number of charges exceeds the well-capacity of a pixel, the additional charges will spill to neighboring pixels (blooming) and lead to data corruption.

One therefore expects that hardware binning is beneficial under the following scenarios: (1.) When the signal is weak (comparable to the magnitude of the readout noise); (2.) When the desired spatial resolution is lower than the constraints due to pixel size.

One natural question to ask is therefore, under low signal conditions, whether we can intelligently design hardware binning schemes to mitigate the contribution of read noise, enhancing the signal-to-noise ratio and gaining more information than we would normally do with single pixel measurements. This shall be the topic of the following sections.

3.5.2 Prerequisite: A noise model for analysis

Before we can examine the optimal use of CCD cameras through hardware binning, a data model is needed to describe how different noise contributions can be considered together. As discussed in the previous part, the major noise sources under ordinary operation conditions are read noise and shot noise. We therefore start with a noise model consisting of additive Poisson shot noise and Gaussian readout noise, simplified from that introduced by Snyder et al. [42]. This noise model is also used in several other literature [44–48], to be described in this section.

A condition is further imposed that the baseline / mean of the Gaussian readout noise is set to be zero. This is often the practical case; the baseline of a pixel can be determined beforehand by repeatedly taking dark frames and taking the average, because this baseline is in most cases a constant over time and independent of flux. As the most significant contribution to readout noise is usually the on-chip amplifier and charge-to-voltage converter [54], minimizing the number of read outs reduces the amount of readout noise corruption in the measurement. Hardware binning does exactly that, and therefore can lead to higher signal-to-noise ratio at the expense of spatial resolution. As discussed previously, the readout noise introduced is approximately independent of the signal and binning condition [4, 51–53], so this is assumed in the subsequent analysis.

For each measurement i , whether it be a single pixel or an effective super-pixel (a group of hardware binned pixels), assuming that the flux $F(i, t)$ is a constant from $t = 0$ to $t = T$, the count

on this measurement can be considered to be a sum of two random variables, one representing the Gaussian readout noise, and the other one representing the signal with a Poissonian shot noise.

So, the detected count $C(i)$ for the measurement i can be modeled as:

$$C(i) = Q(i) + G(i) \quad (3.3)$$

where $Q(i) \sim \text{Poisson}(F(i) * t)$ is a Poissonian random variable with mean $F(i) * t$ and variance $F(i) * t$ and $G(i) \sim \text{Normal}(0, \sigma_R^2)$ is a Gaussian random variable with mean 0 and variance σ_R^2 . Since the read noise component is independent of the signal, these two random variables are independent of each other, therefore:

$$\begin{aligned} E[C(i)] &= E[Q(i)] + E[G(i)] = F(i)t + 0 = F(i)t \\ V[C(i)] &= V[Q(i)] + V[G(i)] = F(i)t + \sigma_R^2 \end{aligned} \quad (3.4)$$

where $E[X]$ denotes the expectation value of a random variable X over an ensemble measurement and $V[X]$ denotes the variance of X .

With this noise model, we are now equipped to proceed to analyze the application of hardware binning to two problems: pixel intensity reconstruction and localization.

3.6 Pixel Intensity Reconstruction

Image pixel intensity reconstruction, also known as denoising, is a popular research topic in many fields including astrophysics [41, 42, 55], computerized tomography [56, 57], and computer vision [58, 59]. The problem can be phrased in the following way: Given a set of measurements corrupted by noise and the knowledge of the underlying noise distribution, how can the actual intensities of pixels be estimated?

Expectation-Maximization algorithm (EM algorithm), first explained and named by Dempster et al. [60], is a commonly used algorithm in denoising [41, 42, 55, 56, 61]. It is a numerical method based on maximum likelihood estimation (MLE), which is one of the most widely used methods in

statistics to infer hidden parameters from observed data. In particular, the ordinary least square fitting of data in linear regression is a special case application of maximum likelihood estimation when the errors are governed by a normal distribution. In essence, MLE is an estimation of the underlying parameters of a model that maximizes the probability of resulting in the observed data. It is popular in statistics because it is intuitive and has many desired properties in estimation (e.g. sufficiency, consistency, and efficiency [62]).

This algorithm is also widely used in other fields like data clustering where it can be used as a soft-clustering method as an alternative to the k-means algorithm [63], and in natural language processing where it is part of the Baum-Welch algorithm for hidden Markov models [64].

The details of the EM algorithm applied to more complicated CCD models have been discussed thoroughly [41,42,65,66]. In this manuscript we simplify these approaches and just focus on recovering the ground truth flux of each pixel with hardware binned data. To the author's best knowledge, there has been no application of this widely popular EM algorithm on hardware-binned data, allowing information to be extracted from multiple measurements with varying super-pixel sizes. In the following sections, we first derive the likelihood function corresponding to the measured counts $c(i)$. The EM algorithm is then applied to pixel intensity reconstruction.

3.6.1 Derivation of the Likelihood Function

In all that follows, to simplify notations, all probabilities are written assuming that $c(i)$ and σ_R are known, and the conditional $f(i)t$ is implied. Starting from equation 3.3, by the law of total probability, for measurement i , the probability of detecting $c(i)$ is given by:

$$P(C(i) = c(i)) = \sum_{q(i)=-\infty}^{\infty} P(C(i) = c(i)|Q(i) = q(i)) * P(Q(i) = q(i)) \quad (3.5)$$

The probability of $Q(i) = q(i)$ is simply the probability mass function for a Poisson distribution given ground truth flux $f(i) \times t$:

$$P(Q(i) = q(i)) = e^{-f(i) \times t} \frac{(f(i) \times t)^{q(i)}}{q(i)!} \quad (3.6)$$

While the other part $P(C(i) = c(i)|Q(i) = q(i))$ is simply the probability of the random variable originating from the read noise returning $c(i) - q(i)$, which can be from ∞ to $-\infty$.

$$P(C(i) = c(i)|Q(i) = q(i)) = \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{\frac{-(c(i)-q(i))^2}{2\sigma_R^2}} \quad (3.7)$$

So, the probability of measurement i measuring $c(i)$ is (noting that the Poisson random variable only returns non-negative numbers):

$$P(C(i) = c(i)) = \sum_{q(i)=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{\frac{-(c(i)-q(i))^2}{2\sigma_R^2}} e^{-f(i)t} \frac{(f(i)t)^{q(i)}}{q(i)!} \quad (3.8)$$

The log-likelihood function, corresponding to all measurements $i = 0, 1, \dots, m$ is therefore:

$$\begin{aligned} \ln(L(f; C)) &= \ln \left(\prod_{i=0}^m P(C(i) = c(i)) \right) \\ &= \sum_{i=0}^m \ln \left(\sum_{q(i)=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{\frac{-(c(i)-q(i))^2}{2\sigma_R^2}} e^{-f(i)t} \frac{(f(i)t)^{q(i)}}{q(i)!} \right) \end{aligned} \quad (3.9)$$

Because of its complicated form, there is no analytical closed form solution to the optimization problem of maximum likelihood. We therefore resort to the EM algorithm which is commonly used for intricate likelihood functions.

3.6.2 Mathematical Details of the EM Algorithm Applied to Denoising

The EM algorithm consists of two steps: the expectation step and the maximization step. In the expectation step, the likelihood function is rewritten by assuming that the latent variable (the $q(i)$ corresponding to the Poisson signal part) is known, and the latent variable is then replaced by its expectation value using the value of the parameters from previous iteration (or, initialization value if it's the first step). In the maximization step, the likelihood function is maximized with respect to the values of the parameters, giving the values for next iteration.

In our case the expectation step is done in the following way: we rewrite the likelihood function by assuming that the latent variable is known. So, the equation 3.9 can be simplified by removing

the sum over $q(i)$:

$$\begin{aligned} \ln L(f; C|Q(i) = q(i)) &= \ln \left(\prod_{i=0}^m \frac{e^{-f(i)t} (f(i)t)^{q(i)}}{q(i)!} \frac{e^{-\frac{(c(i)-q(i))^2}{2\sigma_R^2}}}{\sqrt{2\pi\sigma_R^2}} \right) \\ &= \sum_{i=0}^m \left(-f(i)t + q(i) \ln f(i)t - \ln q(i)! - \frac{(c(i)-q(i))^2}{2\sigma_R^2} - \ln \sqrt{2\pi\sigma_R^2} \right) \end{aligned} \quad (3.10)$$

Since the goal of the likelihood maximization is to find the flux, all terms that are not dependent on flux can be neglected. Maximizing the log-likelihood then simplifies to maximizing:

$$\sum_{i=0}^m (-f(i)t + E[q(i)] \ln(f(i)t)) \quad (3.11)$$

The log-likelihood function is now in a much simpler form which is very easy to optimize. In the case where there is no overlapping $f(i)$, direct differentiation of the expression w.r.t. each $f(i)$ gives $q(i)$ as the solution. When there are overlapping $f(i)$, the solution becomes a bit more complicated. It is seldom an obstacle as numerical optimization is very efficient.

The $q(i)$ is then replaced by the expectation value of $q(i)$, using the values of $c(i)$ and that of previous iteration (or initialization value) of $f(i)t$:

$$E[q(i)] = \sum_{q(i)=0}^{\infty} q(i) P(Q(i) = q(i) | C(i) = c(i), f(i)) \quad (3.12)$$

This is the expectation step. Now, by Bayes' theorem, the desired probability can be written as:

$$\begin{aligned} P(Q(i) = q(i) | C(i) = c(i)) &= \frac{P(Q(i) = q(i), C(i) = c(i))}{P(C(i) = c(i))} \\ &= \frac{P(Q(i) = q(i)) P(C(i) = c(i) | Q(i) = q(i))}{P(C(i) = c(i))} \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma_R^2}} e^{-\frac{(c(i)-q(i))^2}{2\sigma_R^2}} e^{-f(i)t} \frac{(f(i)t)^{q(i)}}{q(i)!}}{\sum_{q'(i)=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{-\frac{(c(i)-q'(i))^2}{2\sigma_R^2}} e^{-f(i)t} \frac{(f(i)t)^{q'(i)}}{q'(i)!}} \end{aligned} \quad (3.13)$$

With the numerical value of the expectation value of $q(i)$, the value of $f(i)$ is obtained by maximizing the simplified expression of the likelihood function. This is the maximization step.

One then iterates between these two steps until a satisfactory convergence is met.

For discussions involving different distribution of exposure time for different measurements, the same procedure can be applied by setting t to be a function of i , that is, replacing t by $t(i)$.

We now apply this EM algorithm in pixel intensity reconstruction.

3.6.3 Application of EM algorithm in pixel intensity reconstruction

The most common, default measurement scheme used taking images with a CCD is to take single pixel images for as long as the experiment time permits, to achieve the best available spatial resolution, such that the pixels reach the shot noise dominated regime and the signal-to-noise ratio is as high as possible. This is true when the photon flux is large compared to read noise. However, when there is a time constraint and the signal is weak, is this still the optimal measurement scheme?

By using hardware binning, the number of readouts can be decreased, in turn decreasing the readout noise introduced. The effect is particularly prominent when signal level is comparable to the read noise. Fig 3.11 shows the effect of hardware binning in decreasing the variance when the read noise is large.

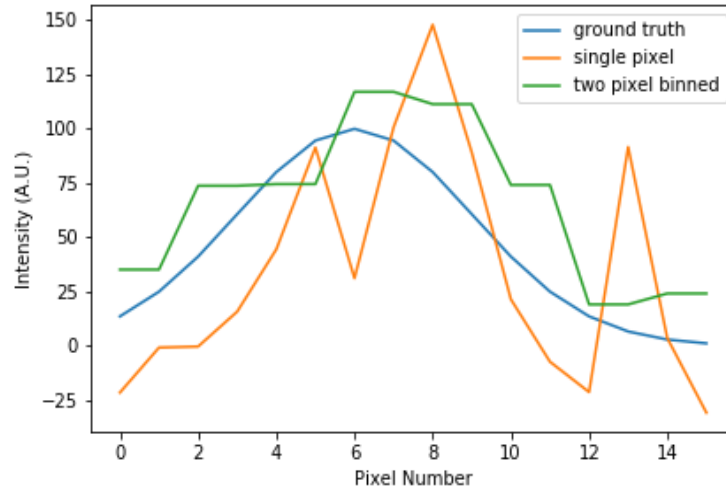


Figure 3.11: Simple illustration showing a Gaussian ground truth spectrum, plotted with single pixel and two pixel binned measurements. Both are corrupted first with Poisson noise and then readout noise of $SD = 50$, implemented with numpy [5].

However, the trade-off is a decreased spatial resolution. In the following sections, we investigate if the decrease in spatial resolution can be solved by using the EM algorithm on data combining

single pixel measurements with binned pixel measurements.

Note that in the traditional full frame CCD, having more than one measurement means the readout time also has to be considered (as measurement and charge shifting cannot be done in parallel), effectively having an idle time in which the photon flux is wasted. When the photon flux is low and there is a time constraint in the experiment, having a time where the scarce signal is not measured is unlikely to improve the signal-to-noise ratio.

Therefore, we restrict our analysis to frame transfer cameras, in which two consecutive full CCD measurements can be conducted before readout. As the shifting time is often orders of magnitude lower than the readout time, it is ignored in the subsequent analysis. We investigate whether combining a single pixel measurement and a binned measurement can give a more accurate prediction of the underlying single pixel ground truth values than just a single pixel measurement, under the same measurement time.

Verification of the EM algorithm

After developing a theoretical framework for applying the EM algorithm, it's time to apply it to practical use. A Python program is written to implement the EM algorithm, using Scipy's optimization package for the maximization step [67].

As a start, we validate whether the EM algorithm predicts the maximum likelihood estimator. In the two pixel case, the actual calculation of the likelihood for all parameter space is simple enough. So we can use this scenario to validate the output of the EM algorithm program.

One hundred simulations with different random seeds are conducted, resulting in different ground truths for each simulation. In each experiment, single pixel measurements and two-pixel binned measurements are taken. Poisson shot noise and Gaussian readout noise of SD of 30 are applied to the measurements. Convergence is set to be 0.01 count.

In all 100 experiments, all of the output falls within 5% of the maximum likelihood estimation within 50 steps. Fig 3.12 shows the results of two random experiments, with the left density plots

showing the values of the likelihood function as a function of the first and second pixel intensity, and the right plots showing the convergence of prediction towards the maximum likelihood estimators.

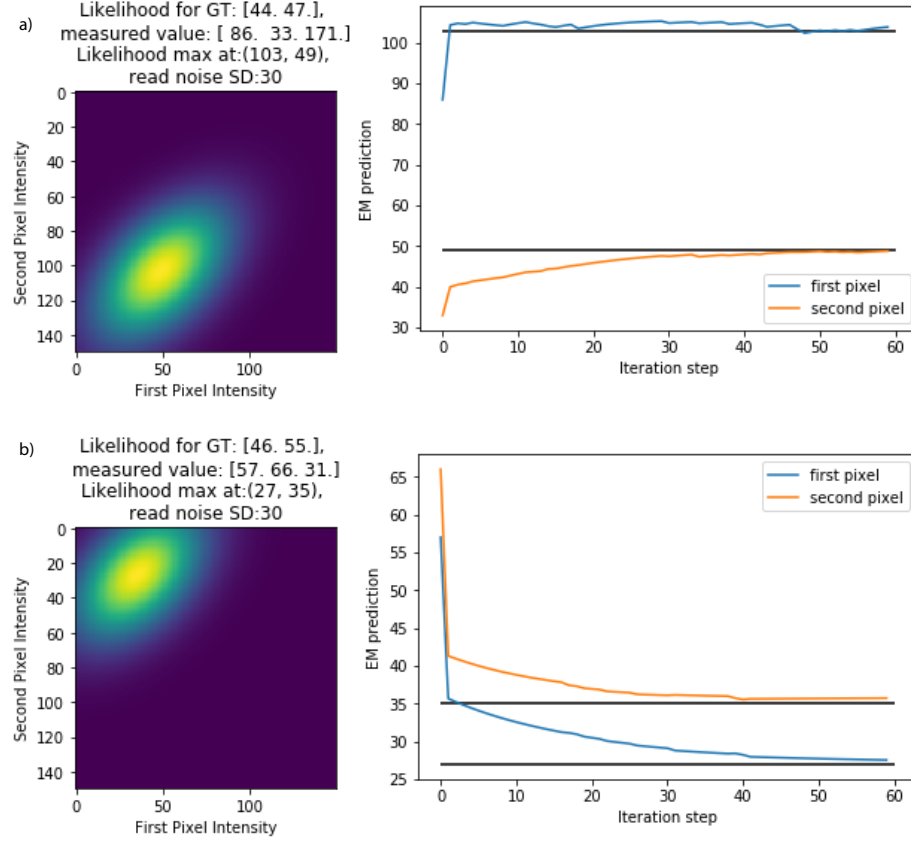


Figure 3.12: Verification that the EM algorithm converges to the maximum likelihood estimation in the case of two pixel: Left figures show the explicit calculation of the likelihood function for the measurement. Right figures show the predicted ground truth value for each pixel as a function of iteration steps of the EM algorithm. Black line indicates the maximum likelihood estimation. a) and b) show the results for two random experiments.

This validates that the implemented EM algorithm does arrive at the maximum likelihood estimation values after the convergence is met. In the next section, we apply the EM algorithm on data consisting of single pixel and binned measurements.

Single Pixel Reconstruction from Duo-Measurement

Under the same measurement time, the predictions based on the EM algorithm on single and binned pixel, and the predictions based on single pixel measurement only, are compared. As it is expected that spatial information is lost quickly as the bins get larger, we focus on binned measurements from 2-pixel, 4-pixel, and 8-pixel. And as discussed previously, as maximally two frames can be stored in a frame transfer CCD before readout, we restrict our analysis to measurements involving two frames.

Three different scenarios are investigated:

1. Single pixel measurement of time t_0 s and 2-pixel binned measurement of time $(1 - t_0)$ s.
2. Single pixel measurement of time t_0 s and 4-pixel binned measurement of time $(1 - t_0)$ s.
3. Single pixel measurement of time t_0 s and 8-pixel binned measurement of time $(1 - t_0)$ s.

Note that the choice of total measurement time being 1 s is just for convenience. The analysis could be applied in other values of measurement time by rescaling the photon flux accordingly.

In each scenario, the output from the EM algorithm on these combined measurement is compared to spending all time (1 s) on single pixel measurement, by calculating the mean squared error of prediction w.r.t. the underlying ground truth.

The ground truths are set to be random integers from 0 to 100 and the maximum intensity normalized to 100 subsequently. This is to ensure the intensity across the whole chip is of similar magnitude, so that the comparison between the read noise SD can be made.

As it is unclear that how much time should be spent on the respective measurements (t_0 and $1 - t_0$), it is left as a variable together with the standard deviation of the read noise in the simulation.

5000 different simulations with different random seeds are conducted for each pair of exposure time distribution and read noise SD. The left graphs in Fig. 3.13, Fig. 3.14, and Fig. 3.15 show respectively the mean squared error of prediction by the EM algorithm applied to the combined measurement data with respect to the underlying ground truth for the case of 2 pixels, 4 pixels,

and 8 pixels. In the figures the right graph shows the mean squared error from the EM algorithm subtracted by the mean squared error of the single pixel measurement. That is why the last row has values of zeros, because that is the row where all the time is spent on single pixel measurement, and the original value is the maximum likelihood estimation when there is only one measurement. So any negative values in the right graph signify that that particular combination of read noise and exposure time distribution favors the combined measurement over spending all time on single pixel measurement in terms of accuracy to the ground truth value.

In the case of two pixels, it can be observed that spending all time on single pixel measurement is better, as the values in the right graph are never negative and all columns have their minimum at the bottom.

In the case of four pixels, the scenario is similar, but at high read noise (≈ 110), a local minimum appears before a local maximum as the exposure time on single pixel measurement is increased.

In the case of eight pixels, the observation of the local minimum is more prominent. And the mean square error difference became negative at $t_0 = 0.4s$ and read noise $SD = 110$. The difference is 7.6, about 7% of the read noise SD . Fig. 3.16 shows two example spectra at that particular combination, with both the EM and SP predictions being very noisy, but the EM prediction being slightly more accurate.

In view of the eight pixel result, we also attempted applying the EM algorithm on 16-pixel case (combining 16 pixel binned measurement with single pixel) and it was compared to spending all time on single pixel measurement. However, no improvement in accuracy was observed. This can be explained by 16 pixel bins losing too much spatial information to be useful to single pixel reconstruction.

From these results, it can be concluded that only under very high read noise (comparable to the maximum intensity in the signal) and specific exposure time distribution, this combination of single pixel and binned pixel measurement would defeat spending all time on single pixel measurement.

This can be loosely explained by the expectation that hardware binning sacrifices spatial resolution to favor higher signal-to-noise ratio. However, single pixel intensity reconstruction effectively corresponds to single pixel resolution when there is no spatial correlation between the pixels.

To further understand whether that is the case, we resort to the Cramér-Rao lower bound to analyze the case for two pixels.

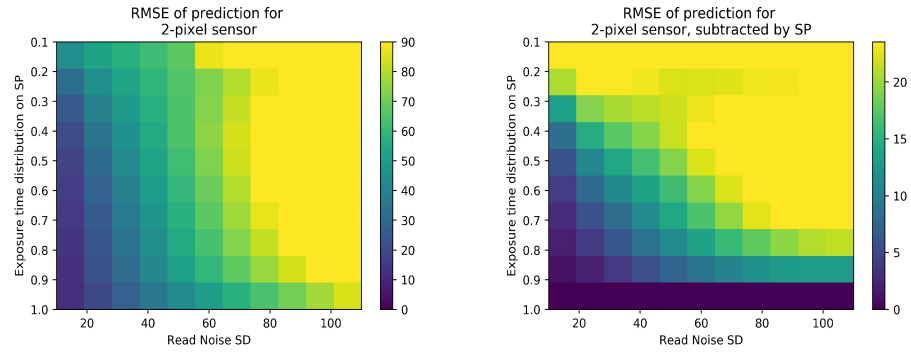


Figure 3.13: Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 2 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 2 pixels.

Cramér-Rao Lower Bound

In this section, we inspect the conclusion made in the previous section by resorting to estimation theory. Denoting the ground truth fluxes for first and second pixel by f_1 and f_2 , the associated Fisher information matrix is:

$$\mathcal{I}(f_1, f_2) = \begin{pmatrix} \left\langle \frac{\partial \log(L)}{\partial f_1} \frac{\partial \log(L)}{\partial f_1} \right\rangle & \left\langle \frac{\partial \log(L)}{\partial f_1} \frac{\partial \log(L)}{\partial f_2} \right\rangle \\ \left\langle \frac{\partial \log(L)}{\partial f_2} \frac{\partial \log(L)}{\partial f_1} \right\rangle & \left\langle \frac{\partial \log(L)}{\partial f_2} \frac{\partial \log(L)}{\partial f_2} \right\rangle \end{pmatrix} \quad (3.14)$$

where $\log(L)$ is the log-likelihood function in eq. 3.9, and the expectation value is taken with respect to the probability density function corresponding to the all possible measured values (from $-\infty$ to ∞ .) In our particular case, since the measurements are independent of each other, the

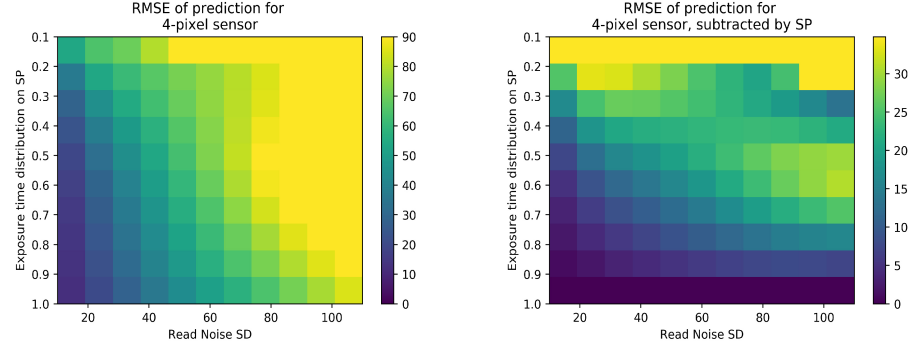


Figure 3.14: Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 4 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 4 pixels.

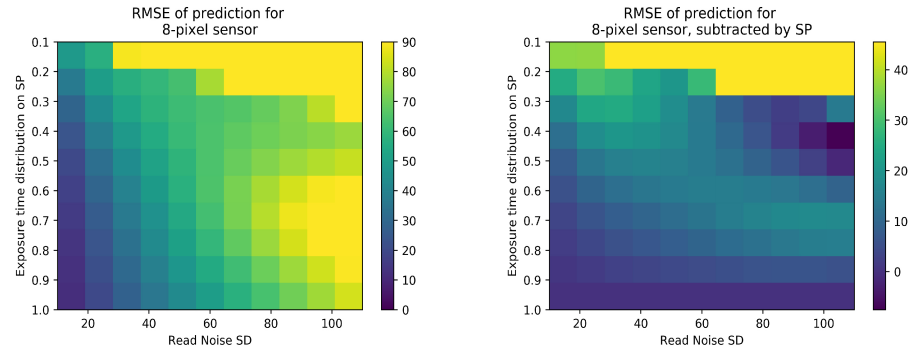


Figure 3.15: Left: RMSE of prediction by the EM algorithm using the combined measurement with respect to the underlying ground truths for 8 pixels. Right: Mean squared error of prediction for the EM algorithm using the combined measurement with respect to the underlying ground truth subtracted by the RMSE of the prediction by spending all time on single pixel measurement for 8 pixels.

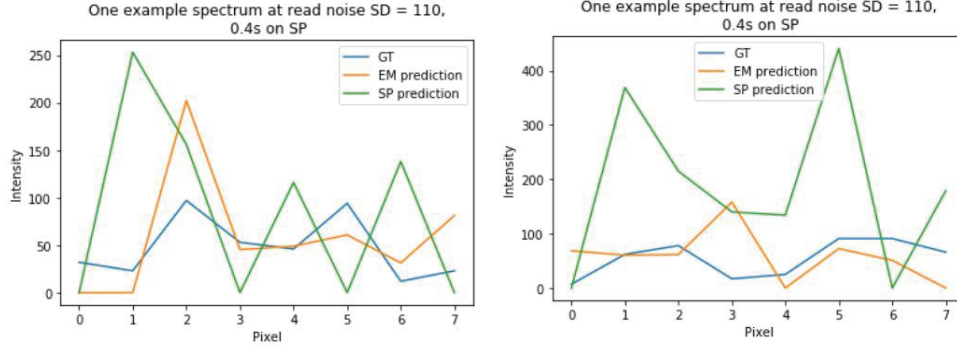


Figure 3.16: Example spectra at the particular read noise and exposure time distribution that combined measurement defeats single pixel measurement.

Fisher information matrix can be computed as a sum of three matrices, each of which denoting the information obtained from each measurement.

The Cramér-Rao lower bound for the covariance matrix of any unbiased estimator T for the fluxes f_1, f_2 is given by:

$$\text{cov}_{f_1, f_2}(T) \geq \mathcal{I}(f_1, f_2)^{-1}. \quad (3.15)$$

This lower bound is numerically calculated as a function of f_1, f_2 , readout noise σ , and t (time spent on single pixel measurement, again, total measurement time is chosen to be 1s for convenience), as shown in Fig 3.17. As observed in the figures, the minimum lower bound is always to the right-most end of the graph for a fixed σ , meaning that most information about the underlying ground truth can be obtained when all time is spent on single pixel measurements. So adding another binned measurement actually makes the measurement less informative, if the total measurement time is fixed.

To validate the calculation, the Cramér-Rao lower bound calculated for a specific f_1, f_2 and σ is compared to the variance in the single pixel case. Several comparisons were made at different combination of values, and they all converge to the variance in the single pixel case as all the time is spent on single pixel measurement.

This analysis is just for two pixels, but it sheds light on the fact that taking away time from the

single pixel measurement to conduct binned pixel measurement is not desirable.

In the subsequent section, we therefore direct our attention to the case where there is significant spatial correlation between neighbouring pixels, that is, localization of spectrum.

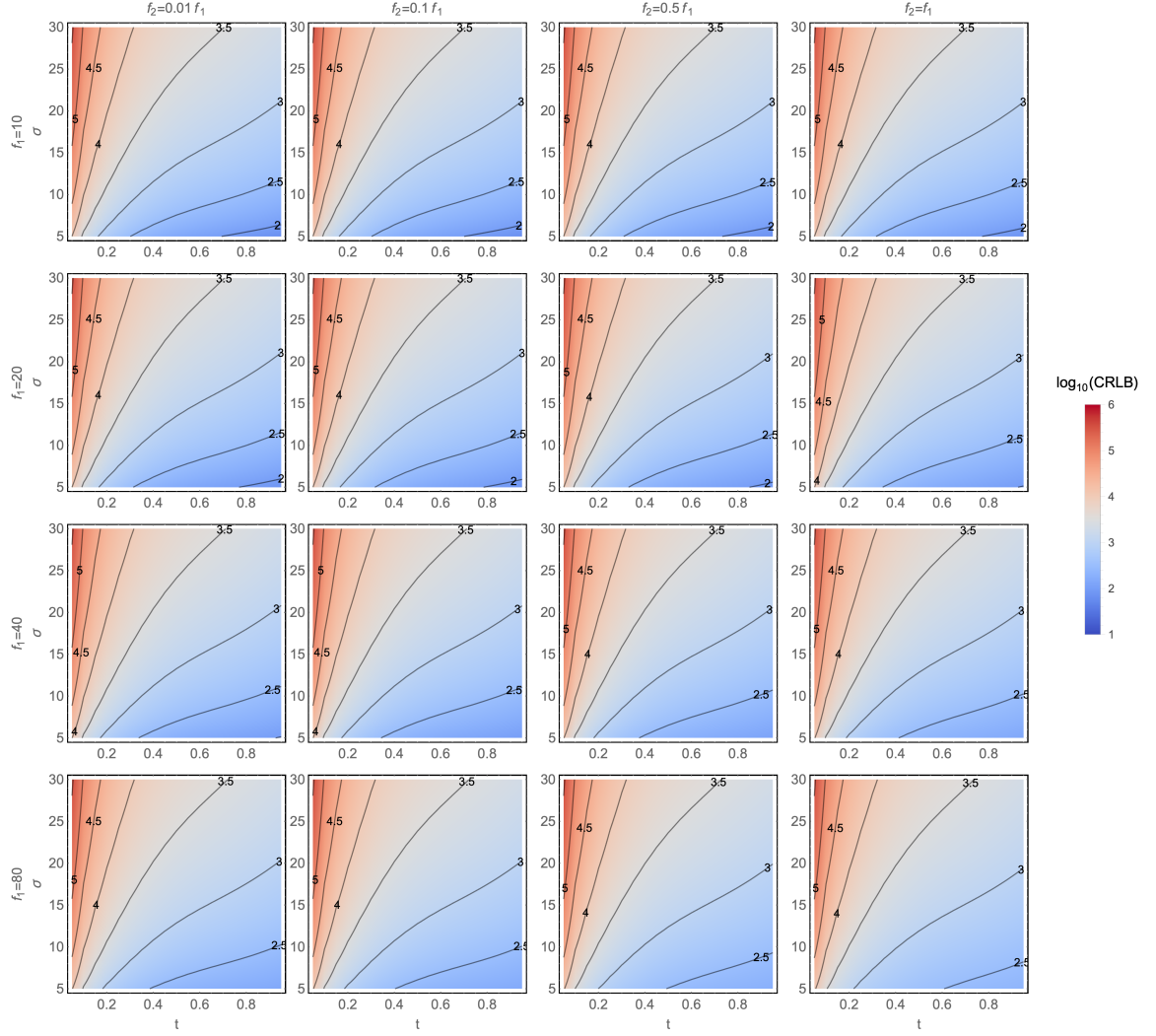


Figure 3.17: Cramér-Rao lower bound computed over a range of single-pixel exposure times t , fluxes f_1 and f_2 , and readout noise σ . The values for f_1 are constant along the rows, and the values for f_2 are the fractions of f_1 shown on the top.

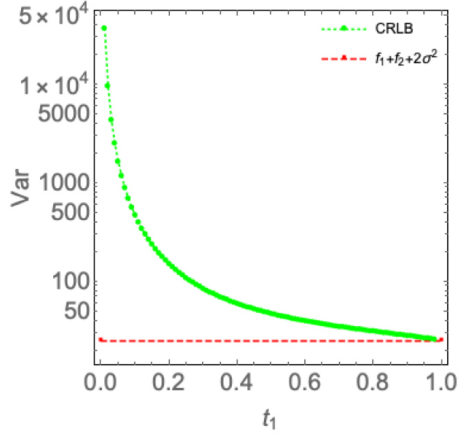


Figure 3.18: Comparison between the Cramér-Rao lower bound computed and the expected variance when all time is spent on single pixel measurement. In this particular example, $f_1 = 10$, $f_2 = 7$, and $\sigma = 2$

3.7 Spectrum Localization

Spectral lines seldom have vanishing line widths. There are different mechanisms that broaden spectral lines into Gaussians and Lorentzians, including lifetime broadening, thermal fluctuations in crystal host, Doppler broadening, pressure broadening, etc. [68]. The combination of them can result in complicated line shapes like the Voigt profile.

We turn our investigation to the application of hardware binning to localization of the maxima of spectral features. With higher spatial correlation between neighbouring pixels, the loss of resolution due to hardware binning should be less impactful.

3.7.1 Weighted Least-Squares Regression of a Gaussian Spectrum

We first consider the simplest case: having one Gaussian spectrum on the sensor. The ground truth intensity of the pixels can then be written as a function of the intensity A , peak position μ and spread σ :

$$F(x; \sigma, \mu, A) = Ae^{-(x-\mu)^2/(2\sigma^2)} \quad (3.16)$$

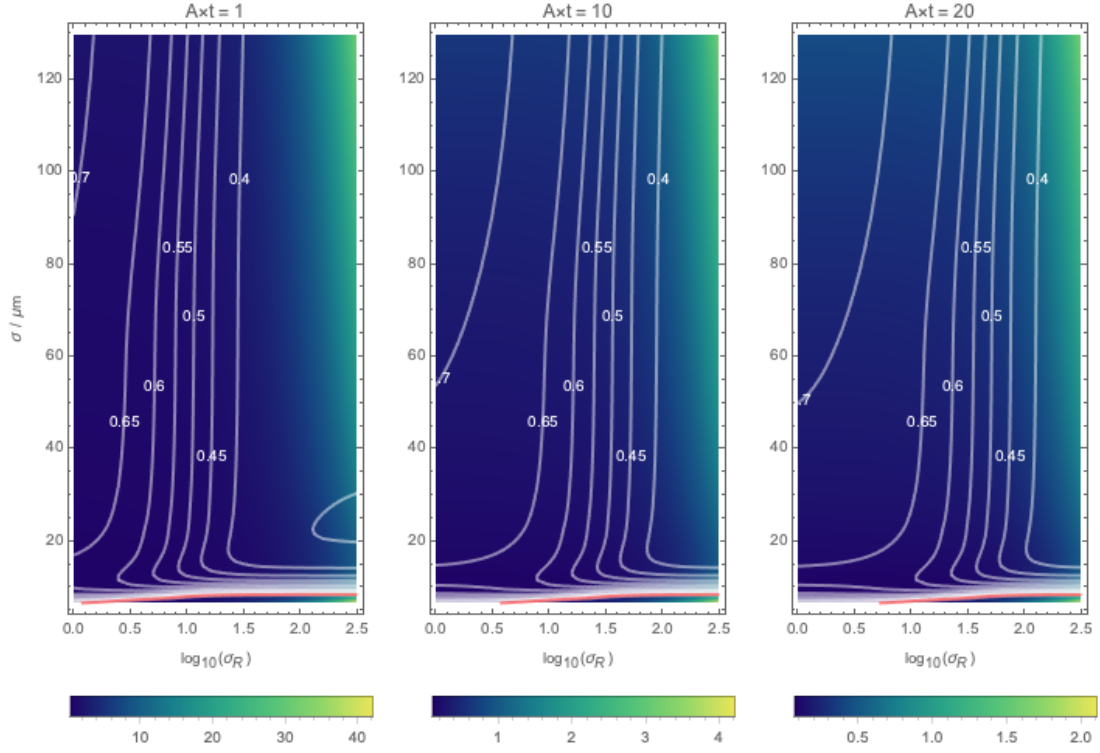


Figure 3.19: Density plot: Uncertainty in peak localization in binned pixel measurement. Contour plot: Ratio of the uncertainty in peak localization in binned pixel measurement to that in single pixel measurement. Number lower than 1 means binned measurement gives lower uncertainty. Red line: Ratio of uncertainty = 1.

The equation can be converted into a quadratic sum in x by taking logarithms:

$$\ln F(x; \sigma, \mu, A) = (\ln A - \frac{\mu^2}{2\sigma^2}) + \frac{\mu}{\sigma^2}x - \frac{x^2}{2\sigma^2} \quad (3.17)$$

This can then be considered a linear regression problem for the parameters $(\ln A - \frac{\mu^2}{2\sigma^2})$, $\frac{\mu}{\sigma^2}$ and $\frac{1}{2\sigma^2}$. The uncertainties in peak localization (the square root of the variance) can be obtained by propagating the uncertainties in pixel values. The noise in the pixels is assumed to be that of a Poisson-Gaussian variable, with parameters σ_R , and λ corresponding to the photon flux which varies across the sensor as the Gaussian profile. The scenarios for single pixel and 2-pixel binned measurements are considered. Fig. 3.19 shows as contour lines the ratio between the uncertainties for the binned and single-pixel scenarios for different overall amplitude of the Gaussian profile $A \times t$,

width of the Gaussian spectrum σ , and readout noise σ_R . The sensor consists of 32 pixels, each of them being $13\mu\text{m}$ in size. The same figure has a density plot that shows in color the uncertainty in peak localization (the uncertainty in the parameter μ of the Gaussian profile) of the binned measurement case. The red line shows the location where the ratio becomes 1, marking the boundary where binning would no longer yield a lower uncertainty than the sensor without any hardware-binning.

At a constant width of the peak, it can be observed that as the readout noise increases, the binned measurement is more and more advantageous than the single pixel measurement. And for a constant readout noise, as the width of the peak shrinks below the pixel size ($13\mu\text{m}$), single pixel measurement quickly dominates, due to the lack of single pixel resolution in binned measurement.

However, in general, the binned pixel measurement is superior to single pixel measurement. This is to be expected to be due to the high spatial correlation between neighbouring pixels.

3.7.2 Numerical Fitting of Mixture of two Gaussian Peaks

We now turn to a slightly more complicated case where two Gaussians profiles are superposed on the linear sensor. In this case, the log linearization cannot be conducted anymore. We therefore resort to more general regression algorithms to look at the effect of binning.

To look at a case more similar to actual scenarios in spectroscopy where the spectral features to be identified are more than one and not known beforehand, the two Gaussian peaks are randomly placed on the sensor, with their peaks being two random numbers between 0 and 31, and spread of the Gaussian to be a random number from 2 (half of the largest bin) to 16 (half of the sensor size). The relative intensity is also a random number, with the maximum intensity of the spectrum normalized to 200.

Three modes of measurement are compared: single pixel measurement, 2-pixel binned measurement, and 4-pixel binned measurement. All of them spend the same amount of measurement time. Two Gaussians are then fitted on top of the measurement with non-linear least squares using Scipy's

optimization module [67] to obtain the peak locations of the two Gaussians.

Fig 3.20a shows the root-mean-squared error in peak location prediction for this simulation. It can be observed that quadruple pixel bins give the least error in peak localization, but the advantage decreases as the read noise SD decreases, followed by double pixel bins, and finally single pixel measurement.

Another simulation which inspects thinner peaks, with random σ being from 1 to 4 (size of largest bin). The results are shown in fig 3.20b. The advantage of higher pixel bins is less, but the order still stays the same. Also, the three graphs converge at higher read noise, as a higher spatial resolution from single pixel measurement benefit more when the peaks are thinner.

The results of these simulations agree with the expectation that hardware binning enhances signal-to-noise ratio when there is significant spatial correlation between neighbouring pixels. However, these hinge on a fact that the exact shape and number of peaks are known beforehand. In practical spectroscopy where there are unknown number of peaks and they are of unknown shape, more sophisticated algorithms would be necessary.

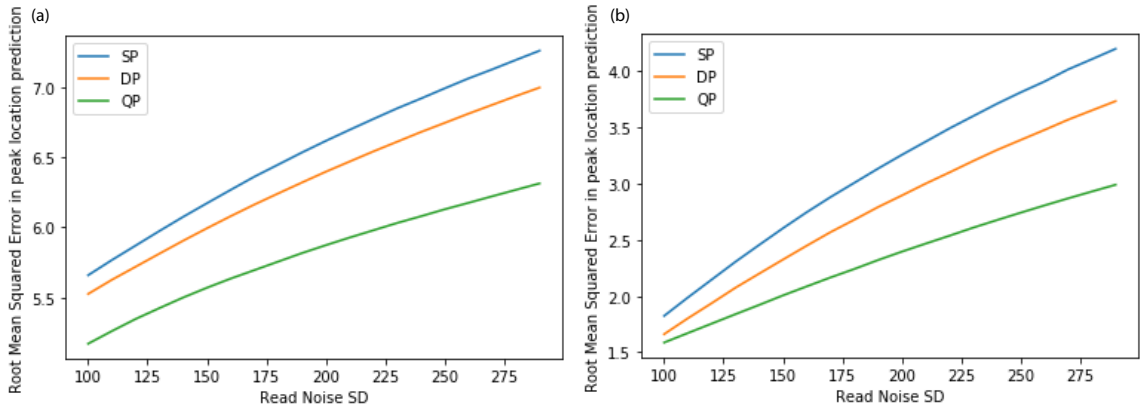


Figure 3.20: RMSE for peak localization for single pixel, double pixel, and quadruple pixel binned measurements. Left: Peak width being a random number from 2 to 16. Right: Peak width being a random number from 1 to 4.

3.8 Conclusion

In this chapter, we reviewed the essentials of CCDs, including the different types available, and most importantly their noise sources and ability to perform hardware-binning. We then showed how characterization of a CCD camera can be conducted as an illustrative example. Then, with a noise model composed of Poisson shot noise and Gaussian readout noise, we applied the EM algorithm on single pixel intensity reconstruction. It was shown that generally combining single pixel and binned pixel measurements is not beneficial, evidenced by the reconstruction results based on two pixel and four pixel cases. Subsequent analysis with the Cramér-Rao lower bound also hints at information content being lost when a given measurement time is shared between two measurements instead of just a single pixel measurement. Only under a very specific regime where the readout noise is comparable to the signal and the exposure time distribution is optimal, the EM algorithm applied to the combination of a hardware binned measurement and single pixel measurement can be slightly more accurate than a single pixel measurement in rather restricted circumstances. We then switched to look at cases where there is significant spatial correlation between neighbouring pixels which is often the case in spectroscopy. We investigated the cases for one and two Gaussian peaks, and in both cases the binned measurement was superior. This is because the sacrifice in spatial resolution is no longer a problem when there is strong spatial correlation, while the enhancement in signal-to-noise ratio improves the localization accuracy.

3.9 Acknowledgement

This work was done in collaboration between the thesis author and Juan David Lizarazo Ferro, who contributed to the analysis about Cramér-Rao lower bound and the weighted least-squares regressions.

Chapter 4

Learning from Our Mistakes - Expanding Reviews of Graduate Admission Decisions

4.1 Introduction

Here, we apply these methods to a broader scientific challenge. Graduate school admission is a process with large variances and further exacerbated by limited data. Given the long duration for graduate study, feedback information acquisition is very slow. In this work, we report a method and develop software tools for obtaining and analyzing academic outcome data for graduate students, resulting in generally an order of magnitude more data for strengthening the feedback mechanism and enabling a more thorough review of admission practices.

4.2 Background

With the recent heated debates, such as, the Harvard's admission controversy, the long running argument on whether GRE scores are a good predictor of success [69–90], and efforts to make GRE optional in various departments at Brown, the continuous improvement of admission processes becomes even more important.

Experiment is all about setting hypotheses, changing independent variables, and observing the subsequent changes in the dependent variables to test the hypothesis. Can the whole admission process be analyzed like an experiment, and can we bring about insights to what would be ideally the best admission strategies based on statistics and data? If we want to statistically analyze academic admission, a feedback mechanism is required. Is there an effective feedback mechanism available to schools to examine whether they have made good admission decisions?

Traditionally, universities evaluate their graduate school admissions processes based on examining only matriculated students - the students who are most readily accessible to them. This limited viewpoint can make visible the challenges of admitted students who may struggle at a given institution, but has two main problems:

Firstly, it misses the exceptional performance of students who were not admitted or who chose not to matriculate. From the data pool of admitted students, there is no information about students that are not matriculated. This limits the number of students for analysis and leaves out important information that can be revealed when looking at those students.

Secondly, due to selection bias, limiting reviews to students who have already been admitted can also compound any existing departmental or institutional preferences, leading to overconfidence in admission decisions. A typical example is that, professors in some universities find that most of their colleagues pursued PhDs in topic X or school Y. They may as a result conclude that similar individuals are more capable and suitable for being professors. However, this may also be due to the fact that there has been existing bias in past recruiting culture. If that is indeed the case,

such departmental or institutional preference / bias would only continue this situation and create a vicious cycle.

It is therefore vital that institutions seek access to a more complete and unbiased feedback mechanism. In the following, we report not only a programmatic way for institutions to obtain an order of magnitude more students data by including both matriculated and non-matriculated students, but we also provide toolkits that would allow the schools to acquire a spectrum of post-matriculation metrics for those students.

4.3 Method

4.3.1 Data Source

Rather than focusing solely on matriculated students, we provide tools to examine the subsequent enrollment and achievements of *all* applicants to our doctoral programs as a measure of their success. This approach has two significant advantages:

First and foremost, it can provide meaningful feedback on our admission criteria. In order to improve practices, admissions committees must be able to recognize mistakes. However, if we restrict admissions analysis to matriculated students, the only mistake that we can observe is the poor performance of a student who matriculates. In the traditional approach of examining students at our own institution, we cannot observe the stellar performance of students whom we did not admit. This approach cannot answer questions like, did we let go of students that are potential stars? If so, why did we make those decisions? Did we admit students because of a certain preference, at the cost of losing more capable students? These questions can all be examined in part through statistical analysis with the proposed set of data.

Second, it allows us to examine a larger number of students across a range of institutions. The size of the applicant pool is typically more than an order of magnitude larger than the pool of matriculated students. At selective institutions where the selectivity (admissions rate) is $\sim 10\%$ and

yield (matriculation rate) is $\sim 50\%$, our approach produces a sample size that is approximate 20 times larger than the conventional approach. As statistical inferences are inherently probabilistic, a larger sample size enables more confidence in conclusions.

Student enrollment and graduation data can be obtained through the StudentTracker service provided by the National Student Clearinghouse. More than 3,600 colleges and universities, covering 98% of all students in these schools, report enrollment and graduation data to the Clearinghouse [91]. By querying graduate applicants to Brown, we are able to obtain data of more than 10,000 graduate applicants, including in which graduate programs and institutions they matriculated in and when they graduated. This data then allows us to obtain more publicly available data like publications, awards, and job placement for each individual student.

4.3.2 GradART

To acquire outcome data on awards and publications, we developed a Python package called Graduate Admission Reflection Toolkit (GradART), whose four components are illustrated in Fig 4.1. In addition to postprocessing the data format obtained from the NSC StudentTracker service, the toolkit consists of wScholar, Awards Tracker, and Thesis Finder. The wScholar package retrieves publications and analyses them. Award Tracker queries NIH and NSF databases for awards, and Thesis Finder interfaces with Proquest to provide titles and DOIs / URLs for the PhD theses of a given group of doctoral students.

In some disciplines, for instance in the life and physical sciences as well as in engineering, publication records are used to measure the success of graduate students. However, not all scholars have unique researcher identifiers such as Open Researcher and Contributor IDs (ORCID), and obtaining accurate and complete publication records can therefore be difficult. Here, we present a search engine that takes NSC data as input, extracts the publication record of each student in the NSC dataset from the Web of Science (WOS) using their API, and applies custom algorithmic filters, consisting of co-author and journal filters, to improve the accuracy of the search results.

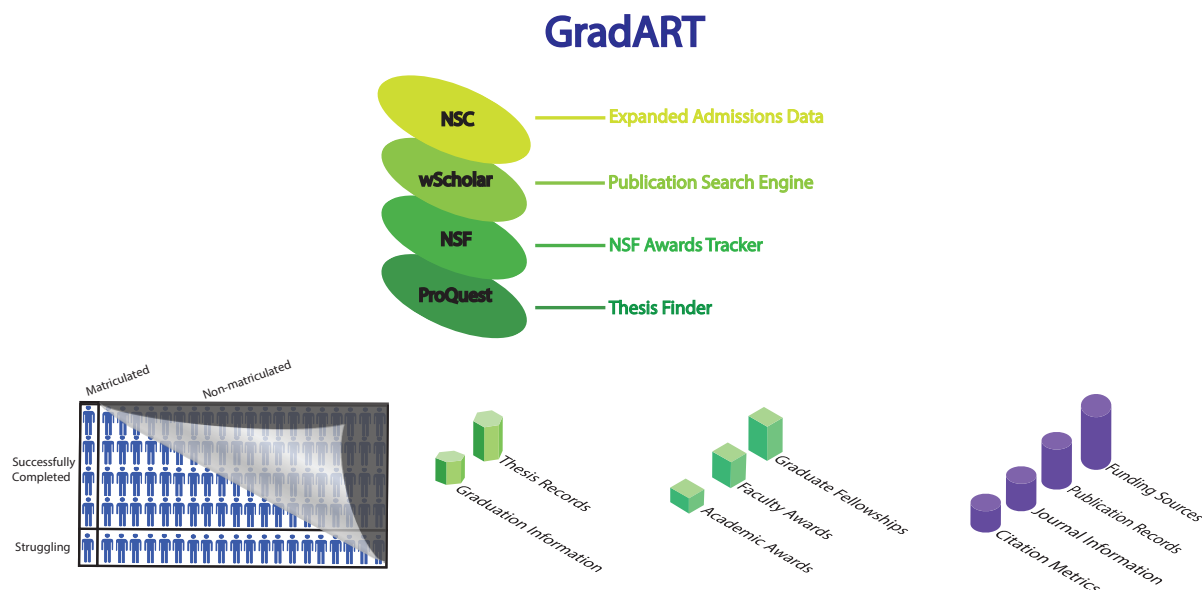


Figure 4.1: The four main components of GradART: a) NSC: Expands admission data by adding information on non-matriculated students (as discussed in the text, the ratio of applications to matriculated students is approximately twenty; the ratio of students who struggled vs completed successfully is based on [6,7]. b) wScholar: Allows programmatic search of publication records for all students. c) Awards Tracker: Retrieves fellowship and grant records from NIH and NSF. d) Thesis Finder: Allows quick access to the PhD theses of graduated students for additional analysis.

To accurately attribute publications to a given individual, wScholar uses two different queries that are then matched. The first query is an exact-name search that retrieves only those publications whose author data match the full name and the full affiliation of the given individual. The second partial-name query retrieves publications using an abbreviated form of the name and the most uncommon word in the affiliation. The exact-name query is often too restrictive, while the partial-name query may include wrongly attributed publications. For instance, carrying out an exact-name record using “Rashid Zia” as author with affiliation “Stanford University” returns three papers; using “R Zia” or “R. Zia” as the name and “Stanford” as the affiliation name returns the complete list of nine papers published by this author during his graduate studies. In other cases, the partial-name query may produce false results as many authors may match a given partial name and affiliation.

We address this ambiguity with two filters.

The coauthor filter uses results of the exact-name search (full name and full affiliation) as the ground truth and constructs a coauthor list for each given individual. It then filters the partial-name search results by retaining only those records with at least one known coauthor. The assumption behind this filter is that students commonly publish consistently with their advisor.

Since not all people can be uniquely resolved with the exact-name search, we also filter for journals. We use the student's discipline as given by the NSC data to remove all publications from the coauthor-filtered partial-name search results that were published in journals outside the student's discipline.

Finally, wScholar can also retrieve the citations for each paper as a partial measure for the scholarly influence and impact of a student's research throughout their PhD studies. This information is supplemented by impact factors of the respective journals.

4.3.3 Additional data acquisition tools

Thesis Finder: Publication records are not always useful. For example, in the humanities and social sciences, the PhD thesis is often the major outcome of graduate work. To retrieve information about PhD theses, we provide Thesis Finder, which is a tool to find thesis titles and DOIs/URLs for each student in a given NSC dataset using ProQuest. ProQuest is a large curated collection of dissertations and theses, and it provides access to students scholarly work by matching their names, schools, and years of study. The output from Thesis Finder allows users to locate thesis information and allows subsequent purchase of the actual thesis through ProQuest. These theses can then be further analyzed to extract department specific insights or metrics of success as feedback on admission strategies.

Award Tracker: Fellowships and grants are also important indicators of academic success. Award Tracker retrieves information about NIH and NSF awardees and matches it with output from NSC

data. While grants are more suitable for measuring the success of researchers after graduation, fellowship information is often very relevant for PhD students and postdoctoral fellows.

In conclusion, GradART enables the automated compilation of a dataset that contains both admission information and multiple proxies of success for applicants over the duration of their academic careers, thus allowing for a comprehensive reflection on the admission strategies used by a given graduate program.

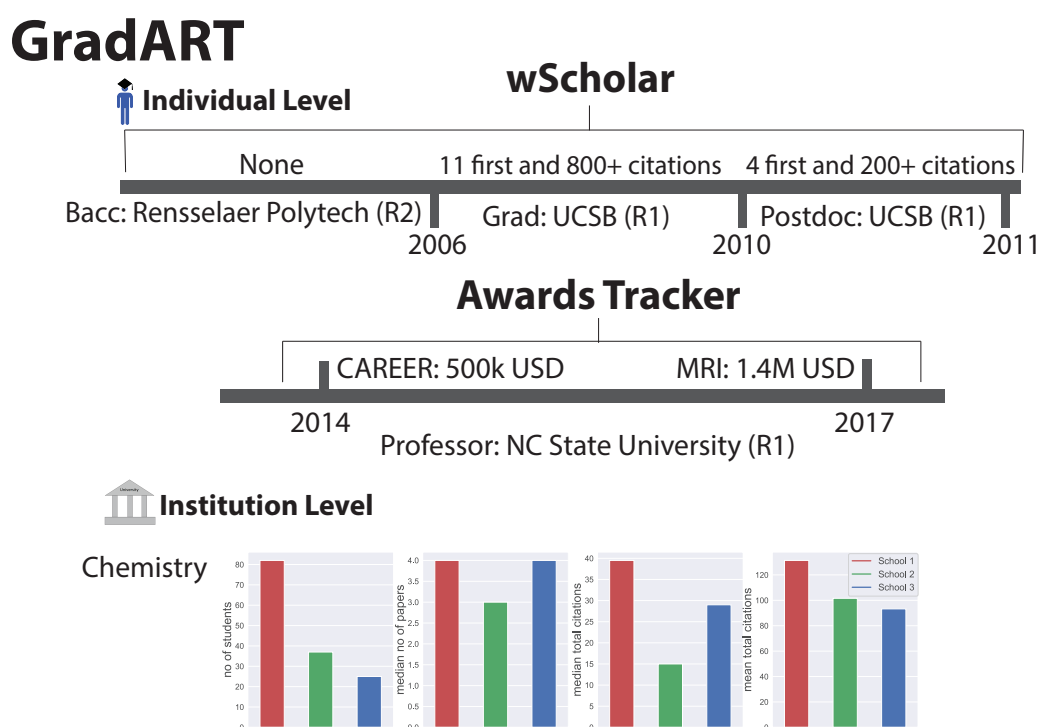


Figure 4.2: Showing the actual results from wScholar and Awards Tracker: illustrating the two extreme levels. The top part shows the academic achievements of an anonymous individual, while the bottom part shows aggregate statistics for chemistry students from three anonymous schools. Both individual level and institution level views can be generated, allowing small and large scale decisions to be scrutinized.

4.4 Sample application

To illustrate the use of this package, we used a publicly available list of students who were awarded NSF Graduate Research Fellowships (GRFP) during 2005-2012 [92]: this resulted in over 25,000 students who received GRFP awards or an Honorable Mention. We reformatted these data to match the format of data provided by the NSC to its member institutions and then used the resulting dataset as input to wScholar and Award Tracker.

To validate the output of wScholar, we carried out manual checks on 50 randomly chosen NSF Graduate Research Fellowship awardees. We found that wScholar achieved an accuracy of 82%. The errors are due to various reasons including overestimating publication counts because of other students with the exact name within the same institution, students published under a different name as registered by the NSC or under an organization other than their Ph.D. program hosts. With wScholar, the performance of students can be measured by four metrics: the number of papers published, the number of first authored papers, the sum of impact factor of journals that students publish first authored papers in, and the total citations of all first authored papers of a student.

Table 5.1 shows the performance statistics obtained by running wScholar on the GRFP data set. One can observe that the number of papers found for social sciences, computer science and mathematical sciences are fewer compared to other fields. For social sciences, it is often the norm that the quality of the thesis is the most important measure of graduate success, and students generally publish less often in journals. For computer science, due to the fast pace of research, results are generally published in conferences proceedings. As the quality of conference proceedings can vary drastically due to the less formal peer review process and the fact that web of science does not support conference proceeding search as well, conference proceedings are not counted as papers in wScholar, resulting in the lower number of published papers as shown in the table for computer science students. For mathematics it may be due to two reasons: Firstly, the publication cycle for

Table 5.1: Table showing summary statistics of the performances of students from the wScholar results ran with the GRFP data set

Major	No. of papers		No. of first authored papers		First authored papers IF sum		cites		norm cites		No. of students	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Life Sciences	4.27	3.54	1.69	1.67	8.92	11.43	63.81	287.81	9.74	28.76	5444	
Engineering	4.52	4.01	1.72	1.81	8.24	12.30	62.74	162.52	9.28	20.34	4572	
Chemistry	4.93	3.99	1.93	1.95	14.89	18.53	97.33	256.45	14.76	31.77	1446	
Psychology	4.65	3.88	2.08	2.10	7.51	8.94	62.51	115.67	9.47	15.28	1443	
Social Sciences	2.69	3.16	0.96	1.32	2.56	5.15	24.19	73.33	3.52	9.63	1217	
Physics and Astronomy	4.91	4.41	1.58	2.04	9.46	14.36	60.24	132.95	9.28	19.39	1067	
Comp/IS/Eng	2.68	2.89	0.85	1.13	2.98	5.59	23.13	94.58	2.95	9.60	801	
Geosciences	3.82	3.42	1.62	1.68	6.77	8.55	44.18	120.69	7.09	16.00	776	
Mathematical Sciences	2.74	3.07	1.02	1.32	2.28	4.57	21.06	116.36	3.11	15.63	456	
Materials Research	4.69	4.24	1.48	1.60	13.30	18.19	55.58	161.24	13.49	33.97	114	

mathematics students is usually much longer and require a lengthy peer review process. Secondly, mathematics students sometimes publish papers without their advisors, rendering the coauthor filter inapplicable. In addition to these three majors, there are also variations in number of publications in each field. It is therefore less convincing to compare students from different fields directly through the number of publications or citations. In the subsequent analysis, therefore, we mostly restrict to comparison within the same major.

4.4.1 Comparison between honorable mention and awardees

This data set allows us to compare the performance of students who got honorable mentions in the GRFP with actual awardees. As a case study, we look at life sciences, the broad field with the highest number of students. Normalized histograms are plotted and statistical tests including Mann-Whitney U test and t-test are used to assess whether the populations are statistically different.

From the figures 4.3, 4.4, 4.5 and 4.6, it can be observed that there is no significant difference in distributions between the two groups. Both tests also reveal that the two populations are not significantly different, as evidenced by the high p-values.

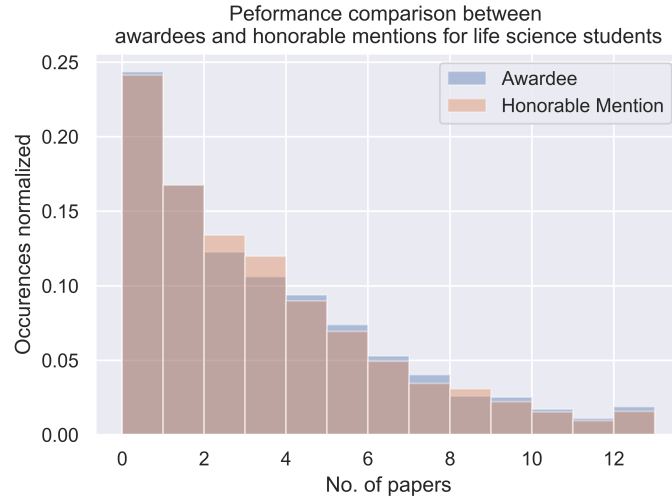


Figure 4.3: Showing the normalized histogram for number of papers for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.13 and 0.09 respectively. The mean values are 3.38 for awardees and 3.23 for honorable mention students.

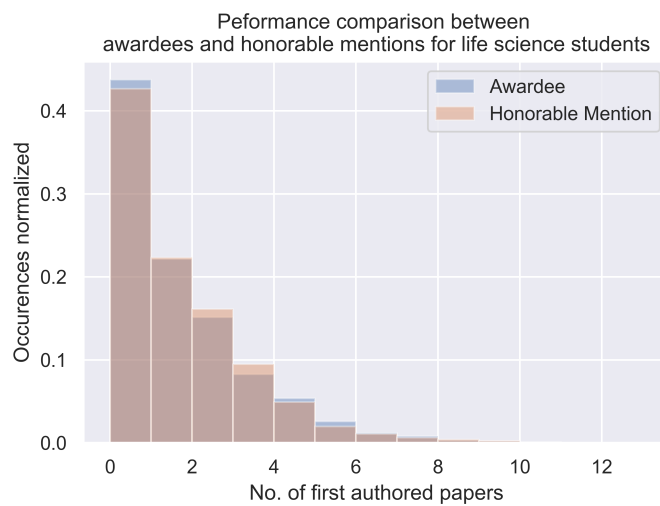


Figure 4.4: Showing the normalized histogram for number of first authored papers for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.28 and 0.86 respectively.

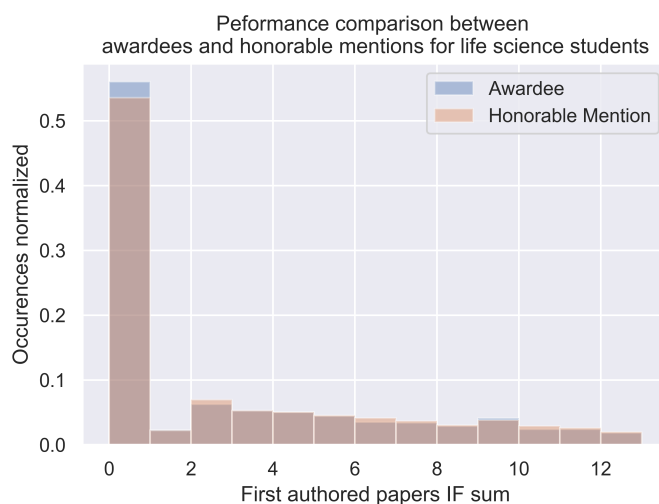


Figure 4.5: Showing the normalized histogram for impact factors for first authored paper summed for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.37 and 0.18 respectively.

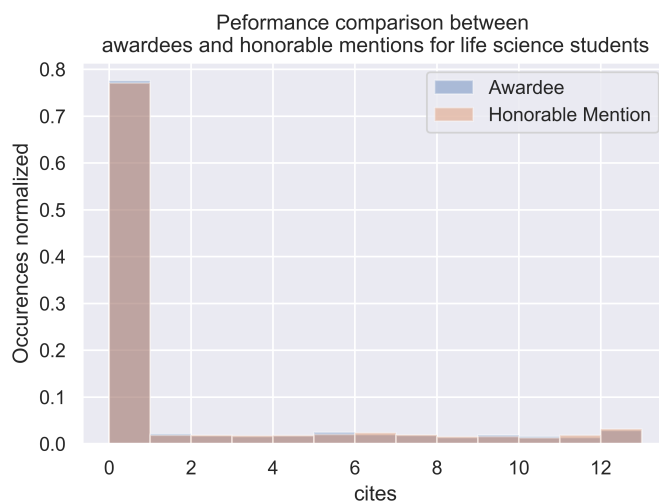


Figure 4.6: Showing the normalized histogram for number of citations for first authored papers summed for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are 0.15 and 0.79 respectively.

4.4.2 Matching students to the Carnegie Classification

The Carnegie Classification [93] is a well known classification system for describing the universities and colleges in the United States. This framework has been widely used in the study of higher education since its start in 1970. We used the Carnegie Classification published in 2015 to classify the baccalaureate and graduate institutions into different classes and investigate whether there are interesting differences between schools from different groups.

Matching Undergraduate Institutions to Carnegie Classification

By matching the baccalaureate institutions to Carnegie Classifications, we can see how students from different types of undergraduate schools perform. Table 5.2 shows the summary statistics for students from each class. Again, since different schools may have different number of students from different fields, and the fact that number of papers or citations cannot be directly compared across majors, it is unfair to compare these metrics directly between classes of schools. As a result, we focus on comparison between the same major between different schools. Fig.4.7 shows the comparison between the largest two groups of undergraduate schools: Doctoral Universities (Highest Research Activity), traditionally known as R1 schools, and Baccalaureate Colleges (Arts & Sciences Focus). It can be observed that the number of students from R1 is many more than the latter.

To compare the two groups statistically, we again restrict the comparison to the broad field with the most number of students - life sciences, and see the performances between students from these two groups are statistically different. As seen in figures 4.8, 4.9, 4.10 and 4.11, there are no statistically significant differences between the two groups except the number of first authored papers. Perhaps surprisingly, students whose undergraduate institutions are Baccalaureate Colleges (Arts & Sciences Focus) have slightly more first authored papers than those from Doctoral Universities (Highest Research Activity). However, as the number of citations summed have no statistical difference, this

Table 5.2: Table showing summary statistics of the performances of students from different classes of baccalaureate institution from the wScholar results ran with the GRFP data set

Category	No. of papers		No. of 1st au. papers		1st au. papers IF		sum citations		No. of students	
	mean	std	mean	std	mean	std	mean	std	mean	std
Doctoral Universities: Highest Research Activity (R1)	3.01	3.74	1.12	1.65	5.81	10.90	39.31	130.87	11561	
Baccalaureate Colleges: Arts & Sciences Focus	2.93	3.64	1.21	1.68	6.09	11.37	45.85	338.21	3410	
Doctoral Universities: Higher Research Activity (R2)	2.71	3.66	1.09	1.65	5.34	10.59	34.02	139.12	1581	
Master's Colleges & Universities: Larger Programs	2.80	3.47	1.12	1.63	5.09	9.60	28.21	90.24	1034	
Doctoral Universities: Moderate Research Activity	2.61	3.71	1.06	1.70	5.29	10.55	25.51	59.47	421	
Master's Colleges & Universities: Medium Programs	2.40	3.07	1.01	1.52	5.13	13.10	25.56	80.36	237	
Master's Colleges & Universities: Small Programs	2.37	2.84	1.05	1.34	4.22	7.41	22.60	61.50	155	
Baccalaureate Colleges: Diverse Fields	3.27	3.45	1.42	1.61	6.91	10.72	31.89	51.77	104	
Special Focus Four-Year: Engineering Schools	2.38	4.05	0.88	1.71	2.97	5.95	16.23	46.32	48	
Tribal Colleges	0.12	0.35	0.00	0.00	0.00	0.00	0.00	0.00	8	
Special Focus Four-Year: Faith-Related Institut...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4	
Special Focus Four-Year: Medical Schools & Centers	2.67	2.31	0.67	0.58	2.29	1.98	22.00	19.05	3	

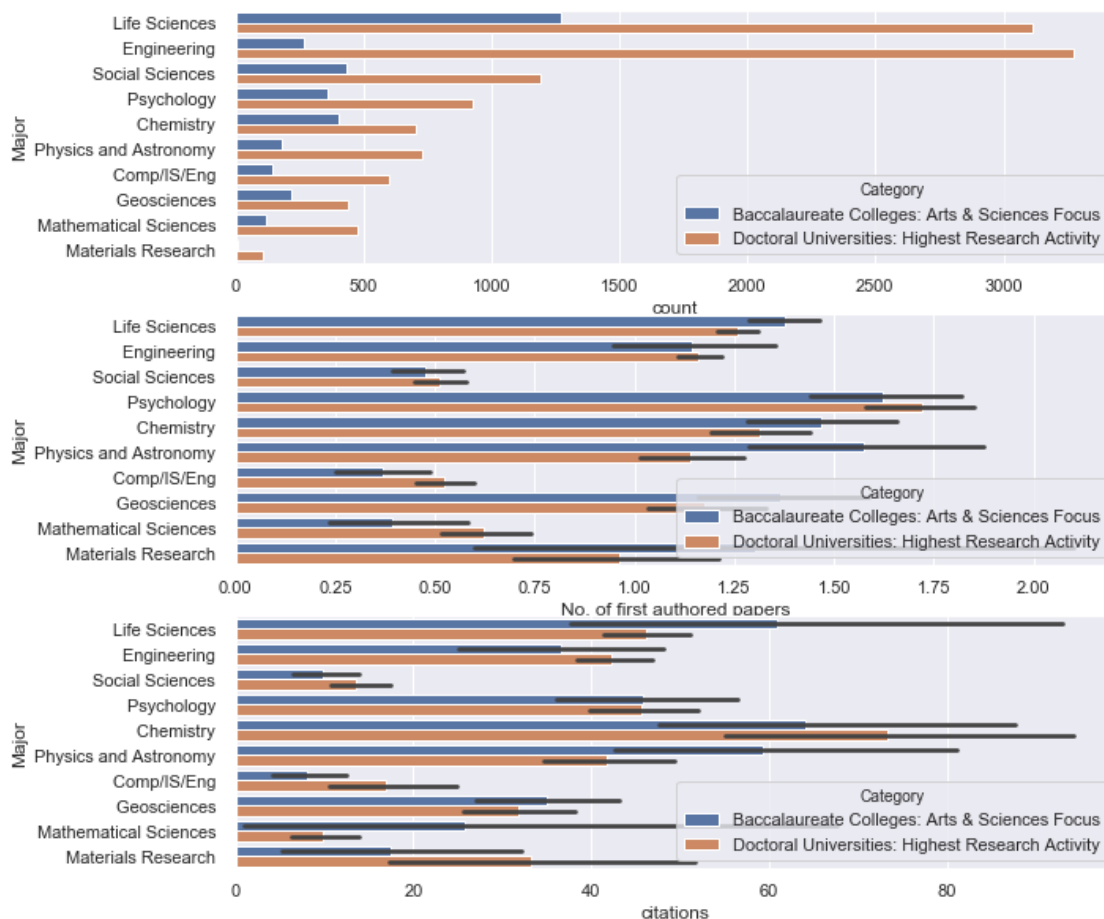


Figure 4.7: Showing the total number of students, average number of first authored papers and average number of citations for each major from each category of undergraduate schools.

result should not be overly interpreted. In any case, this shows that there should not be strong bias in favor of students from universities with more active research activity. It also questions the use of the type of undergraduate institution as a criterion in admission decisions.

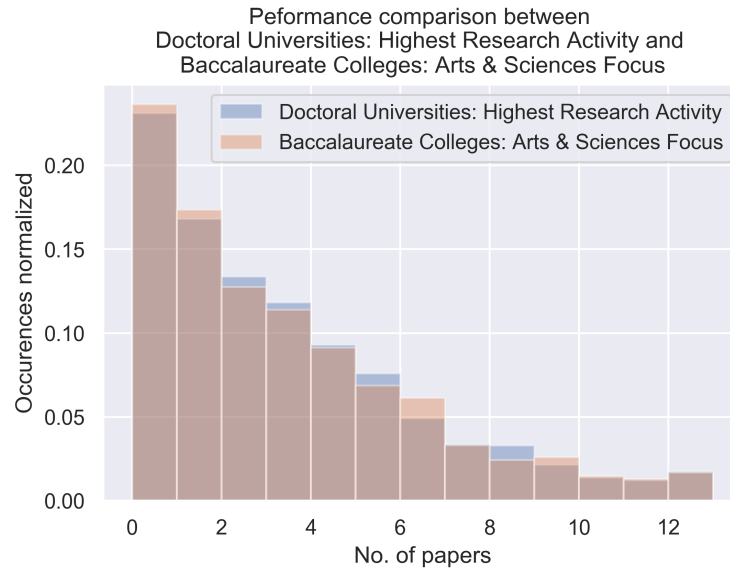


Figure 4.8: Showing the normalized histogram for number of papers for students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.48 and 0.63 respectively.

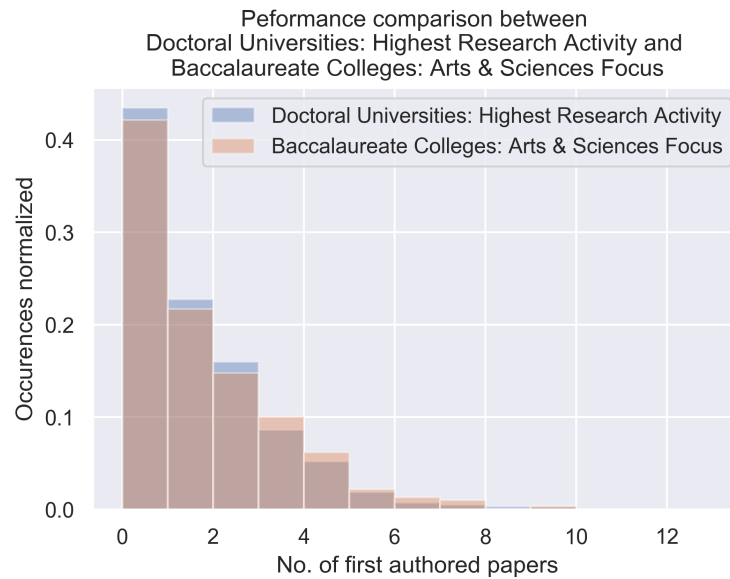


Figure 4.9: Showing the normalized histogram for number of first authored papers for students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.04 and 0.03 respectively. The respective means for R1 and Baccalaureate Colleges are 1.26 and 1.38 respectively.

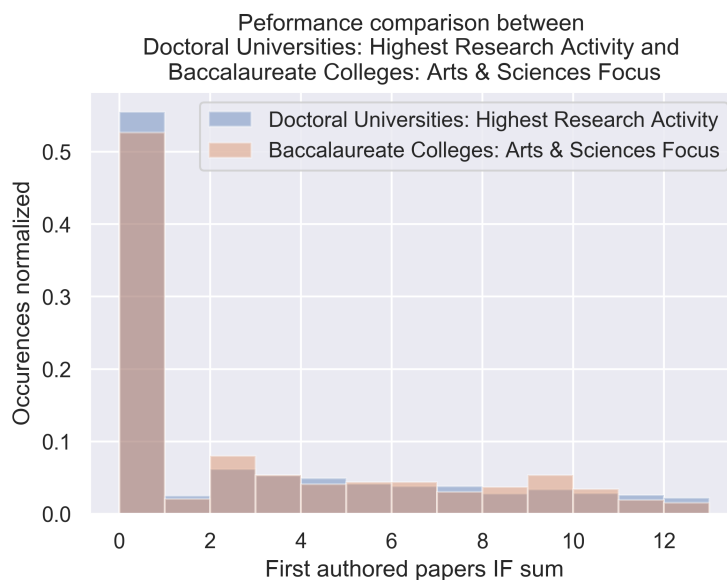


Figure 4.10: Showing the normalized histogram for sum of impact factors of first authored papers by students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.46 and 0.69 respectively.

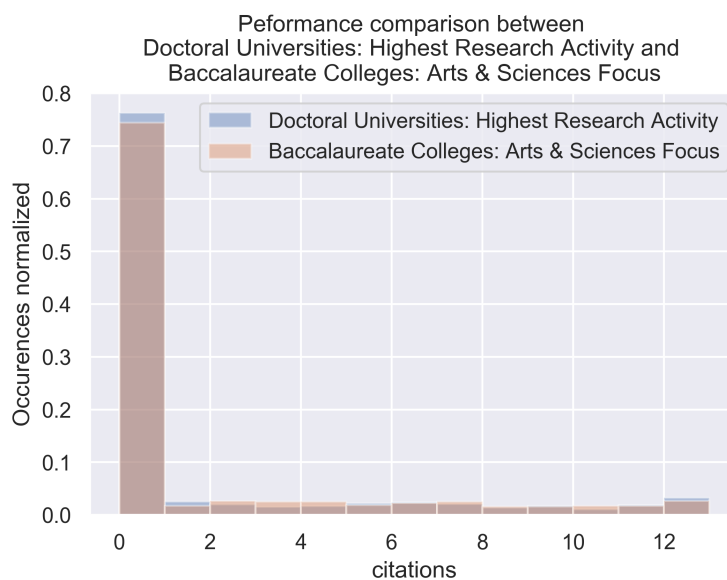


Figure 4.11: Showing the normalized histogram for sum of citations of first authored papers by students from the two groups of undergraduate universities. Mann-Whitney U test and t-test p-values are 0.44 and 0.33 respectively.

Matching Graduate Institutions to Carnegie Classification

The same analysis can be applied to their graduate institutions. Note that this data is noisy in the sense that only the “Current Institutions” is available, and students can be awarded the GRFP as undergraduates. As it is the only proxy for the graduate institution, we treat it as the graduate institution with caution. Table 5.3 shows the summary statistics for each school. It can be seen that the dataset is highly skewed in the sense that most students belong to R1 schools.

Again, as a case study, we restrict the analysis to the largest broad field - life sciences. Figures 4.13, 4.14, 4.15 and 4.16 show the comparison between the students from doctoral universities: Highest vs Higher Research Activity. The differences between the two groups are all statistically significant (with a p-value < 0.05).

The performances of students from R1 schools are all better than R2 schools as measured by the four metrics. This agrees with the general perception that research at R1 schools is more abundant and more cited than at R2 schools.

Table 5.3: Table showing summary statistics of the performances of students from different classes of graduate institution from the wScholar results ran with the GRFP data set

Category	No. of papers		No. of 1st au. papers		1st au. papers IF sum		citations		No. of students	
	mean	std	mean	std	mean	std	mean	std		
Doctoral Universities: Highest Research Activity (R1)	3.06	3.72	1.16	1.66	6.09	11.10	38.58	118.54	14572	
Doctoral Universities: Higher Research Activity (R2)	2.11	3.06	0.96	1.56	3.82	8.21	25.91	135.37	946	
Special Focus Four-Year: Medical Schools & Centers	2.73	2.62	0.97	1.07	7.65	10.54	50.37	119.96	509	
Baccalaureate Colleges: Arts & Sciences Focus	0.80	1.42	0.27	0.71	0.89	2.73	6.23	38.94	497	
Master's Colleges & Universities: Larger Programs	1.18	1.99	0.42	0.87	1.16	3.14	8.47	52.94	197	
Doctoral Universities: Moderate Research Activity	1.10	1.87	0.45	1.00	1.49	5.34	9.85	36.23	137	
Master's Colleges & Universities: Medium Programs	0.48	0.74	0.21	0.61	0.76	2.58	2.55	10.16	42	
Master's Colleges & Universities: Small Programs	0.66	1.29	0.25	0.51	0.59	1.28	3.81	13.53	32	
Baccalaureate Colleges: Diverse Fields	0.60	0.84	0.30	0.67	1.41	3.56	3.90	8.62	10	
Special Focus Four-Year: Engineering Schools	0.50	0.76	0.12	0.35	0.54	1.53	0.38	1.06	8	
Tribal Colleges	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4	

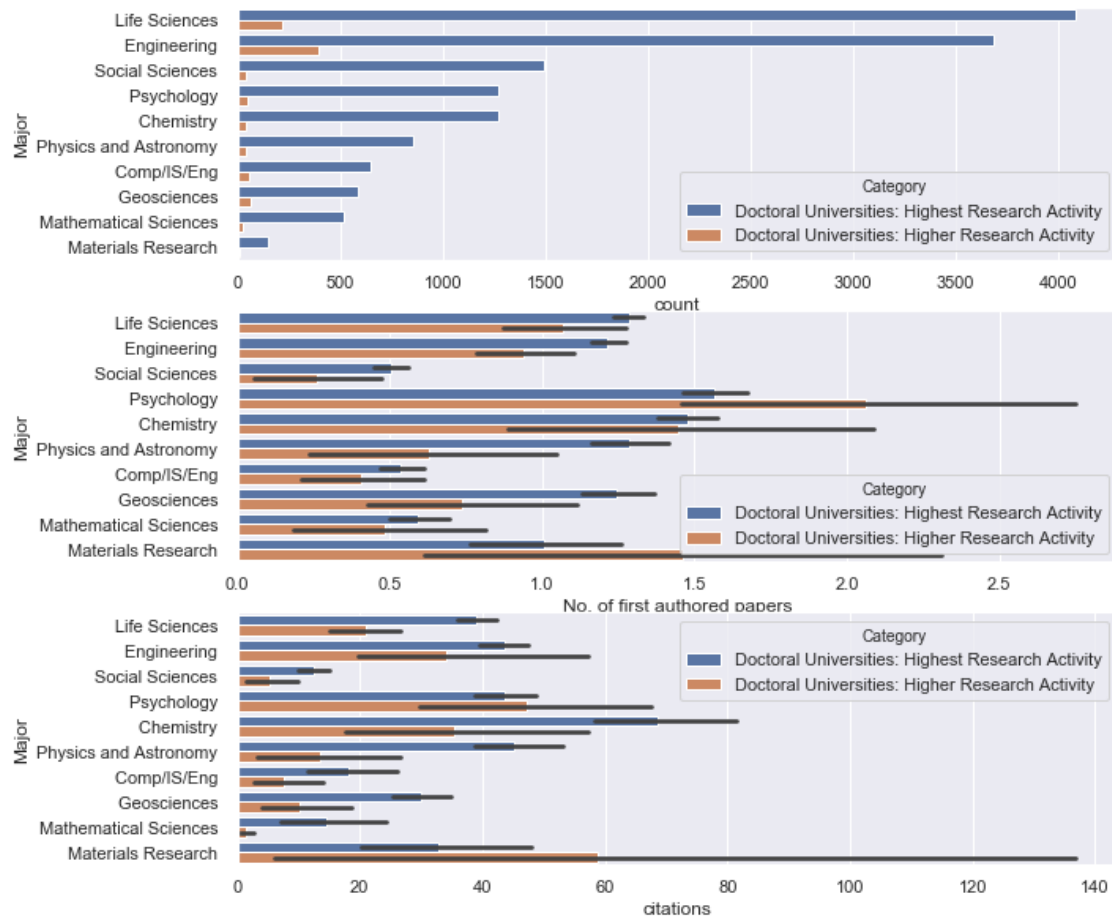


Figure 4.12: Showing the total number of students, average number of first authored papers and average number of citations for each major from each category of graduate schools.

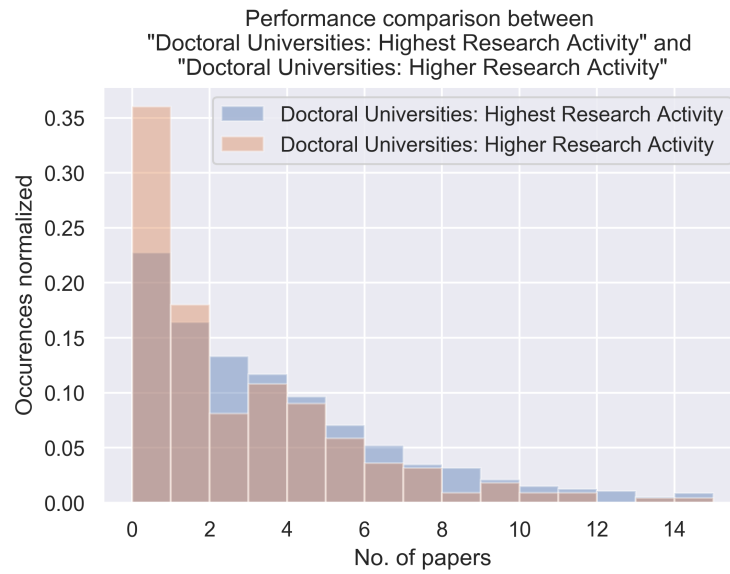


Figure 4.13: Showing the normalized histogram for number of papers for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are $2e-6$ and $9e-7$ respectively. The mean values are 3.3 for R1 and 2.3 for R2 students respectively.

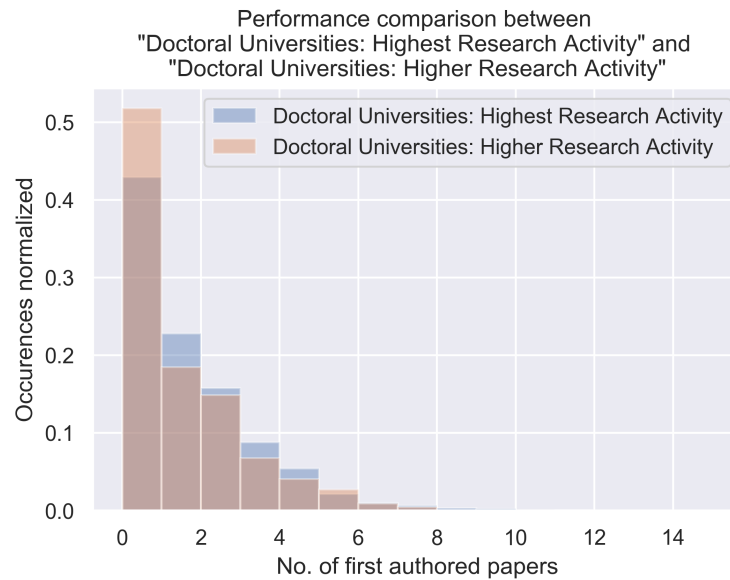


Figure 4.14: Showing the normalized histogram for number of first authored papers for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are 0.01 and 0.03 respectively. The mean values are 1.3 for R1 and 1.1 for R2 students respectively.

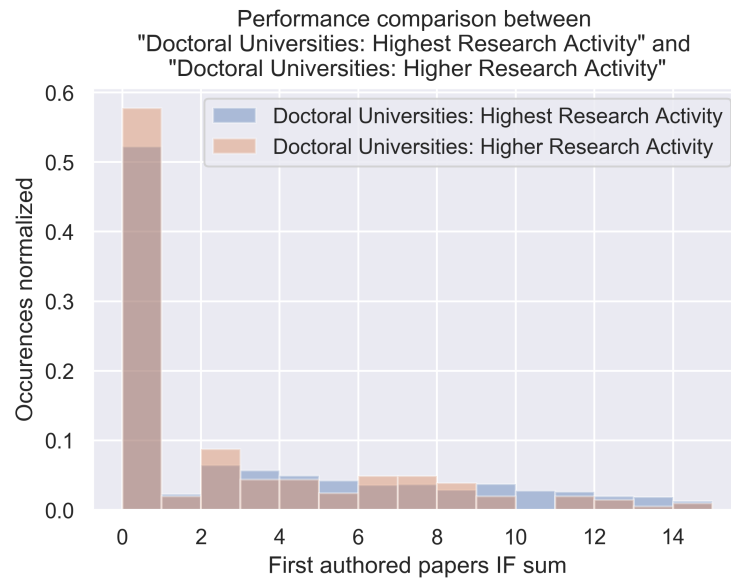


Figure 4.15: Showing the normalized histogram for impact factors for first authored paper summed for students from the two classes of graduate institutions. Mann-Whitney U test and t-test p-values are $1e-4$ and $6e-7$ respectively. The mean values are 6.7 for R1 and 4.2 for R2 students respectively.

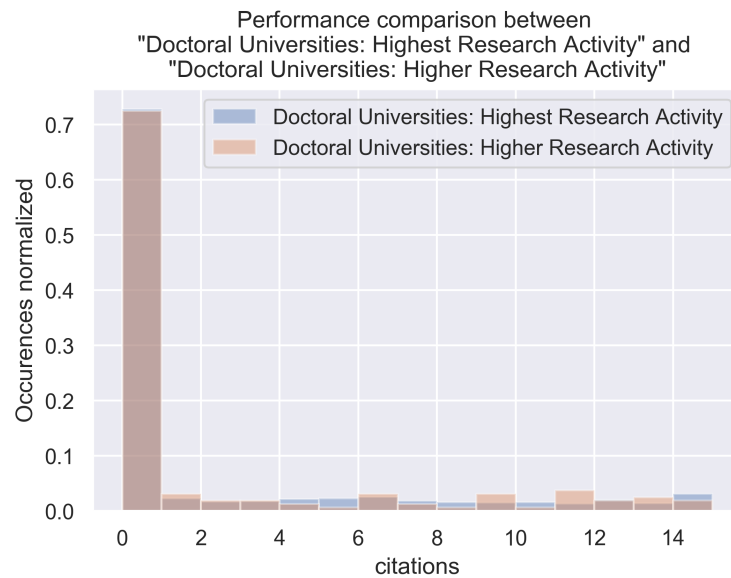


Figure 4.16: Showing the normalized histogram for number of citations for first authored papers summed for awardees and honorable mentions. Mann-Whitney U test and t-test p-values are $2e-3$ and $1e-7$ respectively. The mean values are 39 for R1 and 21 for R2 students respectively.

Interweaving awards information with Carnegie Classification

Next, we discuss the outcomes of wScholar and Award Tracker by breaking down the results by the type of institution using the Carnegie Classification (2015). Of the 25,000 students used in this study, $\sim 11\%$ were subsequently awarded NSF grants, and $\sim 2\%$ of students were subsequently awarded NSF CAREER awards. Students whose graduate institutions are classified as R2 (“Higher Research Activity Universities”) have a higher probability of obtaining an CAREER award ($\sim 2.5\%$) than those with their graduate institutions classified as R1 (“Highest Research Activity Universities”) ($\sim 1.8\%$). Even though the p-values of this finding under Fisher's exact test is 0.11 (one-sided), this outcome is opposite to the usual expectation that students from R1 institutions should perform better than those from R2 institutions. The sample applications summarized above are written in Python Jupyter Notebooks and available on GitHub for ease of reproducibility.

4.5 Conclusion

We hope that the availability of the GradART toolkit introduced here will help facilitate further examinations of different admission strategies. As an illustration, we applied the software to the NSF Graduate Research Fellowships data set. It was found that there is no statistically significant difference between students whose undergraduate institutions are Baccalaureate Colleges (Arts & Sciences Focus) and those from Doctoral Universities (Highest Research Activity) in terms of academic performances, while the performances of students from R1 graduate schools are all better academically than those from R2 graduate schools as measured by our metrics. Interestingly, comparing in terms of probability of obtaining the CAREER award, R2 graduate schools are better than R1 graduate schools, though the difference is not statistically significant according to Fisher's exact test.

With the methodology and toolkits we developed, we have also reviewed outcome data for more than 20,000 PhD graduates who previously applied to Brown University graduate schools. The

analysis of this dataset provided us with unique insights into the merits of GRE scores and the effectiveness of various admission strategies. Unfortunately, due to the confidential nature of this data set, we are unable to illustrate our findings as a demonstration. However, it is our hope that sharing these method will allow others to replicate this review at their own institution.

With the recent active debate (e.g. [94]) about the legal and ethical implications of various criteria used for admissions in higher education, it is our obligation as educators and scientists to continuously examine our own admission practices. We must ensure that the judgment we use to accept or reject applicants produces outcomes that are consistent with our goals and values. Based on the insights from our own internal studies, we strongly encourage other institutions to adopt a similar data-driven methodology to scrutinize the effectiveness of their admissions decisions.

4.6 Acknowledgement

This work is a collaboration between the thesis author and Yang Wang, who developed part of the GradART package, manual verified its accuracy and contributed to the writing of the manuscript to be submitted.

Chapter 5

Conclusions

In this thesis, we discussed experimental design under low signal conditions under three different contexts. We first discussed the demonstration of enhancement in photoluminescence from weak silicon defects emitters proximal to all-dielectric symmetry-broken high contrast gratings. A set of photonic devices with a variety of geometrical parameters which enhances light emission into particular wavelengths and directions by modifying the local density of states was characterized. The engineered dispersions, including Dirac cone, bound state in the continuum, and flat bands, were measured both through partially automated energy momentum spectroscopy with an InGaAs camera, and through PL lifetime measurements with an InGaAs single photon avalanche diode. Even though the noises in InGaAs detectors are inherently large, no specifically designed experiment was required besides long measurement times. The experiments were partially automated to efficiently characterize all devices.

The low-signal experimental conditions encountered prompted us to consider whether the design of the experiment could be optimized. We then proceeded to scrutinize the way spectroscopy and imaging are done. After reviewing the essentials of CCDs, we illustrated the characterization of a CCD camera to obtain its noise statistics. Then, with a noise model composed of Poisson shot noise and Gaussian readout noise, the EM algorithm was applied to denoising. It was shown that

combining single pixel and binned pixel measurements is generally inferior to spending all time on single pixel measurements. The Cramér-Rao lower bound applied to the two pixel case also suggests that information content is lost when a given measurement time is shared between two measurements instead of just a single pixel measurement. We then looked at cases where there is significant spatial correlation between neighbouring pixels, which is usually the case in spectroscopy. In both the cases for one and two Gaussian peaks, the binned measurement was superior. We expected that was because the sacrifice in spatial resolution was no longer a problem when there was strong spatial correlation, while the enhancement in signal-to-noise ratio enhanced the localization accuracy.

Finally, we looked at a broader scientific challenge. As graduate school admission is a process complicated with inherently large variances and restricted by limited data, we reported a method and developed software tools for obtaining and analyzing academic outcome data for graduate students, resulting in generally an order of magnitude more data for strengthening the feedback mechanism and enabling a more thorough review of admission practices. As an illustration, we applied the software to the NSF Graduate Research Fellowships data set. It was found that there is no statistically significant difference between students whose undergraduate institutions are Baccalaureate Colleges (Arts & Sciences Focus) and those from Doctoral Universities (Highest Research Activity) in terms of academic performances, while the research output of students from R1 graduate schools generally exceeded those from R2 graduate schools as measured by our metrics. Interestingly, comparing in terms of probability of obtaining the CAREER award, R2 graduate schools are better than R1 graduate schools, though the difference is not statistically significant according to Fisher's exact test. We hope the developed tools and the illustration will encourage other institutions to adopt a similar data-driven methodology to scrutinize the effectiveness of their admissions decisions.

Bibliography

- [1] Hai Son Nguyen, Florian Dubois, Thierry Deschamps, Sébastien Cueff, Antonin Pardon, J-L Leclercq, Christian Seassal, Xavier Letartre, and Pierre Viktorovitch. Symmetry breaking in photonic crystals: On-demand dispersion from flatband to dirac cones. *Physical Review Letters*, 120(6):066102, 2018.
- [2] Ralf Widenhorn, Morley M Blouke, Alexander Weber, Armin Rest, and Erik Bodegom. Temperature dependence of dark current in a CCD. In *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III*, volume 4669, pages 193–201. International Society for Optics and Photonics, 2002.
- [3] Princeton Instruments. PIXIS: 1024 eXcelon Rev. N.3.1. [Online]. Available from: https://www.princetoninstruments.com/userfiles/files/assetLibrary/Datasheets/Princeton_Instruments_PIXIS_1024_eXcelon_N3_1_10_27_14.pdf, 2019.
- [4] James R Janesick, Kenneth P Klaasen, and Tom Elliott. Charge-coupled-device charge-collection efficiency and the photon-transfer technique. *Optical Engineering*, 26(10):261072, 1987.
- [5] Travis Oliphant. NumPy: A guide to NumPy. USA: Trelgol Publishing, 2006–. [Online; accessed today].

- [6] Cornell University Graduate School. Cornell University Doctoral Program Statistics. [Online]. Available from: <https://gradschool.cornell.edu/degrees-fields/program-metrics-assessments-and-outcomes/facts-figures/>, 2019.
- [7] Robert Sowell. PhD completion and attrition: Analysis of baseline data. *Washington, DC: Council of Graduate Schools*, pages 1–23, 2008.
- [8] David Roxbee Cox and Nancy Reid. *The theory of the design of experiments*. Chapman and Hall/CRC, 2000.
- [9] Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.
- [10] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.
- [11] Luc Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44(2):303–325, 2008.
- [12] Jack PC Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3):707–716, 2009.
- [13] Wolfram. Hadamard Matrix. [Online]. Available from: <http://mathworld.wolfram.com/HadamardMatrix.html>, 2019.
- [14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [15] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

- [16] Benjamin Letham, Brian Karrer, Guilherme Ottoni, and Eytan Bakshy. Constrained bayesian optimization with noisy experiments. *Bayesian Anal.*, 14(2):495–519, 06 2019.
- [17] Mitchell McIntire, Tyler Cope, Stefano Ermon, and Daniel Ratner. Bayesian Optimization of FEL Performance at LCLS. In *Proceedings, 7th International Particle Accelerator Conference (IPAC 2016): Busan, Korea, May 8-13, 2016*, page WEPOW055, 2016.
- [18] E. M. Purcell. Spontaneous emission probabilities at radio frequency. *Phys.Rev.*, page p. 681, 1946.
- [19] K. H. Drexhage, H. Kuhn, and F. P. Schfer. Variation of the fluorescence decay time of a molecule in front of a mirror. *Berichte der Bunsengesellschaft fr physikalische Chemie*, 72(2):329–329, 1968.
- [20] Gunnar Björk, Susumu Machida, Yoshihisa Yamamoto, and Kazuhiro Igeta. Modification of spontaneous emission rate in planar dielectric microcavity structures. *Phys. Rev. A*, 44:669–681, Jul 1991.
- [21] Dirk Englund, David Fattal, Edo Waks, Glenn Solomon, Bingyang Zhang, Toshihiro Nakaoka, Yasuhiko Arakawa, Yoshihisa Yamamoto, and Jelena Vučković. Controlling the spontaneous emission rate of single quantum dots in a two-dimensional photonic crystal. *Phys. Rev. Lett.*, 95:013904, Jul 2005.
- [22] Matthew Pelton. Modified spontaneous emission in nanophotonic structures. *Nature Photonics*, 9(7):427, 2015.
- [23] Jacob B Khurgin. How to deal with the loss in plasmonics and metamaterials. *Nature Nanotechnology*, 10(1):2, 2015.
- [24] Yasushi Takahashi, Hiroyuki Hagino, Yoshinori Tanaka, Bong-Shik Song, Takashi Asano, and Susumu Noda. High-q nanocavity with a 2-ns photon lifetime. *Opt. Express*, 15(25):17206–17213, Dec 2007.

- [25] Connie J Chang-Hasnain. High-contrast gratings as a new platform for integrated optoelectronics. *Semiconductor Science and Technology*, 26(1):014043, 2010.
- [26] Adam C Overvig, Sajan Shrestha, and Nanfang Yu. Dimerized high contrast gratings. *Nanophotonics*, 7(6):1157–1168, 2018.
- [27] Lukas Novotny and Bert Hecht. *Principles of nano-optics*. Cambridge University Press, 2012.
- [28] R. Lo Savio, S. L. Portalupi, D. Gerace, A. Shakoor, T. F. Krauss, L. O’Faolain, L. C. Andreani, and M. Galli. Room-temperature emission at telecom wavelengths from silicon photonic crystal nanocavities. *Applied Physics Letters*, 98(20):201106, 2011.
- [29] Jonathan A Kurvits, Mingming Jiang, and Rashid Zia. Comparative analysis of imaging configurations and objectives for Fourier microscopy. *JOSA A*, 32(11):2082–2092, 2015.
- [30] EN Bulgakov, DN Maksimov, PN Semina, and SA Skorobogatov. Propagating bound states in the continuum in dielectric gratings. *JOSA B*, 35(6):1218–1222, 2018.
- [31] Zin Lin, Lysander Christakis, Yang Li, Eric Mazur, Alejandro W. Rodriguez, and Marko Lončar. Topology-optimized dual-polarization dirac cones. *Phys. Rev. B*, 97:081408, Feb 2018.
- [32] Sabyasachi Barik, Hirokazu Miyake, Wade DeGottardi, Edo Waks, and Mohammad Hafezi. Two-dimensionally confined topological edge states in photonic crystals. *New Journal of Physics*, 18(11):113013, nov 2016.
- [33] Juntao Li, Thomas P. White, Liam O’Faolain, Alvaro Gomez-Iglesias, and Thomas F. Krauss. Systematic design of flat band slow light in photonic crystal waveguides. *Opt. Express*, 16(9):6227–6232, Apr 2008.
- [34] T. Jacqmin, I. Carusotto, I. Sagnes, M. Abbarchi, D. D. Solnyshkov, G. Malpuech, E. Galopin, A. Lemaître, J. Bloch, and A. Amo. Direct observation of dirac cones and a flatband in a honeycomb lattice for polaritons. *Phys. Rev. Lett.*, 112:116402, Mar 2014.

- [35] Grating dispersion calculator. [Online]. Available from: <https://www.princetoninstruments.com/calculators/grating-dispersion.cfm>. Accessed: 2019-09-24.
- [36] PicoQuant. Time-correlated single photon counting. [Online]. Available from: https://www.picoquant.com/images/uploads/page/files/7253/technote_tcspc.pdf, 2019.
- [37] Dongfang Li, Sinan Karaveli, Sébastien Cuff, Wenhao Li, and Rashid Zia. Probing the combined electromagnetic local density of optical states with quantum emitters supporting strong electric and magnetic transitions. *Phys. Rev. Lett.*, 121:227403, Nov 2018.
- [38] Sébastien Cuff, Dongfang Li, You Zhou, Franklin J. Wong, Jonathan A. Kurvits, Shriram Ramanathan, and Rashid Zia. Dynamic control of light emission faster than the lifetime limit using VO₂ phase-change. *Nature Communications*, 6(1):8636, 2015.
- [39] Teledyne. CCD vs. CMOS. [Online]. Available from: <https://www.teledynedalsa.com/en/learn/knowledge-center/ccd-vs-cmos/>, 2019.
- [40] Jerry Chao, E Sally Ward, and Raimund J Ober. Fisher information theory for parameter estimation in single molecule microscopy: tutorial. *JOSA A*, 33(7):B36–B57, 2016.
- [41] Donald L Snyder, Carl W Helstrom, Aaron D Lanterman, Mohammad Faisal, and Richard L White. Compensation for readout noise in CCD images. *JOSA A*, 12(2):272–283, 1995.
- [42] Donald L Snyder, Abed M Hammoud, and Richard L White. Image recovery from data acquired with a charge-coupled-device camera. *JOSA A*, 10(5):1014–1023, 1993.
- [43] Philip Hobbs. *Building Electro-Optical Systems*. Wiley, 2 edition, 2000.
- [44] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical Journal*, 86(2):1185–1200, 2004.

- [45] Jerry Chao, E. Sally Ward, and Raimund J. Ober. Fisher information theory for parameter estimation in single molecule microscopy: tutorial. *J. Opt. Soc. Am. A*, 33(7):B36–B57, Jul 2016.
- [46] Henri Lantéri and Céline Theys. Restoration of astrophysical images: the case of Poisson data with additive Gaussian noise. *EURASIP Journal on Applied Signal Processing*, 2005:2500–2513, 2005.
- [47] Núñez, J. and Llacer, J. Bayesian image reconstruction with space-variant noise suppression. *Astron. Astrophys. Suppl. Ser.*, 131(1):167–180, 1998.
- [48] Jorge Nunez and Jorge Llacer. A general Bayesian image reconstruction algorithm with entropy prior: Preliminary application to HST data. *Publications of the Astronomical Society of the Pacific*, 105:1192, Oct 1993.
- [49] Princeton Instruments. Frame Transfer Cameras. [Online]. Available from: <http://depts.washington.edu/keck/pg22.pdf>, 2019.
- [50] James R Janesick. Photon transfer. *Bellingham: SPIE*, 2007.
- [51] Patrick M Epperson and M Bonner Denton. Binning spectral images in a charge-coupled device. *Analytical Chemistry*, 61(14):1513–1519, 1989.
- [52] Atik Cameras. [Online]. Available from: <https://www.atik-cameras.com/news/binning-the-differences-between-cmos-and-ccd/>, 2019.
- [53] Andor. [Online]. Available from: <https://andor.oxinst.com/learning/view/article/ccd-binning>, 2019.
- [54] H. Nasibov, A. Kholmatov, B. Akselli, A. Nasibov, and S. Baytaroglu. Performance analysis of the CCD pixel binning option in particle-image velocimetry measurements. *IEEE/ASME Transactions on Mechatronics*, 15(4):527–540, Aug 2010.

- [55] M Bertero and P Boccacci. Application of the OS-EM method to the restoration of LBT images. *Astronomy and Astrophysics Supplement Series*, 144(1):181–186, 2000.
- [56] Ming Yan, Alex AT Bui, Jason Cong, and Luminita A Vese. General convergent expectation maximization (EM)-type algorithms for image reconstruction. *Inverse Problems & Imaging*, 7(3):1007–1029, 2013.
- [57] R Sivakumar. Denoising of computer tomography images using curvelet transform. *ARPN Journal of Engineering and Applied Sciences*, 2(1):21–26, 2007.
- [58] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65 vol. 2, June 2005.
- [59] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.
- [60] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [61] Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on Medical Imaging*, 1(2):113–122, 1982.
- [62] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90 – 100, 2003.
- [63] Greg Hamerly and Charles Elkan. Alternatives to the K-means algorithm that find better clusterings. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management, CIKM '02*, pages 600–607, New York, NY, USA, 2002. ACM.

- [64] Kristie Seymore, Andrew McCallum, and Roni Rosenfeld. Learning hidden Markov model structure for information extraction. In *AAAI-99 workshop on machine learning for information extraction*, pages 37–42, 1999.
- [65] F Benvenuto, A La Camera, C Theys, A Ferrari, H Lantéri, and M Bertero. The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise. *Inverse Problems*, 24(3):035016, 2008.
- [66] A. Jezierska, C. Chaux, J. Pesquet, and H. Talbot. An EM approach for Poisson-Gaussian noise modeling. In *2011 19th European Signal Processing Conference*, pages 2244–2248, Aug 2011.
- [67] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001. [Online; accessed 2019].
- [68] J Michael Hollas. *Modern spectroscopy*. John Wiley & Sons, 2004.
- [69] Joshua D Hall, Anna B OConnell, and Jeanette G Cook. Predictors of student productivity in biomedical graduate school applications. *PLoS One*, 12(1):e0169121, 2017.
- [70] Liane Moneta-Koehler, Abigail M Brown, Kimberly A Petrie, Brent J Evans, and Roger Chalkley. The limitations of the GRE in predicting success in biomedical graduate school. *PloS One*, 12(1):e0166742, 2017.
- [71] Casey Miller and Keivan Stassun. A test that fails. *Nature*, 510(7504):303–304, 2014.
- [72] Nathan R Kuncel, Sarah A Hezlett, and Deniz S Ones. A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: implications for graduate student selection and performance. *Psychological Bulletin*, 127(1):162, 2001.
- [73] Nathan R Kuncel, Serena Wee, Lauren Serafin, and Sarah A Hezlett. The validity of the Graduate Record Examination for masters and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, 70(2):340–352, 2010.

- [74] Todd Morrison and Melanie Morrison. A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with graduate grade point average representing the criterion of graduate success. *Educational and Psychological Measurement*, 55(2):309–316, 1995.
- [75] Lisa M Schneider and Jacqueline B Briel. *Validity of the GRE: 1988-89 summary report*. Educational Testing Service, 1990.
- [76] Robert J Sternberg and Wendy M Williams. Does the Graduate Record Examination predict meaningful success in the graduate training of psychology? a case study. *American Psychologist*, 52(6):630, 1997.
- [77] Sarah E Newton and Gary Moore. Undergraduate grade point average and Graduate Record Examination scores: the experience of one graduate nursing program. *Nursing Education Perspectives*, 28(6):327–331, 2007.
- [78] David M Klieger, Frederick A Cline, Steven L Holtzman, Jennifer L Minsky, and Florian Lorenz. New perspectives on the validity of the GRE® general test for predicting graduate school grades. *ETS Research Report Series*, 2014(2):1–62, 2014.
- [79] Inge TL Schwager, Ute R Hülshager, Brent Bridgeman, and Jonas WB Lang. Graduate student selection: Graduate Record Examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment*, 23(1):71–79, 2015.
- [80] Orion D Weiner. How should we be selecting our graduate students? *Molecular Biology of the Cell*, 25(4):429–430, 2014.
- [81] Cooper B Holmes and Megan J Beishline. Correct classification, false positives, and false negatives in predicting completion of the Ph.D. from GRE scores. *Psychological Reports*, 79(3):939–945, 1996.

- [82] Amanda J Rockinson-Szapkiw, Oliver R Bray Jr, and Lucinda S Spaulding. Examining the predictive validity of GRE scores on doctoral education: Students' success and methodology choices in the dissertation process. *Journal of College Student Retention: Research, Theory & Practice*, 16(2):203–217, 2014.
- [83] Nancy W Burton and Ming-mei Wang. Predicting long-term success in graduate school: A collaborative validity study. *ETS Research Report Series*, 2005(1):i–61, 2005.
- [84] Andrew J Thacker and Robert E Williams. The relationship of the Graduate Record Examination to grade point average and success in graduate school. *Educational and Psychological Measurement*, 34(4):939–944, 1974.
- [85] Kenneth Oldfield and Janet R. Hutchinson. Predictive validity of the Graduate Record Examination with and without range restrictions. *Psychological Reports*, 81(1):211–220, 1997.
- [86] Mildred L Sleeper. Relationship of scores on the Graduate Record Examination to grade point averages of graduate students in occupational therapy. *Educational and Psychological Measurement*, 21(4):1039–1040, 1961.
- [87] Michael M Omizo and William B Michael. The prediction of performance in a counselor education master's degree program. *Educational and Psychological Measurement*, 39(2):433–437, 1979.
- [88] William B Schrader. Admissions test scores as predictors of career achievement in psychology. *ERIC*, 1978.
- [89] George Yule and Paul Hoffman. Predicting success for international teaching assistants in a US university. *TESOL Quarterly*, 24(2):227–243, 1990.
- [90] Nancy N Vacc and Rosalind Picot. Predicting success in doctoral study. *College Student Journal*, 1984.

- [91] National Student Clearinghouse. National Student Clearinghouse. <https://studentclearinghouse.org>, 2019.
- [92] National Science Foundation. Graduate Research Fellowship Program. [Online]. Available from: <https://www.research.gov/grfp/AwardeeList.do?method=loadAwardeeList>, 2019.
- [93] The Carnegie Classification of Institutions of Higher Education. Carnegie Classification. [Online]. Available from: <http://carnegieclassifications.iu.edu/>, 2019.
- [94] The New York Times. Harvard Rated Asian-American Applicants Lower on Personality Traits, Suit Says. [Online]. Available from: <https://www.nytimes.com/2018/06/15/us/harvard-asian-enrollment-applicants.html>, 2018.